

Journal of **Geophysics** Zeitschrift für **Geophysik**

Volume 50 1981–82

Managing Editors

W. Dieminger, J. Untiedt

Editorial Board

K.M. Creer Edinburgh, Scotland
W. Dieminger Lindau, F.R.G.
C. Kisslinger Boulder, Colorado
Th. Krey Hannover, F.R.G.
G. Müller Frankfurt, F.R.G.
G.C. Reid Boulder, Colorado
J. Untiedt Münster, F.R.G.
S. Uyeda Tokyo, Japan

Advisory Board

G. Angenheister , München	I. P. Kosminskaja , Moskwa
A. A. Ashour , Cairo	W. Krauss , Kiel
W. I. Axford , Lindau/Harz	St. Müller , Zürich
J. Behrens , Berlin	A. Roche , Strasbourg
V. Bucha , Praha	O. Rosenbach , Clausthal-Zellerfeld
J. Cain , Greenbelt, MD	S. Saxov , Aarhus
K. Fuchs , Karlsruhe	U. Schmucker , Göttingen
N. Fukushima , Tokyo	M. Siebert , Göttingen
V. Haak , Berlin	H. Soffel , München
B. Haurwitz , Fort Collins, CO	L. Stegena , Budapest
W. Jacoby , Frankfurt a.M.	H. Stiller , Potsdam



Springer International

Journal of Geophysics – Zeitschrift für Geophysik

This journal was founded by the Deutsche Geophysikalische Gesellschaft on the initiative of L. Mintrop in 1924 as the Zeitschrift für Geophysik and edited by G. Angenheister from Vol. 1–18 (1944). It reappeared in 1954 edited by B. Brockamp from Vol. 19–26 (1960), and edited by W. Dieminger from Vol. 27–33 (1961), and from Vol. 34 (1968) edited by W. Dieminger and J. Untiedt. After Vol. 40 (1970) the title was changed to Journal of Geophysics – Zeitschrift für Geophysik.

Published: Vols. 19–39 by Physica-Verlag, Würzburg, from Vol. 40 by Springer Berlin, Heidelberg, New York.

The exclusive copyright for all languages and countries, including the right for photomechanical and any other reproductions, also in microform, is transferred to the Deutsche Geophysikalische Gesellschaft.

The use of registered names, trademarks, etc. in this publications does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Authors of this journal can benefit from library and photocopy fees collected by VG WORT if certain conditions are met. If an author lives in the Federal Republic of Germany or in West Berlin it is recommended that he contacts Verwertungsgesellschaft WORT, Abteilung Wissenschaft, Goethestraße 49, D-8000 München 2, for detailed information.

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, sind vorbehalten. Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung der Deutschen Geophysikalischen Gesellschaft in irgendeiner Form – durch Fotokopie, Mikrofilm oder andere Verfahren – reproduziert oder in eine von Maschinen, insbesondere von Datenverarbeitungsanlagen, verwendbare Sprache übertragen werden.

Auch die Rechte der Wiedergabe durch Vortrag, Funk- und Fernsehsendung, im Magnettonverfahren oder ähnlichem Wege bleiben vorbehalten.

Fotokopien für den persönlichen und sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopien hergestellt werden. Jede im Bereich eines gewerblichen Unternehmens hergestellte oder benutzte Kopie dient gewerblichen Zwecken gem. § 54 (2) UrhG und verpflichtet zur Gebührenzahlung an die VG WORT, Abteilung Wissenschaft, Goethestraße 49, D-8000 München 2, von der die einzelnen Zahlungsmodalitäten zu erfragen sind.

Autoren dieser Zeitschrift können unter gewissen Voraussetzungen in die Individualausschüttung von Mitteln aus der Bibliothekantieme und dem Fotokopieraufkommen mit einbezogen werden. Genaue Informationen erteilt die Verwertungsgesellschaft WORT, Abteilung Wissenschaft, Goethestraße 49, D-8000 München 2.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in dieser Zeitschrift berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, daß solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Springer-Verlag Berlin Heidelberg New York

Printed in Germany by Universitätsdruckerei H. Stürtz AG Würzburg

© by the Deutsche Geophysikalische Gesellschaft, Clausthal-Zellerfeld, 1982

z 1983 4137



Author Index

- Ádám, A. 37
André, D. 194
- Babuška, V. 1
Baumjohann, W. 194
Bykov, A.M. 221
- Cevolani, G. 212
Charleton, P. 202
- Dardi, D. 212
Duma, G. 37
- Eichler, D. 110
Ellison, D.C. 110
- Ferrari, A. 70
Formisano, V. 114
Fuller-Rowell, T.J. 202
- Geranios, A. 124
Goldberg, S. 151
Gombosi, T. 101
- Gringauz, K.I. 101
Gutdeutsch, R. 37
- Hellmann, K.N. 73
Hussain, A.G. 16
- Jacoby, W.R. 89
Jain, A.K. 127
Jones, A.G. 23
Jones, F.C. 110
- Keppler, E. 45, 101
Khorana, R. 127
Kirsch, E. 45
Klootwijk, C.T. 127
Koch, M. 125
Kono, M. 137
Korn, M. 171
Kurt, V.G. 101
- Lippolt, H.J. 73
Loewenthal, D. 151
- März, F. 37
Martinic, N. 45
McWhirter, I. 202
Müller, G. 143, 177
- Neubauer, F.M. 101
Nevanlinna, H. 68
Nishitani, T. 137
Novotný, M. 7
- Orsini, S. 114
- Pfotzer, G. † 123
- Räkers, E. 143
Rees, D. 202
Richter, A.K. 101
Rosenbauer, H. 101
Rostoker, G. 51
Rotstein, Y. 151
Rounce, P.A. 202
- Schmeling, H. 89
Schult, A. 16
Smith, K. 202
Soffel, H.C. 16
Somogyi, A. 101
Stöckl, H. 171
Stolpovsky, V.G. 101
Sucksdorff, C. 68
- Tarantola, A. 159
Temme, P. 177
Tighe, W.G. 51
Toptyghin, I.N. 221
Torkar, K.M. 189
Trussoni, E. 70
- Urban, A. 189
- Valette, B. 159
Verigin, M.I. 101
Verö, J. 37
- Wallner, Á. 37
Wilhelm, K. 189

Subject Index

Acceleration of Particles by Shock Waves

- Turbulent Acceleration of Electrons in Extended Extragalactic Ratio Sources (Ferrari, A., Trussoni, E.) 70
- The 3 January 1978 Interplanetary Shock Event as Observed by Energetic Particle, Plasma and Magnetic Field Devices on Board of Helios-1, Helios-2 and PROGNOZ-6 (Richter, A.K., et al.) 101
- Monte Carlo Simulation of Collisionless Shocks Showing Preferential Acceleration of High A/Z Particles (Ellison, D.C., et al.) 110
- Acceleration Processes at the Earth's Bow Shock (Formisano, V., Orsini, S.) 114
- The Theory of Particle Acceleration in Astrophysical Objects Containing Shock Waves and Turbulent Plasma Motions (Bykov, A.M., Toptyghin, I.N.) 221

Africa

- Palaeomagnetism of Upper Cretaceous Volcanics and Nubian Sandstones of Wadi Natash, SE Egypt and Implications for the Polar Wander Path for Africa in the Mesozoic (Schult, A., et al.) 16

Alps

- Magnetotelluric Sounding in the Transitional Zone Between the Eastern Alps and Pannonian Basin (Ádám, A., et al.) 37

Auroral Zone Physics

- Characteristics of Westward Travelling Surges During Magnetospheric Substorms (Tighe, W.G., Rostoker, G.) 51
- Energy Dispersion and Acceleration of Low Energy Protons and Their Relation to Electrons During an Auroral Breakup (Urban, A., et al.) 189
- Joint Two-Dimensional Observations of Ground Magnetic and Ionospheric Electric Fields Associated with Auroral Currents – 5. Current System Associated with Eastward Drifting Omega Bands (André, D., Baumjohann, W.) 194
- Thermospheric Winds During the Energy Budget Campaign: Ground-Based Fabry-Perot Observations Supported by Dynamical Simulations with a Three-Dimensional, Time-Dependent Thermospheric Model (Rees, D., et al.) 202

Coal Prospecting

- Reflection and Transmission of Love Channel Waves at Coal Seam Discontinuities Computed with a Finite Difference Method (Korn, M., Stöckl, H.) 171
- Numerical Simulation of Vertical Seismic Profiling (Temme, P., Müller, G.) 177

Earthquakes

- The Romanian Earthquake of March 4, 1977 – III. Improved Focal Model and Moment Determination (Räkers, E., Müller, G.) 143

Earth's Bow Shock

- Acceleration Processes at the Earth's Bow Shock (Formisano, V., Orsini, S.) 114

Electromagnetic Induction in the Earth

- Geomagnetic Induction Studies in Scandinavia – II. Geomagnetic Depth Sounding, Induction Vectors and Coast-Effect (Jones, A.G.) 23

Geochronology

- Calibration of the Middle Triassic Time Scale by Conventional K-Ar and ⁴⁰Ar/³⁹Ar Dating of Alkali Feldspars (Hellmann, K.N., Lippolt, H.J.) 73

India

- Palaeomagnetic Constraints on Allochthony and Age of the Krol Belt Sequence, Garhwal Himalaya, India (Klootwijk, C.T., et al.) 127

Interplanetary Medium

- The 3 January 1978 Interplanetary Shock Event as Observed by Energetic Particle, Plasma and Magnetic Field Devices on Board of Helios-1, Helios-2 and PROGNOZ-6 (Richter, A.K., et al.) 101
- Acceleration Processes at the Earth's Bow Shock (Formisano, V., Orsini, S.) 114

Inverse Problems

- Two Methods of Solving the Linearized Two-Dimensional Inverse Seismic Kinematic Problem (Novotný, M.) 7
- An Improved Algorithm for Magnetotelluric and Direct Current Data Interpretation (Goldberg, S., et al.) 151
- Inverse Problems = Quest for Information (Tarantola, A., Valette, B.) 159

Lithosphere

- On Modelling the Lithosphere in Mantle Convection with Non-Linear Rheology (Schmeling, H., Jacoby, W.R.) 89

Magnetotellurics

- Magnetotelluric Sounding in the Transitional Zone Between the Eastern Alps and Pannonian Basin (Ádám, A., et al.) 37
- An Improved Algorithm for Magnetotelluric and Direct Current Data Interpretation (Goldberg, S., et al.) 151

Mantle Convection

- On Modelling the Lithosphere in Mantle Convection with Non-Linear Rheology (Schmeling, H., Jacoby, W.R.) 89

Palaeomagnetism

- Palaeomagnetism of Upper Cretaceous Volcanics and Nubian Sandstones of Wadi Natash, SE Egypt and Implications for the Polar Wander Path for Africa in the Mesozoic (Schult, A., et al.) 16
- Palaeomagnetic Constraints on Allochthony and Age of the Krol Belt Sequence, Garhwal Himalaya, India (Klootwijk, C.T., et al.) 127

Rock Magnetism

- Grain Size Effect on the Low-Temperature Oxidation of Titanomagnetite (Nishitani, T., Kono, M.) 137

Rock Physics

- Anisotropy of V_p and V_s in Rock-Forming Minerals (Babuška, V.) 1

Rocket Experiments

- Energy Dispersion and Acceleration of Low Energy Protons and Their Relation to Electrons During an Auroral Breakup (Urban, A., et al.) 189

Scandinavia

- Geomagnetic Induction Studies in Scandinavia – II. Geomagnetic Depth Sounding, Induction Vectors and Coast-Effect (Jones, A.G.) 23

Secular Variation of Geomagnetic Field

- Impulse in Global Geomagnetic "Secular Variation", 1977–1979 (Nevanlinna, H., Sucksdorff, C.) 68

Seismic Anisotropy

- Anisotropy of V_p and V_s in Rock-Forming Minerals (Babuška, V.) 1

Seismics

- Two Methods of Solving the Linearized Two-Dimensional Inverse Seismic Kinematic Problem (Novotný, M.) 7
- Reflection and Transmission of Love Channel Waves at Coal Seam Discontinuities Computed with a Finite Difference Method (Korn, M., Stöckl, H.) 171
- Numerical Simulation of Vertical Seismic Profiling (Temme, P., Müller, G.) 177

Seismology

- The Romanian Earthquake of March 4, 1977 – III. Improved Focal Model and Moment Determination (Räkers, E., Müller, G.) 143

Solar Particle Events

- Low Energy Solar Particles Observed During the Pre-Phase of Large Particle Events (Kirsch, E., et al.) 45

Theoretical Seismograms

- Reflection and Transmission of Love Channel Waves at Coal Seam Discontinuities Computed with a Finite Difference Method (Korn, M., Stöckl, H.) 171
- Numerical Simulation of Vertical Seismic Profiling (Temme, P., Müller, G.) 177

Thermosphere

- Thermospheric Winds During the Energy Budget Campaign: Ground-Based Fabry-Perot Observations Supported by Dynamical Simulations with a Three-Dimensional, Time-Dependent Thermospheric Model (Rees, D., et al.) 202
- Zonal Meteor Wind Observations at Budrio, Italy (Cevolani, G., Dardi, D.) 212

Winds in the Upper Atmosphere

- Thermospheric Winds During the Energy Budget Campaign: Ground-Based Fabry-Perot Observations Supported by Dynamical Simulations with a Three-Dimensional, Time-Dependent Thermospheric Model (Rees, D., et al.) 202
- Zonal Meteor Wind Observations at Budrio, Italy (Cevolani, G., Dardi, D.) 212

In Memoriam 123*Correction* 124*Book Review* 125

Journal of Geophysics

Zeitschrift für Geophysik

Volume 50 Number 1 1981

Review Article

V. Babuška

Anisotropy of v_p and v_s in Rock-Forming Minerals 1

Original Investigations

M. Novotný

Two Methods of Solving the Linearized Two-Dimensional Inverse Seismic Kinematic Problem 7

A. Schult, A.G. Hussain, H.C. Soffel

Palaeomagnetism of Upper Cretaceous Volcanics and Nubian Sandstones of Wadi Natash, SE Egypt and Implications for the Polar Wander Path for Africa in the Mesozoic 16

A.G. Jones

Geomagnetic Induction Studies in Scandinavia – II. Geomagnetic Depth Sounding, Induction Vectors and Coast-Effect 23

A. Ádám, F. Márcz, J. Verö, Á. Wallner, G. Duma, R. Gutdeutsch

Magnetotelluric Sounding in the Transitional Zone Between the Eastern Alps and Pannonian Basin 37

E. Kirsch, N. Martinic, E. Keppler

Low Energy Solar Particles Observed During the Pre-Phase of Large Particle Events 45

W.G. Tighe, G. Rostoker

Characteristics of Westward Travelling Surges During Magnetospheric Substorms 51

Short Communications

H. Nevanlinna, C. Sucksdorff

Impulse in Global Geomagnetic "Secular Variation", 1977–1979 68

A. Ferrari, E. Trussoni

Turbulent Acceleration of Electrons in Extended Extragalactic Radio Sources 70

Indexed in Current Contents

90.00
8° Z. Nat. 2148

(2 4137)

Niedersächsische Staats-
Universitätsbibliothek
Göttingen

26. Okt. 1981

g4



Springer International

Journal of Geophysics – Zeitschrift für Geophysik

Edited for the Deutsche Geophysikalische Gesellschaft by W. Dieminger and J. Untiedt

This journal was founded by the Deutsche Geophysikalische Gesellschaft on the initiative of L. Mintrop in 1924 as the Zeitschrift für Geophysik and edited by G. Angenheister from Vol. 1–18 (1944). It reappeared in 1954 edited by B. Brockamp from Vol. 19–26 (1960), and edited by W. Dieminger from Vol. 27–33 (1961), and from Vol. 34 (1968) edited by W. Dieminger and J. Untiedt. After Vol. 40 (1970) the title was changed to Journal of Geophysics – Zeitschrift für Geophysik.

Published: Vols. 19–39 by Physica-Verlag, Würzburg, from Vol. 40 by Springer-Verlag GmbH & Co. KG Berlin, Heidelberg, New York.

Manuscripts may be addressed to any of the Editors. For addresses see last cover page. Manuscripts should conform with the journal's accepted practice as described in the Instructions to Authors.

The Journal accepts

- Review Articles (invited by the Editors)
- Original Papers
- Short Communications
- Letters to the Editors
- Book Reviews

in the field of Geophysics and Space Physics.

Copyright

It is a fundamental condition that submitted manuscripts have not been published and will not be simultaneously submitted or published elsewhere. By submitting a manuscript, the authors agree that the copyright for their article is transferred to the publisher if and when the article is accepted for publication. The copyright covers the exclusive rights to reproduce and distribute the article, including reprints, photographic reproductions, microform or any other reproductions of similar nature, and translations.

Photographic reproduction, microform, or any other reproduction of text figures, or tables from this journal is prohibited without permission obtained from the publisher.

The use of general descriptive names, trade names, trade marks, etc., in this publication, even if the former are not specifically identified, is not to be interpreted as exempt from the relevant protective laws and regulations and may accordingly be used freely by anyone.

Special Regulations for the USA

The Article Fee Code on the first page of an article in this journal indicates the copyright owner's consent that in the USA copies may be made for personal or internal use, provided the stated fee for copying beyond that permitted by Section 107 or 108 of the United States Copyright Law is paid through the **Copyright Clearance Center, Inc., 21 Congress Street, Salem, Mass. 01970, USA.**

If a code does not appear, copies of the article may be made without charge, provided permission is obtained from the publisher.

The copyright owner's consent does not extend to copying for general distribution, for promotion, for creating new works, or for resale. Specific written permission must be obtained from the publisher for such copying.

Subscription Information

Volumes 49+50 (3 issues each) will appear in 1981. The price of a volume is DM 158,00 or \$93.00. Prices for backvolumes are available on request. In addition, we also offer microform editions in 16 mm and 35 mm microfilm and microfiche. Correspondence concerning subscriptions should be addressed to the publisher.

Changes of Address: Allow six weeks for all changes to become effective. All communications should include both old and new addresses (with Zip Codes) and should be accompanied by a mailing label from a recent issue.

Members. Members of the Deutsche Geophysikalische Gesellschaft are entitled to purchase the Journal for their own use at a privilege price of DM 98,00 payable with the Membership dues. Orders should be sent to the Society's office at the following address: Postfach 230, D-3392 Clausthal-Zellerfeld.

North America. Annual subscription rate: \$195.00, including postage and handling. Subscriptions are entered with prepayment only. Orders should be addressed to: Springer-Verlag New York Inc., Journal Sales Department Secaucus, 44 Hartz-Way, Secaucus, NJ 07094, USA, Tel. (201) 348-4033, Telex 00 23-125994.

All Other Countries. Annual subscription rate: DM 316,00, plus postage and handling. Airmail delivery on request only. For Japan, postage (Surface Airmail Lifted) and handling is DM 36.00. Orders can either be placed with your bookdealer or sent directly to: Springer-Verlag, Heidelberger Platz 3, D-1000 Berlin 33, Tel. (030) 8207-1, Telex 01-83319.

Offices

Springer-Verlag, Heidelberger Platz 3, D-1000 Berlin 33, Tel. (030) 8207-1, Telex 01-83319.

Springer-Verlag, Journal Production Department II, Postfach 105280, D-6900 Heidelberg 1, Tel. (06221) 487-1, Telex 04-61690.

Springer-Verlag New York Inc., 175 Fifth Avenue, New York, N.Y. 10010, USA, Tel. (212) 477-8200, Telex 0023-23223.

Responsible for Advertisements

G. Sternberg, Kurfürstendamm 237, D-1000 Berlin 15, Tel. (030) 8821031, Telex 01-85411.

Printed in Germany by
Universitätsdruckerei H. Stürtz AG, Würzburg

© the Deutsche Geophysikalische Gesellschaft, Clausthal-Zellerfeld, 1981



Springer International

Review Article

Anisotropy of v_P and v_S in Rock-Forming Minerals

V. Babuška

Geophysical Institute, Czechoslovak Academy of Science, 14131 Praha 4 – Spořilov, Czechoslovakia

Abstract. The compressional and shear wave anisotropy coefficients of 33 minerals of igneous and metamorphic rocks were calculated from published elastic constants and tabulated together with the orientation of velocity extremes in single crystals. The most abundant minerals of crustal crystalline rocks – alkali feldspars, plagioclases, quartz, micas and hornblende – have higher anisotropy coefficients than the upper mantle minerals – olivine, pyroxenes and garnets. Due to the orientation of mineral grains and their velocity extremes in a stress field, however, the olivine-rich ultramafites belong to the most anisotropic rocks and, in contrast, the crack-free anisotropy of crustal crystalline rocks is generally low, with the exception of metamorphic rocks rich in micas, hornblende and calcite.

Key words: Elastic wave velocities – Anisotropy coefficients – Rock-forming minerals – Preferred orientation

Introduction

Elastic anisotropy, usually characterized by differences in compressional wave velocities, is one of the physical parameters which reflects the inner structure of rocks very sensitively. The large-scale seismic anisotropy of the continental upper mantle (Bamford 1977) and some regions of the Earth's crust (Burdock 1980) may yield important information on directions and intensity of the stress fields and geodynamic processes acting in geological history, providing the orientation of the structural elements, especially the preferred orientation of minerals at depth, is resolved from the in situ seismic observations.

There are, generally, only vague ideas among geophysicists about the degree of elastic anisotropy of the individual minerals. The purpose of this paper is, therefore, to calculate the coefficients of elastic anisotropy of the main rock-forming minerals from the elastic constants published by different authors and to discuss the possible elastic anisotropy and the preferred orientation of minerals in a stress field.

Methods of Investigation of Elastic Parameters

The experimental methods are either static or dynamic. The static methods are performed under isothermal conditions and are based on observations of static deformations (bending, twisting, compressing) of oriented crystals which mostly have the shape of a bar or a plate. The number of measurements needed to determine the elastic constants depends on the symmetry of crystals (Nye 1960). The accuracy of the static methods is given by the accuracy of determining small deformations (optical, electrical, mechanical methods) and by the size of the specimens. At present static methods are used mainly for the investigation of metals, glasses and construction materials which provide homogeneous samples of larger size.

The dynamic methods utilize high-frequency vibrations in

crystals. Due to the very high frequencies, the process itself, as well as the elastic constants obtained, can be treated as adiabatic. Ide (1935) introduced the dynamic method of resonance vibrations of cylindrical specimens. The principal modes of vibration are extensional, flexural, and torsional. Recent sophisticated techniques can be used to determine the elastic properties of spherical or parallel-sided specimens as small as 1 mm (Goto et al. 1976). Furthermore, only one piece of properly shaped specimen is sufficient for determining all the elastic parameters, because complete information is included in the resonance spectrum of a specimen. Brillouin scattering measurements yield elastic properties of crystals even smaller than 0.1 mm (Weidner et al. 1975). The technique, however, has only a limited usefulness since the orientation of crystals is not known.

By pulse-transmission methods the transit time of a high-frequency pulse through a specimen is measured in special directions. The pulses, of frequencies from 50 kHz to 10 MHz (Anderson and Liebermann 1966) are generated by crystal or ceramic transducers attached to the specimen. The pulse-transmission technique is probably also the most widely used method of determining compressional and shear wave velocities in specimens of minerals and rocks at high pressures and temperatures. Another method of measuring sound velocity in crystals is the ultrasonic interferometry method. In this method internal reflections of the same wave train are made to interfere so that a pseudo-resonance can be achieved by suitably controlling the wavelength of the imposed pulse. A more detailed account of the methods can be found in Anderson and Liebermann (1966), and Belikov et al. (1970).

The choice of the method of investigation depends mainly on the size and the quality of the specimen and the possibilities for experimental work. Large crystals very often contain various inclusions, cracks and pores, and the results are then less reliable than with relatively small but gem-quality crystals.

Anisotropy of Elastic Wave Velocities

Table 1 contains the elastic parameters of rock-forming minerals – the maximum and the minimum velocities of compressional and shear waves, as well as the coefficients of elastic anisotropy. The elastic constants can be found in the papers referenced, or, in the handbook by Simmons and Wang (1971), for most of the listed minerals. The coefficient $k = (v^{\max} - v^{\min})/v^{\text{mean}}$, where v^{mean} is an average value of the maximum and the minimum velocities. The corresponding directions of wave propagation and the polarization directions of the shear waves are given by Miller indices and those directions which cannot be defined by simple indices, are determined by the angles φ and ϑ (see Fig. 1). The original data are definitely of varying quality, depending mainly on the quality of the investigated crystals and the method of measurement. As mentioned in the previous section, the accurate determination of the elastic parameters requires crystals of a reasonable size. This is one of the reasons,

Table 1. Elastic wave velocities (v_p , v_s) and coefficients of elastic, anisotropy (k) in rock-forming minerals

Mineral	Symmetry	Density ($\text{g} \cdot \text{cm}^{-3}$)	Compressional waves					
			v_p^{\max} (km/s)	Propagation direction	v_p^{\min} (km/s)	Propagation direction	v_p^{mean} (km/s)	$k\%$
Ortho- and Ring Silicates								
Olivine (Fo 92. 72)	Orthorhombic	3.31	9.89	[100]	7.72	[010]	8.81	24.6
Zircon	Tetragonal	4.70	9.00	[001]	7.41	[010]	8.21	19.4
Garnets								
Pyrope (PY-1)	Cubic	3.705	8.90	[001]	8.85	$\vartheta = 55^\circ$ $\varphi = 45^\circ$	8.88	0.6
Almandine (AL-4)		3.930	8.76	[001]	8.73	$\vartheta = 55^\circ$ $\varphi = 45^\circ$	8.75	0.3
Spessartine (SP-1)		4.172	8.55	[001]	8.54	$\vartheta = 55^\circ$ $\varphi = 45^\circ$	8.55	0.1
Grossular (GR-1)		3.659	8.95	[001]	8.74	$\vartheta = 55^\circ$ $\varphi = 45^\circ$	8.85	2.4
Kyanite	Triclinic	3.62	9.88	[001]	9.00	[100]	9.44	9.3
Topaz	Orthorhombic	3.52	10.10	$\vartheta = 90^\circ$ $\varphi = 56.25^\circ$	9.15	[001]	9.63	9.9
Staurolite	Monoclinic	3.369	10.09	[100]	6.61	[001]	8.35	41.7
Epidote	Monoclinic	3.40	8.38	[010]	7.24	[011]	7.81	14.6
Beryl	Hexagonal	2.64	10.61	[100] $\vartheta = 94^\circ$	9.76	$\vartheta = 36^\circ$ $\varphi = 0^\circ$	10.19	8.3
Tourmaline	Trigonal	3.10	9.38	$\varphi = 33.75^\circ$	7.29	[001]	8.34	25.1
Chain Silicates								
Pyroxenes								
Bronzite	Orthorhombic	3.34	8.30	[100]	7.04	[010]	7.67	16.4
Diopside	Monoclinic	3.31	8.60	[001]	6.94	[101]	7.77	21.4
Hedenbergite		3.41	6.55	[001]	6.26	[010]	6.41	4.5
Augite		3.32	8.36	[101]	6.81	[010]	7.59	20.4
Aegirine		3.50	8.21	[001]	6.75	[101]	7.48	19.5
Hornblende	Monoclinic	3.15	8.13	[001]	6.18	[101]	7.16	27.2
Sheet Silicates								
Muscovite	Monoclinic	2.79	8.06	[110]	4.44	[001]	6.25	57.9
Biotite		2.89	7.80	[010]	4.21	[001]	6.01	59.7
Framework Silicates								
Feldspars								
Microcline	Triclinic	2.56	8.15	[010]	5.10	[100]	6.63	46.0
Orthoclase	Monoclinic	2.54	7.64	[010]	4.76	[101]	6.20	46.4
Albite (An 9)	Triclinic	2.61	7.26	[010]	5.31	[101]	6.29	31.0
Oligoclase (An 29)		2.64	7.55	[010]	5.50	[101]	6.53	31.4
Labradorite (An 53)		2.68	7.80	[010]	6.06	[100]	6.93	25.1
Anorthite		2.76	8.61	[010]	6.01	[101]	7.31	35.6
Quartz	Trigonal	2.66	7.00	$\vartheta = 130^\circ$ $\varphi = 90^\circ$	5.36	$\vartheta = 72^\circ$ $\varphi = 90^\circ$	6.18	26.5
Nepheline	Hexagonal	2.62	7.12	[001]	5.61	[100]	6.37	23.7
Non-Silicates								
Rutile	Tetragonal	4.264	10.65	[001]	8.00	[100]	9.33	28.4
Spinel (MgO · 3.5 Al ₂ O ₃)	Cubic	3.63	10.31	[101]	9.10	[001]	9.71	12.5
Magnetite		5.18	7.46	$\vartheta = 55^\circ$ $\varphi = 45^\circ$	7.29	[001]	7.38	2.3
Pyrite (2)	Trigonal	5.013	8.40	[001]	7.78	[110]	8.09	7.7
Calcite		2.717	7.55	$\vartheta = 70^\circ$ $\varphi = 90^\circ$	5.43	[001]	6.49	32.7

$$k = \frac{v_p^{\max} - v_p^{\min}}{v_p^{\text{mean}}} \cdot 100\%$$

Shear waves								Source of data
v_s^{\max} (km/s)	Propagation direction	Polarization direction	v_s^{\min} (km/s)	Propagation direction	Polarization direction	v_s^{mean} (km/s)	$k\%$	
5.53	$\vartheta=45^\circ$ $\varphi=0^\circ$	$\vartheta=135^\circ$ $\varphi=0^\circ$	4.42	[010]	[100]	4.98	22.3	Kumazawa and Anderson 1969
4.87	[010]	[001]	2.94	[110]	[110]	3.91	49.4	Ryzhova et al. 1966
4.96	[101]	[10 $\bar{1}$]	4.91	[001]	[110]	4.94	1	Babuška et al. 1978
4.96	[110]	[1 $\bar{1}$ 0]	4.91	[001]	[1 $\bar{1}$ 0]	4.94	1	
4.82	[110]	[1 $\bar{1}$ 0]	4.80	[001]	[110]	4.81	0.4	
5.45	[101]	[10 $\bar{1}$]	5.20	[001]	[110]	5.33	4.7	
6.76	[010]	[001]	4.24	[001]	[100]	5.50	45.8	Belikov et al. 1970
6.15	[001]	[100]	5.16	[$\bar{1}$ 10]	$\vartheta=90^\circ$ $\varphi=41.25^\circ$	5.66	17.5	Voigt 1928
5.29	$\vartheta=118^\circ$ $\varphi=45^\circ$	$\vartheta=62^\circ$ $\varphi=315^\circ$	3.23	$\vartheta=140^\circ$ $\varphi=45^\circ$	$\vartheta=50^\circ$ $\varphi=86.7^\circ$	4.26	48.4	Aleksandrov and Ryzhova 1961 a
5.11	[011]	[011]	3.39	[001]	[010]	4.25	40.5	Ryzhova et al. 1966
6.25	$\vartheta=44^\circ$ $\varphi=0^\circ$	$\vartheta=136.7^\circ$ $\varphi=0^\circ$	5.34	[001]	[110]	5.80	15.7	Hearmon 1956
6.14	$\vartheta=82^\circ$ $\varphi=33.75^\circ$	$\vartheta=88.5^\circ$ $\varphi=124^\circ$	4.55	$\vartheta=172^\circ$ $\varphi=33.75^\circ$	$\vartheta=389.4^\circ$ $\varphi=119.4^\circ$	5.35	29.7	
4.99	[010]	[001]	4.27	$\vartheta=90^\circ$ $\varphi=45^\circ$	$\vartheta=90^\circ$ $\varphi=135^\circ$	4.63	15.6	Kumazawa 1969
4.83	[011]	[0 $\bar{1}$ 1]	3.94	[110]	[1 $\bar{1}$ 0]	4.39	20.3	Aleksandrov et al. 1963 Volarovich et al., 1975
4.72	[011]	[0 $\bar{1}$ 1]	3.81	[100]	[001]	4.27	21.3	Aleksandrov et al. 1963
4.65	[011]	[0 $\bar{1}$ 1]	3.48	[010]	[100]	4.07	28.7	
4.60	[011]	[0 $\bar{1}$ 1]	3.37	[001]	[100]	3.99	30.8	Aleksandrov and Ryzhova 1961 b
5.01	[010]	[100]	2.03	[001]	[100]	3.52	84.7	Aleksandrov and Ryzhova 1961 c
5.06	[010]	[100]	1.34	[010]	[001]	3.20	116.3	Belikov et al. 1970
4.96	[011]	[01 $\bar{1}$]	2.14	[010]	[001]	3.55	79.4	Aleksandrov and Ryzhova 1962 Ryzhova and Aleksandrov 1965
4.45	[011]	[01 $\bar{1}$]	2.33	[001]	[010]	3.39	62.5	
4.63	[011]	[01 $\bar{1}$]	2.59	[001]	[010]	3.61	56.5	Ryzhova 1964
4.70	[011]	[01 $\bar{1}$]	2.65	[001]	[010]	3.68	55.7	
4.76	[011]	[01 $\bar{1}$]	2.70	[001]	[010]	3.73	55.2	
4.96	[011]	[01 $\bar{1}$]	2.91	[010]	[100]	3.94	52.0	Aleksandrov et al. 1974
5.06	[100]	$\vartheta=147^\circ$ $\varphi=284.5^\circ$	3.35	[100]	$\vartheta=122^\circ$ $\varphi=90^\circ$	4.21	40.1	Huntington 1958
3.92	[011]	[01 $\bar{1}$]	2.83	[100]	[010]	3.38	32.2	Ryzhova and Aleksandrov 1962
6.74	[100]	[010]	3.32	[110]	[110]	5.03	68.0	Birch 1960
6.61	[001]	[1 $\bar{1}$ 0]	4.50	[101]	[10 $\bar{1}$]	5.56	37.9	Verma 1960
4.29	[001]	[1 $\bar{1}$ 0]	4.06	[101]	[10 $\bar{1}$]	4.18	5.5	Hearmon 1956
5.66	[110]	[$\bar{1}$ 10]	4.70	[001]	[110]	5.18	18.5	Aleksandrov and Ryzhova 1961 d
4.77	$\vartheta=126^\circ$ $\varphi=90^\circ$	[100]	2.66	$\vartheta=36^\circ$ $\varphi=90^\circ$	[100]	3.72	56.7	Hearmon 1956

why the elastic constants of single crystals are lacking for important minerals such as chlorite and serpentine.

The minerals in Table 1 are arranged from ortho- and ring silicates, via chain silicates, sheet silicates, and framework silicates to non-silicates, as is usual in systematic mineralogy. Of the group of *ortho- and ring silicates*, olivine and garnets are among the most abundant rock-forming minerals in the lithosphere; they are important constituents of the upper mantle rocks. In contrast to olivine, garnets show the lowest anisotropy of all the minerals listed. Due to a frequent and well-developed preferred orientation of olivine, the olivine-rich ultramafites usually have a relatively high degree of velocity anisotropy, of about 10% for the compressional waves (Babuška 1972a). An admixture of garnets in these ultramafites tends to decrease their elastic anisotropy.

In the important group of *chain silicates* the published data enable the elastic anisotropies to be calculated for only five pyroxenes and one amphibole (hornblende). As regards *sheet silicates*, the elastic constants of single crystals can be determined reliably only for the main members of the mica group. The extremely high values of the coefficients of anisotropy in micas are explained by their structure in which a layer of octahedrally coordinated cations is sandwiched between two identical layers of linked (Si, Al) O_4 tetrahedra (Deer et al. 1966). A remarkable anisotropy of the physical properties, including a perfect basal cleavage, is a consequence of the great difference in the strength of interatomic bonds in the directions parallel and perpendicular to the layered atomic structure. The perfect cleavage, however, may decrease the accuracy of determining the elastic parameters and may cause a considerable scatter of data.

Framework silicates include the most abundant minerals of the Earth's crust – feldspars and quartz. Among feldspars the potassium members belong, along with the mica group, to the most anisotropic minerals. In the plagioclase feldspar series the coefficients of *S*-wave anisotropy are substantially higher than the coefficients for the *P* waves. The elastic anisotropy of quartz, though lower in comparison with feldspars, still classes it with the rock-forming minerals with high coefficients of anisotropy. This is in contrast to the generally very low coefficients of the crack-free elastic anisotropy of the granitic rocks (Birch 1961; Babuška et al. 1977).

Among *non-silicates* probably only calcite, in some regions of the uppermost crust, and spinel in the spinel peridotites may sporadically form substantial volumetric parts of the lithosphere. The high elastic anisotropy of calcite, together with its preferred orientation, results in a high velocity anisotropy for marbles (Babuška 1968). Due to the low *P*-wave anisotropy of spinel, any admixture of the mineral in peridotite or lherzolite would decrease the anisotropy of the rocks.

With the exception of two pyroxenes (bronzite, diopside) and olivine, all the other minerals listed in Table 1 have coeffi-

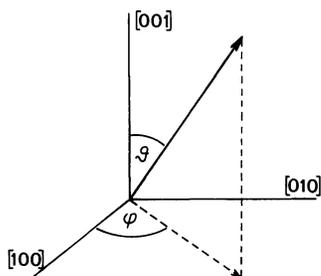


Fig. 1. System of determination of the propagation and the polarization directions given in Table 1

icients of anisotropy of the shear wave velocities higher than the coefficients of the compressional velocities. The largest differences in the *P*- and *S*-wave anisotropy coefficients are observed for micas, feldspars and several other minerals (kyanite, spinel, epidote, rutile, zircon).

Discussion

Let us take from Table 1 the most abundant minerals or mineral groups of crystalline rocks and arrange them in a series according to the values of the coefficient of compressional wave anisotropy. This results in the following sequence: micas ($k=59\%$), alkali feldspars (46%), calcite (33%), plagioclase (31%), hornblende (27%), quartz (26%), olivine (25%), pyroxenes (16%), garnets (1%). There is no significant change in this sequence if we consider the coefficients of the shear-wave anisotropy. It is remarkable that the last three members of the sequence are the main components of the upper mantle rocks. The olivine-rich rocks like dunite and peridotite, however, belong to the most anisotropic rocks. In contrast, granite or granodiorite, being composed of highly anisotropic components and often carrying signs of tectonites, including mineral-preferred orientation, generally have low coefficients of crack-free anisotropy (Babuška et al. 1977).

What then are the main factors which determine the degree of the elastic anisotropy of crystalline rocks? There are three: the values of the anisotropy coefficients of the constituent minerals and their volume in the aggregate; the character and degree of the preferred orientation of minerals and other fabric elements (e.g. oriented micro-discontinuities, alternating layers); the orientation of extremes of the elastic wave velocities in single crystals with respect to the directions which are significant for their orientation in a stress field.

Besides garnets, the anisotropy of which is negligible, all other very abundant minerals named above can contribute, according to their single crystal anisotropy, to the elastic anisotropy of crystalline rocks.

Experiments on the preferred orientation of *mica* carried out by Tullis (1971), showed that during crystallization or recrystallization the (001) plane mean orientation is always normal to the direction of the greatest shortening which, in most cases, coincides with the maximum compressive stress σ_1 . The compressional velocity minimum in micas is perpendicular to the (001) plane and the velocity maximum is parallel to this plane (see Table 1). Thanks to these orientations and to the extremely high single crystal anisotropy, micas contribute substantially to the elastic anisotropy of aggregates. This is true of a number of metamorphic rocks such as mica schists and gneisses.

Deformation of *feldspars* and the question of the origin of their preferred orientation is not a simple problem and the data are insufficient in this respect. Borg and Heard (1970) in their experimental work on deformation of plagioclases found that, if it can be ascertained that the albite and pericline twins are mechanical in origin, it is theoretically possible to deduce the orientation of the principal stresses which produced them. It is, however, difficult to prove the mechanical origin of twin lamellae in naturally deformed, plagioclase-bearing rocks. In the case of such mechanical twins the direction of maximum stress σ_1 is at acute angles (close to 45°) to the crystallographic directions $[0\bar{1}0]$ and $[001]$. However, the maximum velocity in all feldspars is in the $[010]$ direction, the velocity minimum in either $[100]$ or $[101]$ directions. Therefore, we cannot expect any alignment of the velocity extremes and the resulting *P*-velocity anisotropy by this mechanism of preferred orientation.

Experimental deformation of *calcite* aggregates was investigated in a number of studies which can be summarized as fol-

lows: in syntectonic recrystallization the c axes tend to be subparallel to the principal stress σ_1 , or, the c axes concentrations are inclined at 20° – 30° to the σ_1 direction (Turner and Weiss 1963). Because the c axis is the direction of the velocity minimum (also the direction of high compressibility, Birch 1966) and the P -wave velocities in the x – y plane are very close to the maximum velocity (deWitte 1962), the directions of the velocity extremes are often preferentially oriented, which leads to the commonly observed high elastic anisotropy of marbles.

The preferred orientation of coarse grained hornblende in metamorphic rocks is often observed. Schwerdtner (1964) found that the hornblende orientation in a deformed gneiss from Norway corresponds to a pattern predicted by the theory of Kamb (1959), i.e., the c axes tend to be oriented perpendicular to the axis of greatest compressional stress. The c axis is also the direction of maximum P -wave velocity and thus the oriented hornblende grains can contribute to an increase in the elastic anisotropy of rocks. Both in hornblende and in micas the external shape plays an important role in orientation and the resulting anisotropy of metamorphic complexes (Burdock 1980).

The investigation of the preferred orientation of quartz aggregates shows a complex picture of the different types of orientation developed under different experimental conditions. At lower temperatures or faster strain rates a maximum of the c axes of the α -quartz develops parallel to the compression direction σ_1 (c -maximum fabric), and at higher temperatures or slower strain rates a concentration of the normal to r ($10\bar{1}1$) develops parallel to σ_1 (r -maximum fabric, Green et al. 1970). At intermediate temperatures and strain rates, the preferred orientation is a combination of both fabric types (composite fabrics). According to Green et al. (1970) at high temperatures and low strain rates, orthorhombic strain due to temperature gradients produces 'crossed girdles' of c -axes intersecting normal to σ_1 . Taking into consideration these possibilities for the quartz orientations, which are, of course, still incomplete, and the fact that the P -wave velocity extremes in quartz do not coincide either with the c -axis or with the $[10\bar{1}1]$ direction, we cannot expect any pronounced elastic anisotropy of rocks due to the orientation of quartz grains.

Olivine in syntectonically recrystallized aggregates tends to be oriented with the $[010]$ direction parallel to σ_1 and $[100]$ parallel to σ_3 . This orientation conforms both to Kamb's thermodynamic theory (Kamb 1959) and the experiments carried out on natural dunites by Avé Lallemant and Carter (1970). Also other mechanisms like plastic flow or shear along transverse faults, which were discussed, e.g., by Nicolas and Poirier (1976), lead to a distinct preferred orientation of olivine crystals. Olivine has a P -velocity anisotropy of about 25% with the maximum velocity along the $[100]$ direction and the minimum value along $[010]$. These directions tend to be oriented along the minimum and the maximum stresses, respectively. Such favourable conditions result in the fact that a significant elastic anisotropy of olivine-rich ultramafites is a rule rather than an exception.

The experiments of Carter et al. (1972) on syntectonic recrystallization of orthopyroxene (enstatite) showed that recrystallized grains have preferred orientations with $[010]$ oriented parallel to σ_1 and the remaining principal axes arranged in girdles in the σ_2 – σ_3 plane. The fabric is similar to that of olivine. The orientation of velocity extremes in orthopyroxene (Bronzite, see Table 1) is conformable to olivine. Therefore, provided olivine and orthopyroxene recrystallized in the same stress field, the anisotropies due to both minerals are additive for the anisotropy of rocks in bulk (Kumazawa et al. 1971). However, Helmstaedt et al. (1972) observed in a nodule of Iherzolite that the crystallo-

graphic axes a of enstatite and those of olivine form maxima perpendicular to each other, these corresponding to the contrasting non-additive anisotropies of both minerals. As the preferred orientations of orthopyroxene are often weaker than those of olivine (Nicolas and Poirier 1976; Babuška 1972b) and the anisotropy coefficients of orthopyroxene are smaller, the resulting anisotropy of orthopyroxene-rich rocks is relatively low.

Kumazawa et al. (1971) observed that, despite a significant preferred orientation of clinopyroxenes, the velocity anisotropy of eclogites from diatremes of the Colorado plateau is considerably smaller than the anisotropy of peridotites. Although the coefficients of anisotropy of the omphacitic pyroxene have not been determined yet, we can suppose a value of about 20%, as for augite. On the basis of the thermodynamic theory of Kumazawa (1963), Kumazawa et al. (1971) assume that the preferred orientation of clinopyroxenes is controlled by their linear compressibility. Since in clinopyroxenes the directions of extreme linear compressibilities are different from the directions of the compressional velocity extremes, a high velocity anisotropy in eclogites cannot be expected.

Although the shear-wave anisotropy of most rock-forming minerals is larger than the compressional anisotropy, the shear anisotropy of aggregates is a complicated problem, because of the great differences in the shear velocities with different polarizations. For example, in olivine the maximum difference due to polarization for a single direction or propagation is as much as 0.46 km/s (Kumazawa and Anderson 1969) and in dunite, which is composed of 98% of olivine, the difference for a single direction reaches 0.38 km/s (Babuška 1972b), this means 10% and 8%, respectively, of the mean velocity in the corresponding directions. The shear-wave birefringence was also observed in a number of metamorphic rocks (Tilmann and Bennet 1973) and it is believed that polarization studies using three-component records will show diagnostic anomalies in the presence of in situ seismic anisotropy (Keith and Crampin 1977).

Conclusions

According to the elastic anisotropy coefficients the most abundant rock-forming minerals of igneous and metamorphic rocks form the following sequence (the values in brackets are the compressional and the shear-wave anisotropy coefficients in per cent, respectively): micas (59, 100), alkali feldspars (46, 71), calcite (33, 57), plagioclases (31, 55), hornblende (27, 31), quartz (26, 40), olivine (25, 22), pyroxenes (16, 21), garnets (1, 2).

Feldspars, quartz and clinopyroxene, the minerals which are characterized by high anisotropy coefficients, form aggregates with generally low anisotropy, although their grains are often preferentially oriented. This is probably because the velocity extremes in these minerals differ from the directions which are decisive for their orientation in a stress field. The velocity extremes are thus not aligned according to directions of the maximum and the minimum stresses, as observed, e.g., in the aggregates of olivine. Consequently, this means that we cannot expect a high velocity anisotropy in crustal igneous rocks. On the other hand, metamorphites rich in micas, hornblende and calcite have high coefficients of elastic anisotropy.

Acknowledgements. I would like to thank Dr. J. Šílený for his assistance in computing the spatial distribution of velocities in minerals and Dr. J. Vaněk for his critical review of the manuscript.

References

- Aleksandrov, K.S., Alchikov, V.V., Belikov, B.P., Zaslavskii, B.I., Krupnyi, A.I.: Velocities of elastic waves in minerals at atmospheric

- pressure and increasing of precision of elastic constants by means of EVM (in Russian). *Izv. Acad. Sci. USSR, Geol. Ser. No. 10*, 15–24, 1974
- Aleksandrov, K.S., Ryzhova, T.V.: The elastic properties of crystals. *Sov. Phys. Crystallogr.* **6**, 228–252, 1961a
- Aleksandrov, K.S., Ryzhova, T.V.: The elastic properties of rock-forming minerals, I: pyroxenes and amphiboles (in Russian). *Izv. Acad. Sci. USSR, Geophys. Ser. No. 9*, 1339–1344, 1961b
- Aleksandrov, K.S., Ryzhova, T.V.: The elastic properties of rock-forming minerals, II: layered silicates (in Russian). *Izv. Acad. Sci. USSR, Geophys. Ser. No. 12*, 1799–1804, 1961c
- Aleksandrov, K.S., Ryzhova, T.V.: Moduli of elasticity of pyrite (in Russian). *Izv. Acad. Sci. USSR, Sibir. Branch, No. 6*, 43–47, 1961d
- Aleksandrov, K.S., Ryzhova, T.V.: The elastic properties of rock-forming minerals, III: feldspars (in Russian). *Izv. Acad. Sci. USSR, Geophys. Ser. No. 2*, 186–189, 1962
- Aleksandrov, K.S., Ryzhova, T.V., Belikov, B.P.: The elastic properties of pyroxenes (in Russian). *Sov. Phys. Crystallogr.* **8**, 738–741, 1963
- Anderson, O.L., Liebermann, R.C.: Sound velocities in rocks and minerals. VESIAC State-of-the-Art Report. Ann Arbor: Univ. of Michigan 1966
- Avé Lallemant, H.G., Carter, N.L.: Syntectonic recrystallization of olivine and modes of flow in the upper mantle. *Geol. Soc. Am. Bull.* **81**, 2203–2220, 1970
- Babuška, V.: Elastic anisotropy of igneous and metamorphic rocks. *Stud. Geophys. Geod.* **12**, 291–303, 1968
- Babuška, V.: Anisotropy of the upper mantle rocks. *Z. Geophys.* **38**, 461–467, 1972a
- Babuška, V.: Elasticity and anisotropy of dunite and bronzitite. *J. Geophys. Res.* **77**, 6955–6965, 1972b
- Babuška, V., Pros, Z., Franke, W.: Effect of fabric and cracks on the elastic anisotropy in granodiorite. *Publ. Inst. Geophys. Pol. Acad. Sci A-6 (117)*, 179–186, 1977
- Babuška, V., Fiala, J., Kumazawa, M., Ohno, I., Sumino, Y.: Elastic properties of garnet solid-solution series. *Phys. Earth Planet. Inter.* **16**, 157–176, 1978
- Bamford, D.: Pn velocity anisotropy in a continental upper mantle. *Geophys. J. R. Astron. Soc.* **49**, 29–48, 1977
- Belikov, B.P., Aleksandrov, K.S., Ryzhova, T.V.: Elastic properties of rock-forming minerals and rocks. Moscow: Nauka 1970
- Birch, F.: Elastic constants of rutile – A correction to a paper by R.K. Verma, Elasticity of some high-density crystals. *J. Geophys. Res.* **65**, 3855–3856, 1960
- Birch, F.: The velocity of compressional waves in rocks to 10 kilobars. Part 2. *J. Geophys. Res.* **66**, 2199–2224, 1961
- Birch, F.: Compressibility; elastic constants. In: *Handbook of Physical Constants revised edition*, Geol. Soc. Am. Mem. 97, S.P. Clark, Jr., ed.: pp. 97–173, New York: Geol. Soc. Am. 1966
- Borg, I.Y., Heard, H.C.: Experimental deformation of plagioclases. In: *Experimental and Natural Rock Deformation*, P. Paulitsch, ed.: pp. 375–403. Berlin Heidelberg New York: Springer 1970
- Burdock, J.H.: Seismic velocity structure of the metamorphic belt of central California. *Bull. Seismol. Soc. Am.* **70**, 203–221, 1980
- Carter, N.L., Baker, D.W., George R.P., Jr.: Seismic anisotropy, flow, and constitution of the upper mantle. In: *Flow and fracture of rocks*, Geophys. Monograph 16, H.C. Heard, I.Y. Borg, N.L. Carter, C.B. Raleigh, eds.: pp. 167–190, Washington D.C.: Am. Geophys. Union 1972
- Deer, W.A., Howie, R.A., Zussman, J.: An introduction to the rock-forming minerals. London: Longmans, Green and Co. Ltd 1966
- Goto, T., Ohno, I., Sumino, Y.: The determination of the elastic constants of natural almandine-pyrope garnet by rectangular parallel-piped resonance method. *J. Phys. Earth* **24**, 149–156, 1976
- Green, H.W., II, Griggs, D.T., Christie, J.M.: Syntectonic and annealing recrystallization of fine-grained quartz aggregates. In: *Experimental and Natural Rock Deformation*, P. Paulitsch, ed.: pp. 272–335, Berlin, Heidelberg: Springer-Verlag 1970
- Hearmon, R.F.S.: The elastic constants of anisotropic materials II. *Adv. Phys.* **5**, 523–382, 1956
- Helmstaedt, H., Anderson, O.L., Gavasci, A.T.: Petrofabric studies of eclogite, spinel-websterite, and spinel-lherzolite xenoliths from kimberlite-bearing breccia pipes in southeastern Utah and northeastern Arizona. *J. Geophys. Res.* **77**, 4350–4365, 1972
- Huntington, H.B.: The elastic constants of crystals. In: *Solid State Physics*, Vol. 7, F. Seitz, D. Turnbull, eds.: pp. 213–285, New York: Academic Press 1958
- Ide, J.M.: Some dynamic methods for determination of Young's modulus. *Rev. Sci. Instr.* **6**, 296–298, 1935
- Kamb, W.B.: Theory of preferred crystal orientation developed by crystallization under stress. *J. Geol.* **67**, 153–170, 1959
- Keith, C.M., Crampin, S.: Seismic body waves in anisotropic media: synthetic seismograms. *Geophys. J. R. Astron. Soc.* **49**, 225–243, 1977
- Kumazawa, M.: Fundamental theory on the nonhydrostatic thermodynamics and on the stability of mineral orientation and phase equilibrium. *J. Earth Sci., Nagoya Univ.* **11**, 145–217, 1963
- Kumazawa, M.: The elastic constants of single-crystal orthopyroxene. *J. Geophys. Res.* **74**, 5973–5980, 1969
- Kumazawa, M., Anderson, O.L.: Elastic moduli, pressure derivatives, and temperature derivatives of single-crystal olivine and single-crystal forsterite. *J. Geophys. Res.* **74**, 5961–5972, 1969
- Kumazawa, M., Helmstaedt, H., Masaki, K.: Elastic properties of eclogite xenoliths from diatremes of the east Colorado plateau and their implication to the upper mantle structure. *J. Geophys. Res.* **76**, 1231–1247, 1971
- Nicolas, A., Poirier, J.P.: Crystalline plasticity and solid state flow in metamorphic rocks. London: John Wiley and Sons, 1976
- Nye, J.F.: Physical properties of crystals. Oxford: Clarendon Press 1960
- Ryzhova, T.V.: Elastic properties of plagioclases (in Russian). *Izv. Acad. Sci. USSR, Geophys. Ser. No. 7*, 1049–1051, 1964
- Ryzhova, T.V., Aleksandrov, K.S.: The elastic properties of rock-forming minerals, IV: nepheline (in Russian). *Izv. Acad. Sci. USSR, Geophys. Ser. No. 12*, 1799–1801, 1962
- Ryzhova, T.V., Aleksandrov, K.S.: The elastic properties of potassium-sodium feldspars (in Russian). *Izv. Acad. Sci. USSR, Earth Phys. Ser. No. 1*, 98–102, 1965
- Ryzhova, T.V., Aleksandrov, K.S., Korobkova, V.M.: The elastic properties of rock-forming minerals. V: Additional data on silicates (in Russian). *Izv. Acad. Sci. USSR, Earth Phys. Ser. No. 2*, 63–65, 1966
- Schwerdtner, W.M.: Preferred orientation of hornblende in a banded hornblende gneiss. *Am. J. Sci.* **262**, 1212–1229, 1964
- Simmons, G., Wang, H.: Single crystal elastic constants and calculated aggregate properties: A handbook. Cambridge, Mass.: M.I.T. Press 1971
- Tilmann, S.E., Bennet, H.F.: Ultrasonic shear wave birefringence as a test of homogeneous elastic anisotropy. *J. Geophys. Res.* **78**, 2623–2629, 1973
- Tullis, T.: Experimental development of preferred orientations of mica during recrystallization. Ph.D. Thesis, Univ. of Calif., Los Angeles, 1971
- Turner, F.J., Weiss, L.E.: Structural analysis of metamorphic tectonites. New York: McGraw-Hill Book Comp. 1963
- Verma, R.K.: Elasticity of some high-density crystals. *J. Geophys. Res.* **65**, 757–766, 1960
- Voigt, W.: *Lehrbuch der Krystallophysik*. Leipzig: B.G. Teubner 1928
- Volarovich, M.P., Bayuk, E.I., Efimova, G.A.: Elastic properties of minerals at high pressures (in Russian). Moscow: Nauka 1975
- Weidner, D.J.: Elasticity of microcrystals. *Geophys. Res. Lett.* **2**, 189–192, 1975
- Witte, A.J. de: On elastic wave propagation in crystals with applications to calcite and quartz. *Eng. Exp. Station Techn. Rep. No. 4*, Univ. of Illinois 1962

Original Investigations

Two Methods of Solving the Linearized
Two-Dimensional Inverse Seismic Kinematic Problem

M. Novotný

Geofyzika, n.p. Brno, Ječná 29a, 612 46 Brno, Czechoslovakia

Abstract. In a laterally variable seismic medium, the slowness function $s(x, z)$ can be derived from travel-time curves of surface-to-surface refracted waves if the arrival times are interpreted in the form of a two-dimensional function. In the following article two methods are suggested for solving the inverse kinematic problem using a linearized formulation. The arrival times $t(p, q)$ are arranged according to increasing epicentral distance q , and p represents the x -coordinate of the point midway between source and receiver. The first method is based on the Fourier transform of the time function $t(p, q)$ in the variable p . This method can be applied to a system of travel-time curves with epicentral distances small in comparison to the length of the profile interval x_p , i.e., $q \ll x_p$. The second, grid method interprets recursively the arrival times $t(p_j, q_i)$ from the smallest epicentral distance to the largest one. The results of one interpretation step are the grid values of the slowness function $s(x, z)$ on one grid line $z = z_i$. In contrast to the method of Fourier coefficients the recursive grid method does not introduce any limitation of the shape of the solution region. Both methods have been tested using both theoretical and real data.

Key words: Inverse kinematic seismic problem - Laterally inhomogeneous media - Ray interpretation of travel-times

Introduction

The interpretation of travel-time curves of surface-to-surface refracted waves represents one possible method for estimating the velocity distribution in deep-seated structures. If the medium is assumed to be inhomogeneous in the vertical direction only, then the unknown velocity function $v(z)$ of depth can be obtained by the well-known Wiechert-Herglotz integral transformation of the travel-time curve of the surface-to-surface refracted wave $t(r)$. In the majority of computational algorithms based upon the Wiechert-Herglotz theory it is assumed that $t(r)$ is a one-valued function of the epicentral distance r with non-increasing derivative, even though, as early as 1932, the theory was generalized by Slichter (1932) to the case of loop-like travel-time curves occurring in a medium with strongly varying velocity gradient or even with velocity discontinuities. In this case the travel-time curve of the supercritically reflected wave must also be taken into account. The problem of the estimation of the function $v(z)$ in the presence of a low-velocity channel was studied in a paper by Gerver et al. (1966). Practical

difficulties associated with the travel-time curve ambiguity meant that a new theory was needed and this has recently appeared in the form of the tau method (Bessonova et al. 1974), sometimes called the method of extremal inversion (Kennett 1976). The tau method is based on the Legendre transform of the travel-time curve $t(r)$ that transforms the loop-like travel-time curve into the one-valued function $\tau(p)$ (Gerver et al. 1967). This approach also allows the setting of reasonable bounds to variations of the velocity function $v(z)$, deduced from the character of the discretely measured values $t(r)$. In particular, incompletely recorded travel-time curves, as well as the discontinuous ones that are met frequently in interpretation of deep seismic sounding data, can be more reliably interpreted by means of the tau method (Beránek et al. 1979).

Real seismic media are generally also inhomogeneous in the horizontal directions x and y . If the inhomogeneity is negligible in one direction (for instance y), it then makes sense to consider the two-dimensional inverse problem of finding the velocity distribution $v(x, z)$. It is now assumed that an adequately dense system of travel-time curves $t(x_S, x_R)$ is available with sources $S = (x_S, 0)$ and receivers $R = (x_R, 0)$ distributed uniformly over the entire interval under consideration $\langle 0, x_p \rangle$ along the profile (Fig. 1).

Instead of the variables x_S, x_R it is more advantageous to operate with other variables p, q defined by the relations

$$p = (x_S + x_R)/2, \quad q = |x_R - x_S|. \quad (1)$$

This new coordinate frame is also currently used in the CDP method of seismic reflection prospecting. The p variable represents the x coordinate of the midpoint between the source S and the receiver R . The variable q stands for the epicentral distance (offset of measurement). The function $t(p, q)$ then represents, for the constant offset variable q , an isoline of the so called special time field (Puzyrev et al. 1975). The use of the variables p, q is based on the equivalence of the travel-times of seismic waves between reciprocal points, i.e., $t(x_R, x_S) = t(x_S, x_R)$. Since the reciprocal measurements are often performed in practice, we remove this redundancy in the input data.

While methods of solution of the one-dimensional inverse problems are relatively well developed and can thus be immediately applied to solve practical problems (Bessonova et al. 1976; Kennett 1976), the study of multi-dimensional problems is still at the beginning. The inverse kinematic problem for more dimensions is, from the mathematical point of view, a classical ill-posed problem. Its solution has required a new

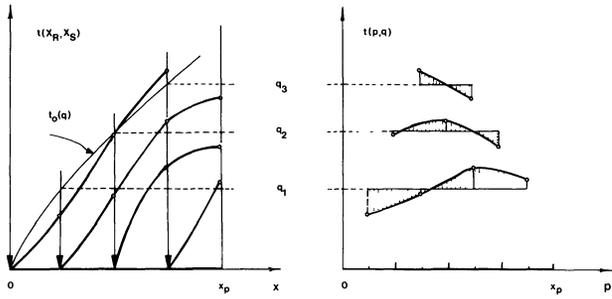


Fig. 1. Two representations of arrival times of surface-to-surface refracted waves. System of travel-time curves $t(x_R, x_S)$ with a constant coordinate x_S of the source S and a variable coordinate x_R of the receiver point R can be arranged into time isolines $t(p, q)$ with a constant offset of measurements $q = |x_S - x_R|$ and with the lateral variable $p = (x_R + x_S)/2$. Travel-time curve $t_0(q)$ is obtained as the average of all arrival times due to common offset q

approach to the question of solution stability (Romanov 1972). The first strict mathematical formulation of the two-dimensional inverse problem as a certain Cauchy problem for the non-linear differential equation of the Hamilton-Jacobi type has been presented by Belonosova and Alexeyev (1967). The travel-time curve of the surface-to-surface refracted wave $t(x_S, x_R)$ and its partial derivatives $\partial t/\partial x_S$, $\partial t/\partial x_R$ represent initial conditions. It has been shown by theoretical examples (Belonosova and Alexeyev 1967) that the finite-difference approximation yields a numerical solution to the problem with an accuracy of 5–10%.

Another approach (Romanov et al. 1978) is based on a combination of solutions of both the inverse and the direct problem. Both approaches require both the two dimensional travel-time curve $t(x_S, x_R)$ and its partial derivatives. However, commonly used systems of measurements do not yield a sufficiently dense set of data that would make possible a fair approximation of the required initial data. In particular, derivatives $\partial t/\partial x_S$ and $\partial t/\partial x_R$ derived from sparsely measured travel-times $t(x_S, x_R)$ would cause considerable inaccuracies in the results obtained by these methods. In such a case, the simplified linearized formulation of the inverse kinematic problem appears as one way out.

The linearization of the inverse problem in a multi-dimensional case has been suggested by Romanov (1972). It represents a particular generalized analogue of the linearized approach to solving one-dimensional problems (Johnson and Gilbert 1972).

Let us consider a two-dimensional problem. The following expression is valid for the travel-time of seismic energy propagated from point S to point R along trajectory $\Gamma(R, S)$ (Fig. 2):

$$t(R, S) = \int_{\Gamma(R, S)} s(x, z) d\gamma \quad (2)$$

where the slowness function $s(x, z)$ represents the reciprocal of the velocity function $v(x, z)$ in the medium. In agreement with the Fermat principle, the curve $\Gamma(R, S)$ connecting points R and S represents the geodesic in the Riemann space with the metric $s(x, z)$. To solve the inverse kinematic problem means, in the geometrical interpretation, to find the unknown metric $s(x, z)$ by means of the metrical distances $t(R, S)$ known between all boundary points S and R of the region D_{xz} under

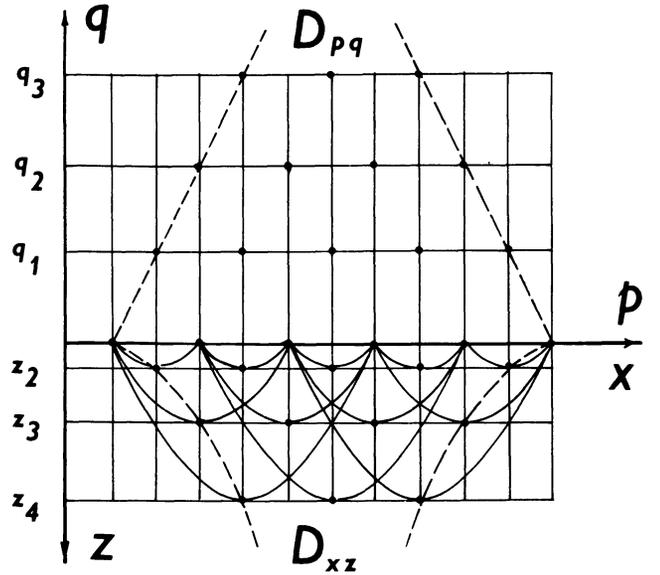


Fig. 2. All grid points (x_j, z_i) in the solution region D_{xz} are defined by the vertices of supporting ray paths $\Gamma_0(p_j, q_i)$, i.e., $x_j = p_j$, $z_i = \xi(q_i)$. The form of the region D_{xz} implies the useful part D_{pq} of input data in the (p, q) coordinate plane

consideration. It is obvious that the expression (2) cannot be inverted with respect to the unknown metric $s(x, z)$ because this metric also defines the geodesics along which the integration is performed. The process of linearization suggested by Romanov (1972) consists in dividing the metric $s(x, z)$ into a regular part $s_0(x, z)$ and a perturbative part $\Delta s(x, z)$, i.e., $s(x, z) = s_0(x, z) + \Delta s(x, z)$, provided that the norms of the functions $s(x, z)$ and $s_0(x, z)$ in the region D_{xz} satisfy the condition $\|\Delta s(x, z)\| \ll \|s_0(x, z)\|$. It can then be proved on the basis of (2) that the travel-time also consists of the regular part $t^0(R, S)$ and the perturbative part $\Delta t(R, S)$, i.e.,

$$t(R, S) = t_0(R, S) + \Delta t(R, S), \quad (3)$$

where $t_0(R, S)$ corresponds to the travel-time of wave propagation along the geodesic $\Gamma_0(R, S)$ generated by the metric $s_0(x, z)$. For the perturbative part it follows that

$$\Delta t(R, S) = \int_{\Gamma_0(R, S)} \Delta s(x, z) d\gamma_0 + \Delta t_2(R, S). \quad (4)$$

The term $\Delta t_2(R, S)$ approaches zero as a quantity of 2nd order with respect to $\Delta t(R, S)$ if the norm $\Delta s(x, z)$ goes to zero. If the term Δt_2 is neglected in expression (4), the relation between Δt and Δs is linear because the geodesic $\Gamma_0(R, S)$ appearing in expression (4) does not depend on Δs . Expression (4), or its differential analogue (Novotný 1980), serves as a base for all algorithms of numerical solution of the linearized inverse kinematic problem.

The Method of Fourier Coefficients

Relation (4) can be used either for the removing of arrival-time deficiency relative to some velocity model obtained in another way (Romanov et al. 1978) or in an iterative mode to find the unknown velocity model with desired accuracy. In the case of weak lateral inhomogeneity we can restrict ourselves to the first iteration and choose the starting model independent of the lateral variable. Romanov (1972) has suggested an algorithm on the basis of expression (4) for a spherically symmetrical starting model, i.e., $s_0(\rho, \varphi) = s_0(\rho)$; ρ ,

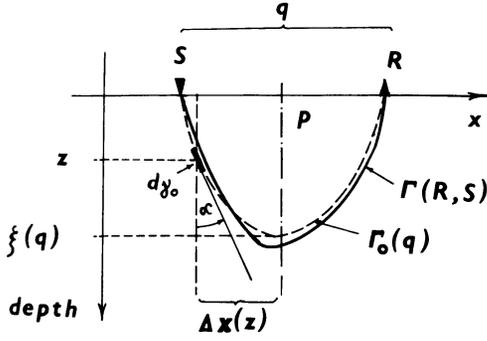


Fig. 3. In a laterally homogeneous medium neither the ray path $\Gamma_0(q)$ nor its characteristics $s_0(z)$, $\Delta x(z)$ and ξ depend on the lateral coordinate p

φ are polar coordinates. The unknown function $\Delta s(\rho, \varphi)$ was sought in the form of finite polynomial expansions in $\exp(i\varphi)$ and ρ . After integration according to Eq. (4), an expression was obtained for $\Delta t(R, S)$ which is linearly dependent on the unknown expansion coefficients. If a satisfactory number of arrival times $t(R, S)$ is available, optimal values of coefficients can be found by the least-squares procedure. Earthquake records along the Pamir-Baykal profile have been interpreted in this way (Alexeyev et al. 1971).

Besides this algorithm, another method of solving the inverse problem is suggested by Romanov (1972). The integral equation relating the Fourier spectra of Δt and Δs functions with respect to the lateral variable p was utilized (Eq. (1)). On the basis of integral equation properties the uniqueness of solution of the inverse linearized problem was proved. However, no numerical algorithm based on the integral equation was suggested.

The Fourier coefficient method presented in this section is based on an expansion of Δt and Δs functions in the lateral variable p by means of Fourier series. This allows derivation of directly solvable system of linear equations for the unknown coefficients of the Fourier series for Δs instead of the integral equation mentioned above.

Let us assume a starting model independent of the lateral variable x . Thus, in a Cartesian coordinate frame we have $s_0(x, z) = s_0(z)$. It is then possible to transform the integration along the geodetics $\Gamma_0(R, S)$ in expression (4) into a simple integral with respect to the variable z :

$$\Delta t(p, q) = \int_0^{\xi(q)} [\Delta s(p - \Delta x(z), z) + \Delta s(p + \Delta x(z), z)] \frac{dz}{\cos \alpha(z)}. \quad (5)$$

Here we used new variables p and q defined by Eq. (1). The maximum depth of the ray path $\Gamma_0(R, S)$ is denoted by ξ , $\alpha(z)$ is the angle between the ray path and the vertical at an arbitrary point (x, z) . The quantity $\Delta x(z)$ (Fig. 3) is given by the following relation:

$$\Delta x(z) = \int_0^z dz' / \tan \alpha(z').$$

Time differences $\Delta t(p, q)$ are determined for individual offsets q_i as the difference between the total values of time isolines $t(p, q_i)$ and the arrival time $t_0(q_i)$ corresponding to the laterally homogeneous starting model $s_0(z)$. It thus follows ($i = 1, 2, \dots, M$)

$$\Delta t(p, q_i) = t(p, q_i) - t_0(q_i). \quad (6)$$

The unknown perturbative part of the metric $\Delta s(x, z)$ will be

now approximated in the region $0 \leq x \leq x_p$, $0 \leq z \leq z_{\max}$ by the finite expansion ($n = -(N-1)/2, \dots, -1, 0, 1, \dots, (N-1)/2$)

$$\Delta s(x, z) = \sum_j \sum_n A_{nj} P_j(z) e^{i2\pi n x / x_p}, \quad (7)$$

where $P_j(z)$, $j = 1, 2, \dots, L$ is a suitable polynomial basis. Now, we want to derive equations for the unknown Fourier coefficients A_{nj} . Therefore, we expand the left-hand side of Eq. (5) in the single variable $p \in \langle 0, x_p \rangle$

$$\Delta t(p, q) = \sum_n B_n(q) e^{i2\pi n p / x_p}. \quad (8)$$

Finally, we obtain from (5), for Fourier coefficients A_{nj} and $B_n(q)$ of expansions (7) and (8), the following linear relation

$$B_n(q) = \sum_j A_{nj} F_{nj}(q), \quad (9)$$

where

$$F_{nj}(q) = 2 \int_0^{\xi(q)} P_j(z) \cos[2\pi n \Delta x(z) / x_p] dz. \quad (10)$$

Expression (9) represents N linear systems from which all Fourier coefficients A_{nj} can be determined in terms of $B_n(q)$ and the starting slowness function $s_0(z)$. Each system consists of as many equations as time isolines are available with different offsets ($q = q_i$, $i = 1, 2, \dots, M$). If the number N of the polynomials used in (7) is equal to the number of time isolines M , i.e., $L = M$, there is one solution of (9). If $L < M$, there is no solution satisfying all equations, but we can perform an optimization process, finding the vector A_{nj} which gives the minimum value of the expression

$$\sum_i |B_n(q_i) - \sum_j F_{nj}(q_i) A_{nj}|^2. \quad (11)$$

Both these possibilities have been included in the numerical realization (Novotný et al. 1979) of the algorithm which enables us to choose the maximum degree L of polynomials independently of the number M of given offsets. The algorithm was tested on an example of a theoretical model with constant velocity gradients.

To characterize the accuracy that is achieved by a linearized method in the case of theoretical models, we used the quantity $\eta(z)$, defined as a ratio of the integral absolute error of the method and the integral absolute lateral deviation of model $s(x, z)$ from the starting model $s_0(z)$ at the depth considered, i.e.

$$\eta(z) = \frac{\int_{x_1(z)}^{x_2(z)} |\Delta s_C(x, z) - \Delta s_T(x, z)| dx}{\int_{x_1(z)}^{x_2(z)} |\Delta s_T(x, z)| dx}. \quad (12)$$

Indices C, T denote successively computed or theoretical values in the interval of interest $\langle x_1(z), x_2(z) \rangle$. It was found that two factors limit the application of the Fourier coefficient method:

(i) Adequacy of the fitting of the input values $\Delta t(p_j, q_i)$ by means of the expansion (8). Figure 4 illustrates the accuracy achieved by the Fourier coefficient method with respect to the different ways of fitting. Curve (1) represents the relative accuracy $\eta(z)$ of the method in the case when the Fourier coefficients were calculated from experimental values $\Delta t(p_i)$ with the help of the discrete Fourier transform. As we could use only a limited number N of the Fourier coefficients ($N \leq 128$), we achieved a better accuracy using the finite Fourier series that approximates the course of $\Delta t(p)$ in the least-squares fashion. Further, we tested the influence of surface velocities on the accuracy of the method. The curve (2a)

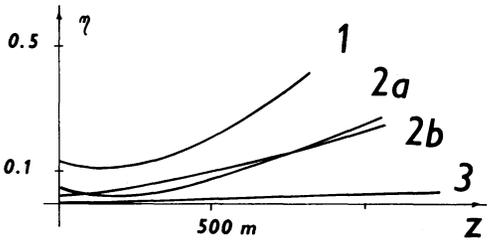
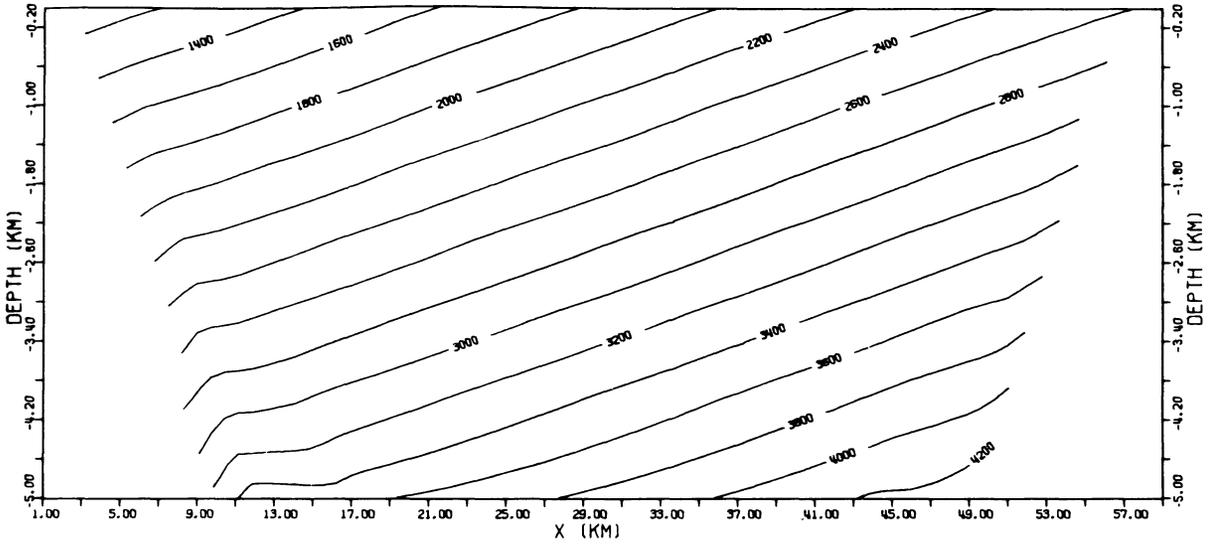
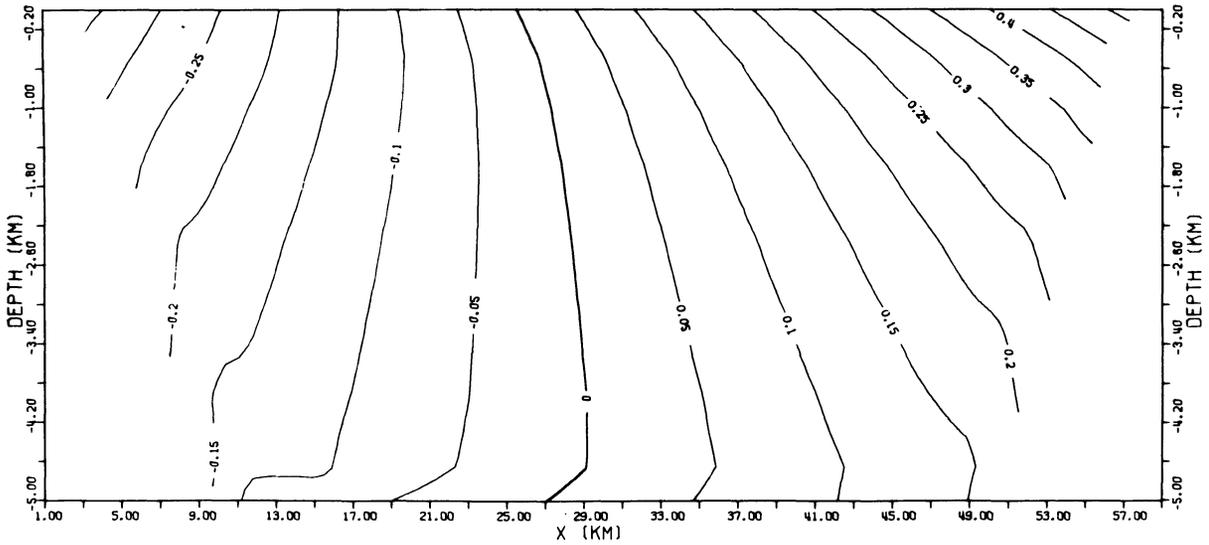


Fig. 4. Curves $\eta(z)$ characterize the relative accuracy of various linearized methods of solving the inverse problem. All the methods were tested for one particular velocity model with constant gradients. Curves 1, 2a, 2b characterize the accuracy of the method of Fourier coefficients, curve 3 the method of direct inversion



THEORETICAL TEST
COMPUTED VELOCITY MODEL

(a)



THEORETICAL TEST
RELATIVE LATERAL DEVIATIONS

(b)

Fig. 5a-b. The test of the direct inversion method for the velocity model $v(x, z) = v_0 + g(x \sin \alpha + z \cos \alpha)$, $v_0 = 1 \text{ km/s}$, $\alpha = 4^\circ$, $g = 4.10^{-4} \text{ s}^{-1}$
a Contour map of velocity distribution calculated on the basis of corresponding time isolines

$$t(p, q) = \frac{1}{v_0 g} \ln \left| \frac{E_1 - gq}{E_1 + gq} \right|, \quad E_1 = \sqrt{E_2 + (qg)^2}, \quad E_2 = 4(1 + pg \sin \alpha)^2 - (qg \sin \alpha)^2.$$

b Contour map of relative deviations between the starting model $s_0(z)$ and the calculated one $s(x, z)$

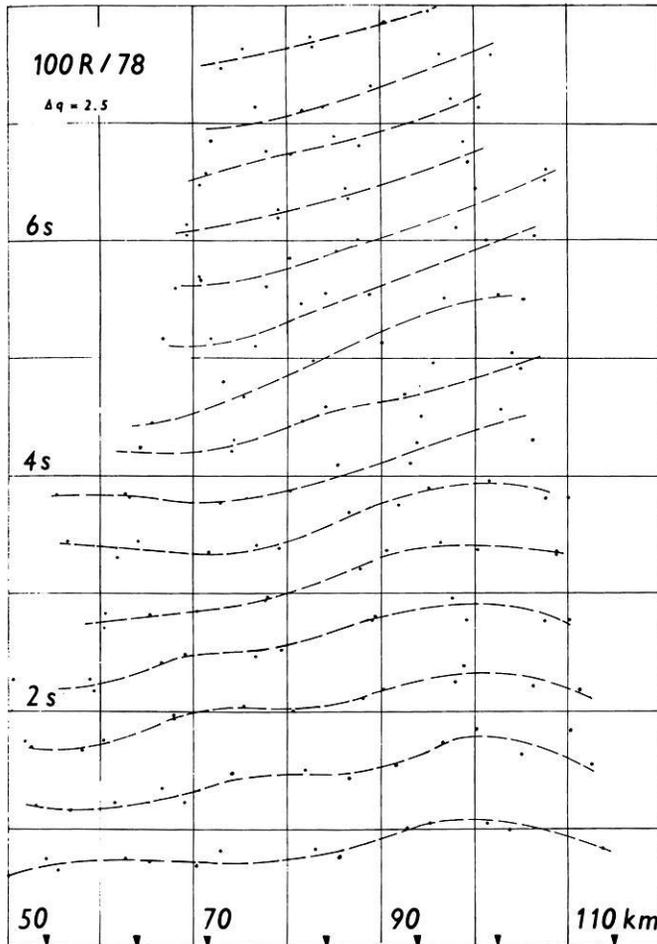


Fig. 6. Time isolines $t(p, q_i)$ obtained from the data, recorded along the Carpathian profile 100R. For the calculation of the velocity section isolines are given with offsets $q_i = i\Delta q$, $i = 1, 2, 4, 6, 8, 11, 13$, and 15, $\Delta q = 2.5$ km

corresponds to the case when the surface velocities are taken into account, while the curve (2b) stands for the relative accuracy $\eta(z)$ in the case when the method works without the input of the surface velocities. We observe that the method is rather insensitive to this influence.

(ii) The shape of the region D_{pq} where values $\Delta t(p, q)$ are located. The natural shape of the input region D_{pq} that enables utilization of all input times is rather trapezium shaped (Fig. 2). However, the method of Fourier coefficients is based on Fourier expansions in the rectangular region $0 \leq x \leq x_p$, $0 \leq q \leq q_M$, eventually in $0 \leq x \leq x_p$, $0 \leq z(q) \leq z(q_m)$. This principal disproportion leads to a considerable deterioration of the convergence of the method. A test has shown that the solution converges only in that part of the region D_{xz} delimited by the rectangle constructed on the shorter base of the trapezium D_{pq} . Therefore, the Fourier coefficient method can only be applied if the maximum measurement offset q_M is much smaller than the length of the interpreted interval x_p , i.e., $q_M \ll x_p$.

The Direct Inversion Method

Using the method of Fourier coefficients the velocity distribution is found over a larger region than it is possible to map

by means of supporting rays $\Gamma_0(p_j, q_i)$ corresponding to the input time data $t(p_j, q_i)$ available. If the region where the solution is sought were limited strictly to the interpretable region D_{xz} formed by the turning points of seismic rays $\Gamma_0(p_j, q_i)$ (Fig. 2) it could be expected that the linearized theory will bring better results. The grid method presented in this section is one possible way to achieve this. Its principle consists in a sequential ray interpretation of time isolines from the smallest offset to the largest. In contrast to the previous method the interpretation of arrival times is based on the knowledge of surface velocities.

Basic Equations

Let us consider a laterally homogeneous starting model $s_0(z)$ such that the arrival times of the surface-to-surface refracted wave form a continuous function $t_0(q)$ with monotonically decreasing derivative. For this class of starting model it is natural to define the grid points in the plane (p, z) to be the turning points of supporting rays $\Gamma_0(p_j, q_i)$ (Fig. 2). Coordinates of grid lines are then as follows

$$p_j = j\Delta p, \quad j = \dots, -1, 0, 1, \dots$$

$$z_i = \xi(q_i), \quad i = 1, 2, \dots, M,$$

where $\xi(q)$ denotes a continuous and monotonically increasing function of offset q . In individual steps of interpretation we seek the unknown grid values $\Delta s(p_j, z_i)$ on individual grid levels z_i . In the first step we determine only the lateral slowness deviations on the surface using the input values of surface velocities and the mean velocity $s_0^{-1}(0)$ following from the chosen initial model $s_0(z)$. In the successive steps we interpret the individual time isolines $\Delta t(p_j, q_i)$ in the following recursive manner.

Let us assume that we have found the velocity distribution to the depth z_i . Now, we shall divide the integration in expression (5) along the ray trajectory $\Gamma_0(q_{i+1})$ into two parts separated by the grid line $z = z_i$. The integral in the region already mapped for $z \leq z_i$ can be performed. Therefore, we write expression (5) in the form

$$\Delta t(p_j, q_{i+1}) - \int_{0 \leq z \leq z_i} \Delta s(x, z) d\gamma_0 = \int_{z_i \leq z \leq z_{i+1}} \Delta s(x, z) d\gamma_0. \quad (13)$$

The left-hand side represents a value of the quantity $\Delta t(p, q)$ related to the subsurface level $z = z_i$. For this we introduce the notation $\Delta t(p_j, q_{i+1}, z_i)$. This quantity can be expressed on the basis of known grid values $\Delta s(p_n, z_m)$, $m = 1, 2, \dots, i$, in a reasonable numerical approximation. The right-hand side of expression (13) depends on unknown grid values $\Delta s(p_n, z_{i+1})$ on the next grid line $z = z_{i+1}$. If the grid spacing is appropriate to the character of the medium studied, one can restrict oneself to a linear interpolation of function $\Delta s(x, z)$ along this part of the ray path. The right-hand side can then be expressed in linear form with respect to unknown grid values. Expression (13) can then be written in the following form

$$\Delta t(p_j, q_{i+1}, z_i) = C_j + \sum_n D_{jn} \Delta s(p_n, z_{i+1}). \quad (14)$$

Vector C_j as well as matrix D_{jn} are entirely determined by values $\Delta s(p_n, z_i)$ on the preceding grid line $z = z_i$, then by the shape of the ray path $\Gamma_0(q_{i+1})$ in the depth interval $z_i \leq z \leq z_{i+1}$ and finally, by the kind of numerical approximation of the line integral. Summation index n goes through the same values as the index j . Matrix D_{jn} is therefore square. It

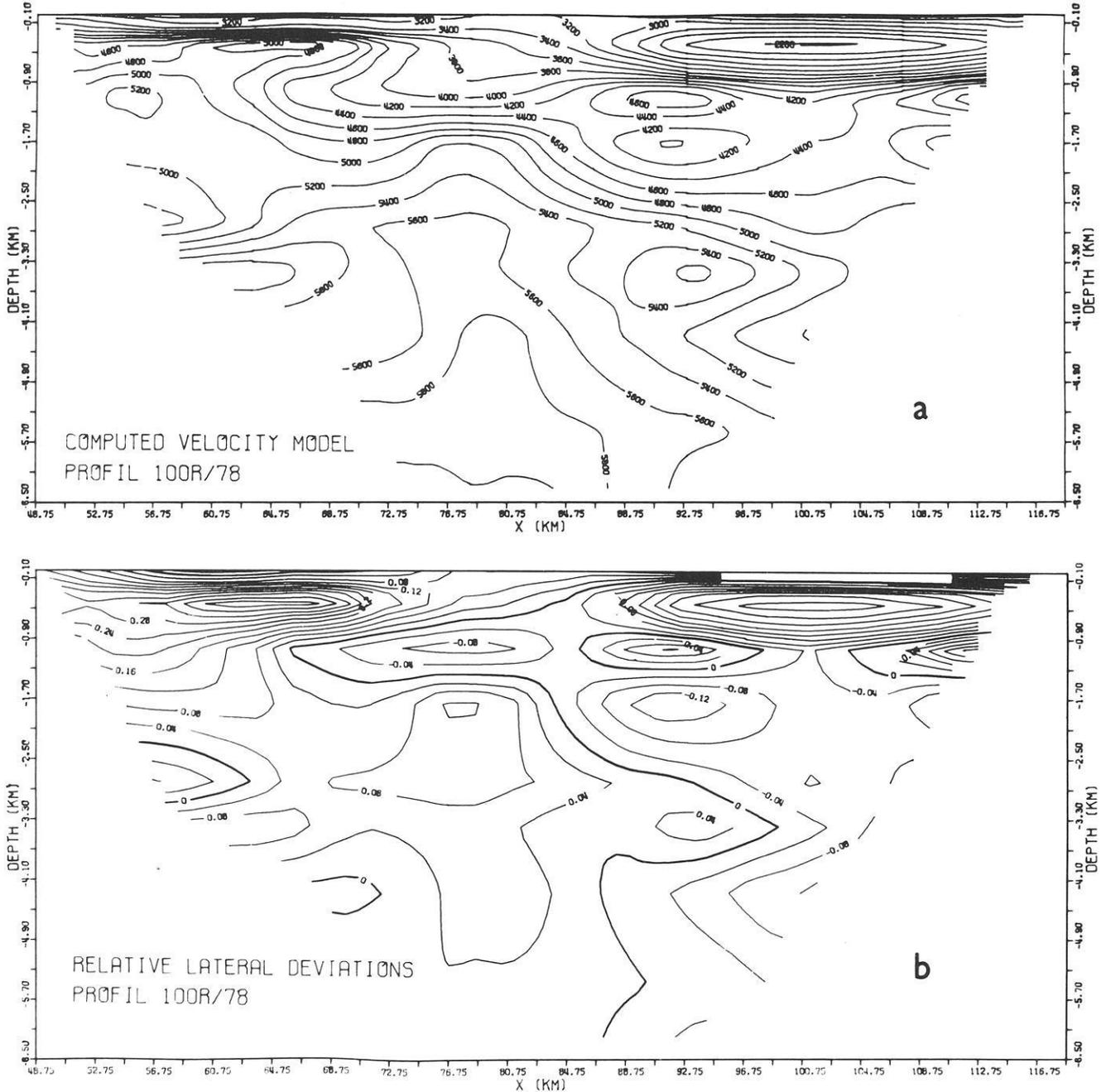


Fig. 7. **a** Contours of velocity distribution along profile 100R calculated by the method of direct inversion on the basis of selected time isolines according to Fig. 6; **b** Contours of relative lateral deviation between the starting model $s_0(z)$ and the calculated one $s(x, z)$

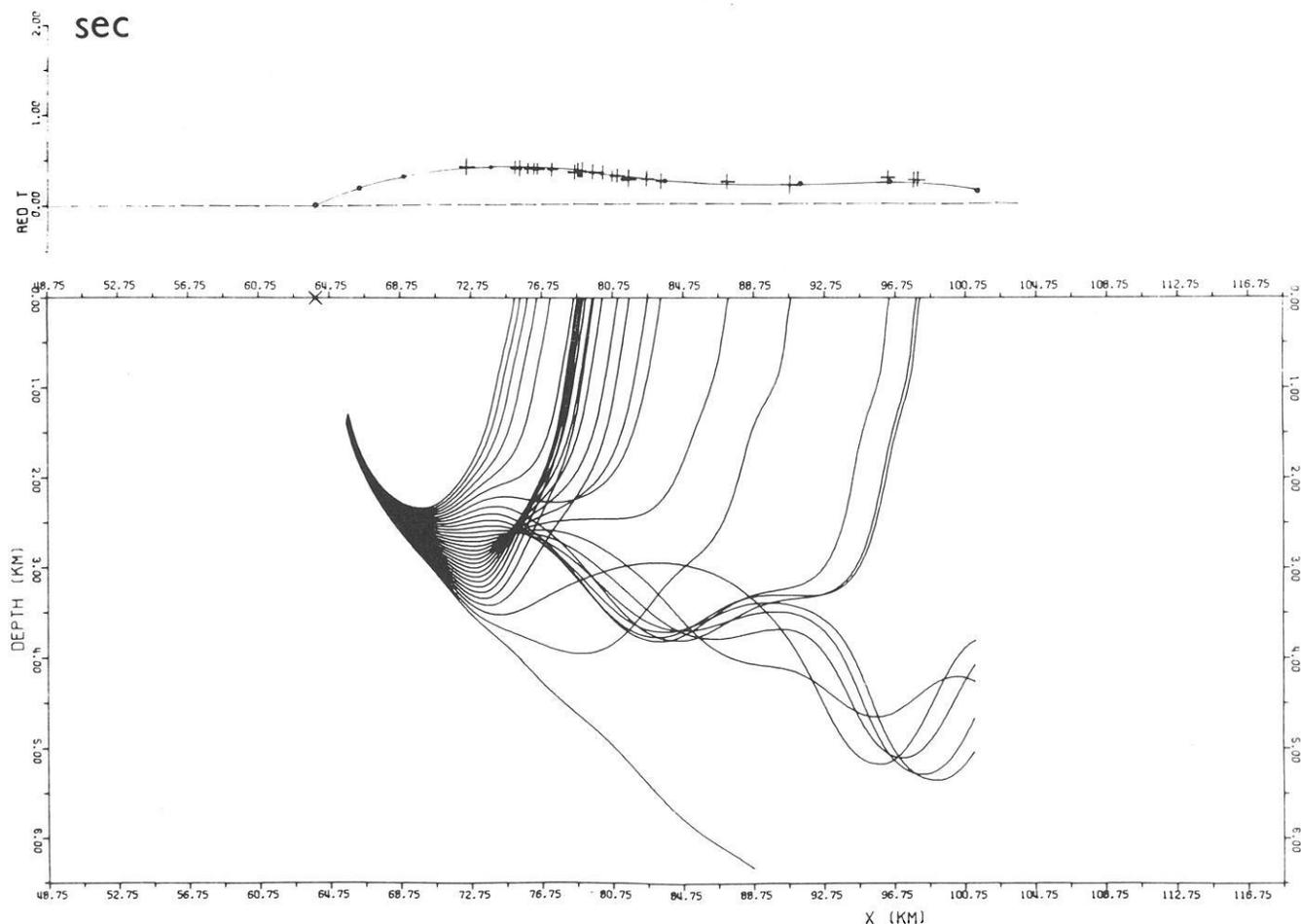
can be shown that it is symmetrical with a limited number of non-zero diagonals given by the number of grid values $\Delta s(p_n, z_{i+1})$ on line $z = z_{i+1}$ necessary to perform the interpolation of $\Delta s(x, z)$ along the deepest part of the ray $\Gamma_0(p_j, q_{i+1})$. In particular, if the interpolation in the lateral direction is suppressed, a diagonal form of matrix A_{jn} is obtained. In this case, one input time value $t(p_j, q_{i+1})$ contained at the left-hand side of the relation (14) results in one grid slowness value $\Delta s(p_j, z_{i+1})$. This property of direct inversion $\Delta t(p_j, q_{i+1}) \rightarrow \Delta s(p_j, z_{i+1})$ led to the name of the method.

The system of linear Eq. (14) thus enables us to state all unknown grid values $\Delta s(p_n, z_{i+1})$ on the basis of the known time differences $\Delta t(p_j, q_{i+1})$ of the interpreted $(i+1)$ -th time isoline. After performing all M steps of the interpretation, the

entire region D_{xz} is then mapped out. Expression (14) is the basic one for the method of direct inversion.

Applications

On the basis of expression (14) a numerical algorithm was derived and a program implementing the method of direct inversion was written. Input data are surface velocities and arrival times of surface-to-surface refracted waves arranged according to individual measurement offsets in the interval investigated $\langle 0, x_p \rangle$. The accuracy of the line integral calculation that occurs in expression (13) depends on the subdivision of the ray path $\Gamma_0(q)$. In the computer implementation the accuracy of a calculation of C_j coefficients as well



PF 100R/78 PKV 64.00

Fig. 8. Test of the calculated velocity model in Fig. 7a for the position of the source $x_s = 64.00$ km. The reduction velocity is 5 km/s. ..., arrival times derived from input time isolines; + + +, arrival times calculated by means of a ray tracing program. The corresponding ray paths are illustrated in the lower part of the figure

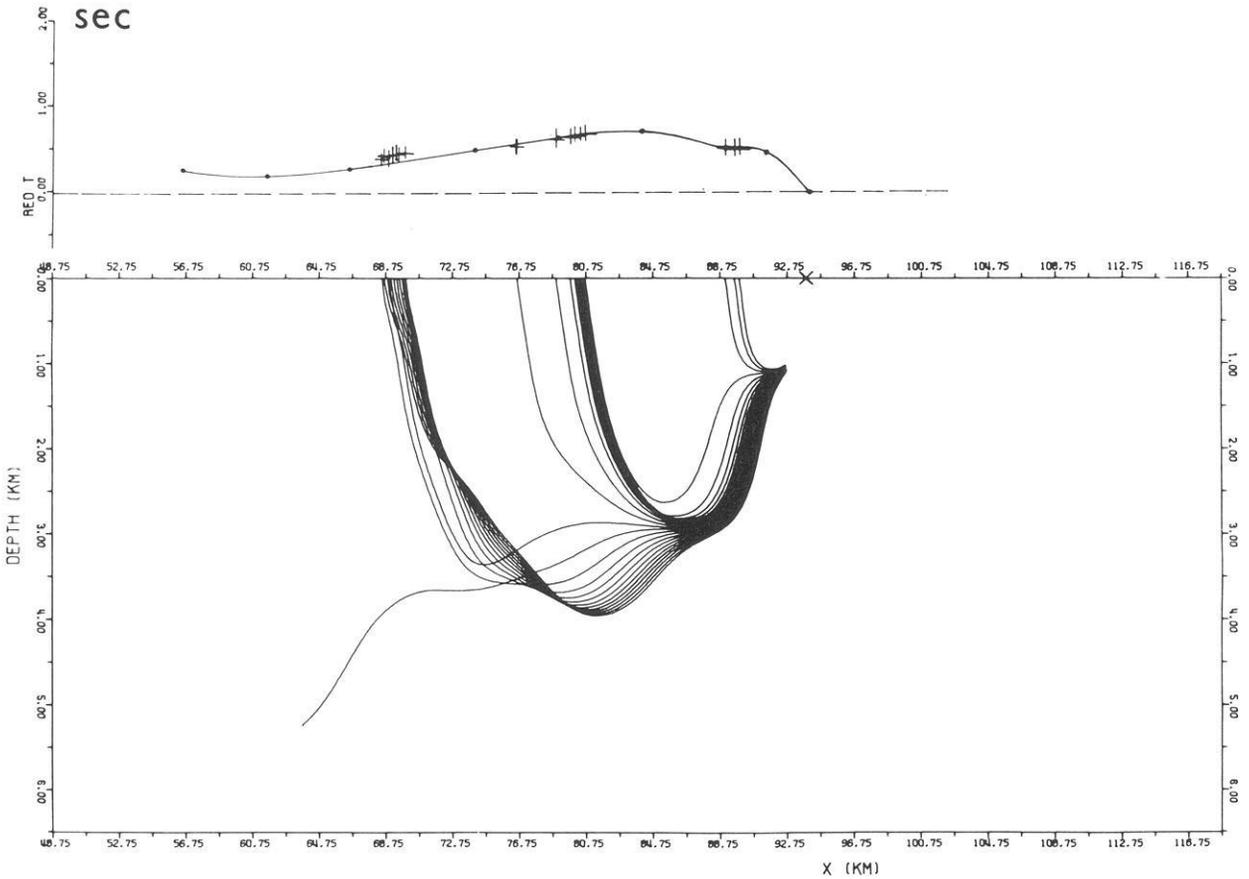
as matrix elements D_{jn} is controlled by choice of an adequate number of subelements of the ray path between neighbouring grid lines z . Further, the calculation of the velocity distribution can be performed at any sub-interval of the basic interval $\langle 0, x_p \rangle$.

The supporting travel-time curve $t_0(q)$ is obtained by centering the individual time isolines within the chosen sub-interval. The starting model $s_0(z)$ is then derived by the Wiechert-Herglotz transformation of a smoothed version of $t_0(q)$. In this way we always interpret the minimum lateral time deviations in the chosen sub-interval, when the linearized theory works best (Fig. 1). By defining sub-intervals with certain overlapping parts and by a mutual comparison of the velocity models calculated here, the consistency of the linearized approach can also be verified.

The method has been tested on theoretical as well as real data. Figure 5 shows the result of an interpretation of theoretical time isolines corresponding to a velocity model with constant velocity gradients. The lateral inhomogeneity reached nearly $\pm 50\%$ in the upper part of the section (Fig. 5b). In spite of this fact, isolines of the calculated velocity distribution are in good agreement with the theoretical equidistant straight lines over the entire region of the solution obtained, with the exception of a narrow region close to the margins (Fig. 5a). A quantitative evaluation of the method's

accuracy is presented in Fig. 4. Curve (3) shows the dependence of relative accuracy $\eta(z)$ on the depth z according to expression (12). In comparison with the one-step method of Fourier coefficients (curves 1 and 2) the method of direct inversion shows a substantially better accuracy.

Time isolines in Fig. 6 were drawn from refraction data recorded along a part of the Carpathian profile 100 R in West Slovakia (Rektořík et al. 1979). In this portion of the profile time isolines exhibit a considerable lateral dependence corresponding to a transition from the region of pre-Neogene formations into Neogene basins. The contour map of the velocity distribution in Fig. 7 was derived from calculated grid data $\Delta s(x, z)$ as well as contour lines of the relative lateral deviation from the velocity starting model. It is obvious from this figure that the lateral deviation varies from approximately $+40\%$ in the region of pre-Neogene formations to -30% in the region of lower velocities in the Neogene sedimentary basins. To verify the results we chose the two longest travel-time curves with shot points fixed at $x = 64.0$ km and 93.9 km (solid lines in Figs. 8 and 9). By means of a ray tracing program (Červený et al. 1977) the direct problem for the obtained velocity model was solved. The ray paths, together with corresponding travel times for the shot points tested, are illustrated in Figs. 8 and 9. In spite of the complex character of the model tested (strongly variable gra-



PF 100R/78 PKV 93.9

Fig. 9. Test of the calculated velocity model for the source position $x_s=93.90$ km. The reduction velocity is 5 km/s. + + +, arrival times calculated by means of ray tracing program. The corresponding ray paths are illustrated in the lower part of the figure; ..., arrival times derived from input time isolines

dients, local velocity inversions) time discrepancies larger than 2% have not been detected in the tests.

Recently the method has been applied on further regional refraction profiles in Czechoslovakia (Novotný et al. 1981).

Concluding Remarks

Methods for solving inverse seismic problems can be divided into two groups:

(i) Evolutional, multi-step methods that, on the basis of surface data, map the seismic medium successively from the surface down to some depth. The result of each step is dependent on results of preceding steps. Included here, for instance, are the so called wave-equation migration methods which form an image of the seismic medium using information carried by the reflected wave field (Claerbout 1971). Of the methods for solving the inverse kinematic problem mentioned above, the differential method (Belonosova et al. 1967) and the method of characteristics (Romanov et al. 1978) can be designated as evolutional, as can the method of direct inversion presented in the section above.

(ii) One-step methods where the solution is sought for the entire region investigated by means of an optimization procedure. The method suggested by Romanov (1972) is a one-step methods as is the method of Fourier coefficients derived above.

As far as the character of the problem allows, it is advantageous to use numerical algorithms of the evolutionary type (lower requirements for computer memory, mostly simpler algorithms, a natural possibility of intermediate results management). In spite of various numerical implementations, common principles of evolutionary algorithms can be found, i.e.,

- the values of the interpreted quantity registered at the surface ($z=0$) are recursively transformed to lower and lower depth levels $0 < z_1 < z_2 < z_3 \dots$,

- the result of the interpretation is yielded in terms of the quantity transformed for the zero travel time of the seismic wave of interest.

In the case of holoseismic techniques the reflected wave field is transformed. While solving the inverse kinematic problems, we transform the travel time of refracted waves directly. In particular, the transformation of time isolines used by the direct inversion method consists in the recalculation of the travel-times $t(R, S)$ from the position of source S and receiver R on the surface $z=0$ recursively to lower and lower depth levels $z=z_i$. In the limit, when, during the recalculation, coincidence of the source and receiver is achieved (i.e., transformed travel-time reaches zero), the value of the slowness function can be obtained by differentiating $t(R, S)$ with respect to the distance q of the source and receiver. In terms of p, q introduced in Eq. (1) this transformation principle assumes the following simple form

$$\Delta s(x, z) = \frac{\partial t(p, q, z)}{\partial q} \Big|_{\substack{p=x \\ q \rightarrow 0}} \quad (15)$$

where $t(p, q, z)$ is the travel time recalculated along the ray path $\Gamma(p, q)$ to the level z . An expression like (15) in a finite-difference form represents the basic equation of the method suggested by Belonosova et al. (1967). To extrapolate the surface arrival times depth, a non-linear differential equation is solved by finite-difference approximation.

In the linearized approach used in the direct inversion method, the extrapolation is realized in a form of a line integral, such as the left-hand side of expression (13) for the deviation $\Delta t(p, q, z)$ of the propagation time. The right-hand side of expression (13) then represents the linearized analogue of the transformation principle (15) in integral form: the limiting transition $q \rightarrow 0$ is replaced by the linear interpolation of the function Δs along the ray path trajectory. It appears that the choice of the natural grid points as the vertices of the supporting seismic ray paths where the relation (15) is fulfilled, in particular, leads to a high accuracy for the linearized method even in the case of a low density of grid points.

References

- Alexeyev, A.S., Lavrentiev, M.M., Mukhometov, R.G., Nersesov, I.L., Romanov, V.G.: Method for numerical investigation of the upper Earth's mantle. In: *Mathematical problems of geophysics*, Vol. 2, M.M. Lavrentiev and A.S. Alexeyev, eds.: pp 143–165 (in Russian). Novosibirsk: Computing center Acad. Sci USSR (Sib. Branch) 1971
- Belonosova, A.A., Alexeyev, A.S.: On a statement of the inverse kinematic seismic problem for two-dimensional continuously inhomogeneous medium. In: *Nekotorye metody i algoritmy interpretatsii geofizicheskikh dannykh*. M.M. Lavrentiev, ed.: pp. 137–154. Moscow: Nauka 1967
- Beránek, B., Červený, V., Konířová, L., Pšenčík, I., Zounková, M.: Interpretational methods in seismic investigation of deep structures. *App. Geophys.* **16**, 63–89, 1979
- Bessonova, E.N., Fishman, V.M., Ryaboyi, V.M., Sitnikova, G.A.:

- The tau method for the inversion of travel times – I, *Deep Seismic Sounding Data*. *Geophys. J.R. Astron. Soc.* **36**, 377–398, 1974
- Bessonova, E.N., Fishman, V.M., Shnirman, M.G., Sitnikova, G.A.: The tau method for inversion of travel times – II, *Earthquake Data*. *Geophys. J.R. Astron. Soc.* **46**, 87–108, 1976
- Červený, V., Molotkov, I., Pšenčík, I.: *Ray method in seismology*. Prague: Charles University 1977
- Claerbout, J.F.: Toward a unified theory of reflector mapping. *Geophysics* **36**, 467–481, 1971
- Gerver, M., Markushevitch, V.: Determination of seismic wave velocity from the travel-time curve. *Geophys. J.R. Astron. Soc.* **11**, 165–173, 1966
- Gerver, M., Markushevitch, V.: On the characteristic properties of the travel-time curve. *Geophys. J.R. Astron. Soc.* **13**, 241–246, 1967
- Johnson, L.E., Gilbert, F.: Inversion and interference for teleseismic ray data. In: *Methods in Computational Physics*, Vol. 12, B. Alder, ed.: pp. 231–266. New York: Academic Press 1972
- Kennett, B.L.N.: A comparison of travel-time inversions. *Geophys. J.R. Astron. Soc.* **44**, 517–536, 1976
- Novotný, M., Firbas, P., Bursa, O.: Solution of linearized inverse kinematic problem by the method of Fourier coefficients. In: *Proceedings of the 23rd Geophysical Symposium*, Krakow 1979
- Novotný, M.: Applications of linearized methods to multi-dimensional inverse kinematic problems. In: *Proceedings of the ESC workshop meeting*. Suzdal, 2nd–12th October, 1980
- Novotný, M., Nehybka, V., Sedlák, P.: First applications of the method DIME to regional refraction profiles in Czechoslovakia. In: *Proceedings of the 25th Geophysical Symposium*, Leipzig 1981
- Puzyrev, N.N., Krylov, C.V., Mishenkin, B.P.: *Methods of regional seismic studies*. Novosibirsk: Nauka 1975
- Rektořík, J., Mayerová, M., Nehybka, V., Sedlák, P.: Refraction prospecting in the West Carpathians in the years 1976–1978 (in Czech). *Archives of Geofyzika n.p.*, Brno 1979
- Romanov, M.E., Alexeyev, A.S.: A characteristic method for numerical solution of the inverse kinematic seismic problems. *J. Geophys.* **48**, 173–180, 1980
- Romanov, V.G.: *Inverse kinematic problems for hyperbolic equations*. Novosibirsk: Nauka 1972
- Slichter, L.B.: The theory of the interpretation of seismic travel-time curves in horizontal structures. *Physics* **3**, 273–295, 1932

Received February 9, 1981; Revised Version May 18, 1981

Accepted May 18, 1981

Palaeomagnetism of Upper Cretaceous Volcanics and Nubian Sandstones of Wadi Natash, SE Egypt and Implications for the Polar Wander Path for Africa in the Mesozoic

A. Schult¹, A.G. Hussain², and H.C. Soffel¹

¹ Institut für Allgemeine und Angewandte Geophysik, Universität München, Theresienstr. 41, 8000 München 2, Federal Republic of Germany

² Helwan Institute of Astronomy and Geophysics, Helwan, Cairo, Egypt

Abstract. Eighteen sites (342 samples) from Upper Cretaceous Wadi Natash volcanics (24.5° N, 34.5° E) yield a mean direction of magnetization $D=345.4^\circ$, $I=16.7^\circ$ with $\alpha_{95}=8.5^\circ$, $k=21.4$, $N=15$ after AF cleaning resulting in a pole at 69.3° N, 258° E with $A_{95}=5.8^\circ$. All sites have normal polarity consistent with their age and the magnetic stratigraphy in the Cretaceous. From 5 sites (85 samples) from Upper Cretaceous Nubian sandstone at Wadi Natash a mean direction $D=358.1^\circ$, $I=32.0^\circ$ with $\alpha_{95}=8.7^\circ$, $k=143$ (mixed polarity) was obtained after thermal demagnetization. Combined with previous investigations on Nubian sandstone at other locations in Southern Egypt (Schult et al. 1978) this yields a pole at 81.8° N, 223° E with $A_{95}=3.3^\circ$, $N=23$. 9 sites from Eocene Baharia iron ores (27.5° N, 29.0° E) yield a mean direction $D=188.0^\circ$, $I=-43.6^\circ$ and $\alpha_{95}=6.4^\circ$, $k=65$ with a pole at 83.5° N, 139° E and $A_{95}=7^\circ$.

In addition the palaeomagnetism of some Tertiary basaltic rocks in Northern Egypt was studied. The polar wander path for Africa in Mesozoic time is presented showing more mobility than in earlier papers. For appropriate reconstructions of South America with respect to Africa the polar wander paths of both continents are substantially in agreement.

Key words: Palaeomagnetism – Africa – Mesozoic polar wander path

Introduction

The Cretaceous polar wander path for Africa is of particular interest because this time coincides with the opening of the Atlantic. An improvement in the time density of known African Cretaceous poles is desirable in order to define the polar wander path in a more detailed fashion for this period of time of about 75 m.y. duration. In this paper the palaeomagnetism of Upper Cretaceous volcanics and Nubian sandstones from Wadi Natash, Egypt is presented. Previous investigations of these formations were carried out by El Shazly and Krs (1973) and by Schult et al. (1978). Palaeomagnetic investigations of Tertiary iron ores and basalts from Baharia Oasis and Qatrani (SW of Cairo) were completed and the results are also presented here. A modification of the apparent polar wander path for Africa in the Mesozoic will be proposed.

Geological Setting

Wadi Natash Volcanics

Figure 1 shows a geological sketch map of the Wadi Natash area and the sampling sites. The volcanics include lava flows

of basaltic, andesitic, phonolitic and rarely trachytic composition (El Ramly 1972). They are interbedded in the Nubian sandstone at some localities particularly in the south-western parts of the Wadi. The volcanics are dated as Upper Cretaceous for stratigraphic reasons (Said 1971) consistent with radiometric age determinations yielding ages between 100 and 86 m.y. (El Shazly and Krs 1973; El Shazly 1977). Samples were collected from 17 sites (Fig. 1). Two of them are located at the ring complex of Gabal El Ghorfa which is believed to be cogenetic with the other volcanics (Said 1971).

Wadi Natash Nubian Sandstone

The Nubian sandstone extends over large parts of Southern Egypt and has previously been studied palaeomagnetically at other localities (Schult et al. 1978). The age of the Nubian sandstone is Upper Cretaceous (Said 1971). Samples were collected from 12 sites (Fig. 1), some of them are associated with volcanics (e.g. sites 14, 15 and 26).

Tertiary Rocks from Qatrani and Baharia Oasis

The basalt from Qatrani (20 km SW of Cairo) was extensively sampled at widely spaced localities. This large flow represents one site only, with a mean radiometric age of 26 m.y. (El Shazly 1977). From basaltic rocks in the Baharia oasis, 5 sites were sampled. The basalts, which overlie the Eocene Baharia formation (El Akkad and Issawi 1963) have a radiometric age of 20 m.y. (El Shazly 1977). The sampling of the Baharia iron ores (Schult et al. 1978) was completed at two sites. An Upper Eocene age is attributed to these ores (Said 1971).

Sampling and Measurements

About six oriented blocks were collected from each site except for the Qatrani basalt and Baharia iron-ores where 15 blocks were taken. Cores of 2.5 cm in diameter were drilled from the blocks and cut into specimens of 2.4 cm length. The natural remanent magnetization (NRM) was measured using a spinner fluxgate or a stationary fluxgate magnetometer. Three or four pilot samples from each site of the volcanics were then subjected to alternating field demagnetization in steps up to 1,000 Oe peak value. In Fig. 2a–d vector diagrams show typical examples for the removal of viscous magnetization by AF treatment. All other samples were subjected to AF demagnetization with appropriate peak values according to the vector diagrams for the pilot samples and the characteristic remanent magnetization (CARM) was measured.

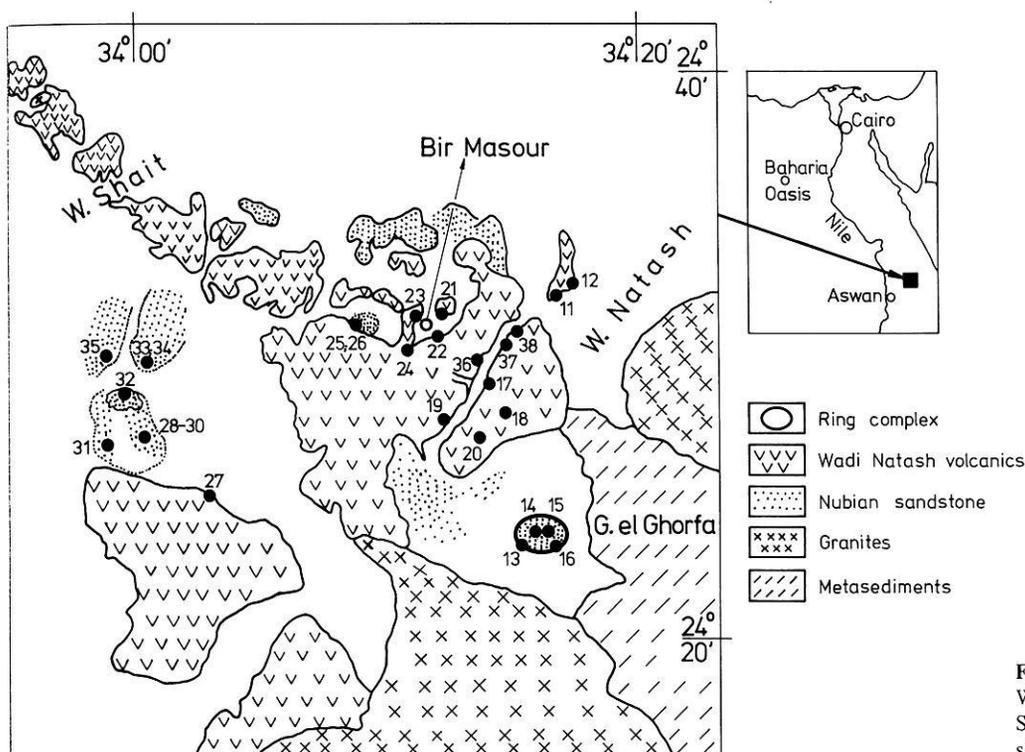


Fig. 1. Geological sketch map of Wadi Natash area (El Ramly 1972; Sabet et al. 1973) showing sampling sites

The samples from Nubian sandstone were subjected to thermal demagnetization because AF treatment had no effect on direction or intensity of the remanent magnetization. Typical vector diagrams of thermal demagnetization of pilot samples are shown in Fig. 3a-c from which the appropriate demagnetization temperature for all samples was chosen.

Results

Wadi Natash Volcanics

The results are summarized in Table 1 and the circles of confidence of the site-mean directions of NRM and CARM are shown in Fig. 4. The within-site scatter as well as the between-site scatter was reduced by the AF treatment. All sites have normal direction of magnetization. The NRM of site 27 (Fig. 2c, Fig. 4) was obviously overprinted by a nearly reversed direction, which was removed by the AF treatment. For three sites no consistent results could be obtained. Giving unit weight to each site the mean direction of CARM of 15 sites is $D=345.4^\circ$, $I=+16.7^\circ$, $\alpha_{95}=8.5^\circ$. The mean palaeopole (69.3° N, 258.1° E, $A_{95}=5.8^\circ$) is rather different from the result of El Shazly and Krs (1973) for the same formation. However their data show a relatively large scatter with $k=4.85$. This would yield a large circle of confidence about the mean giving unit weight to each site, rather than to each sample as was done by El Shazly and Krs (1973). The significance of the difference of the two collections might then be questionable. The angular standard variation of our collection ($s \approx 81 \times k^{-0.5}$, $k=21.4$) equals 17.5 and indicates the usual palaeosecular variation for that latitude (Brock 1971).

The observation of only normal direction of magnetization is consistent with the radiometric age (100–86 m.y.) and the geomagnetic polarity sequence for the Cretaceous. After Lowrie et al. (1980) the long normal polarity sequence lasted from 115–78 m.y. B.P. with some uncertain reversals between 109 and 96 m.y.

Wadi Natash Nubian Sandstone

The results are listed in Table 2 and shown in Fig. 5. For a relatively large number of sites no consistent result could be obtained, and site 26 was discarded because the respective VGP is far away from the mean. The average pole position of the remaining 5 sites (82.9° N, 231° E, $A_{95}=6.4^\circ$) is different from the pole position of the Wadi Natash volcanics and agrees well with the pole position of Nubian Sandstone from other areas of Egypt previously published (80.4° N, 227° E with $A_{95}=4.1^\circ$ and $N=18$) (Schult et al. 1978). The best estimate for the pole position is probably a combination of both collections (giving unit weight to each site). Together with a slight correction of the variance used in the previous paper (involved in the orientation of the samples) the new calculation yields a pole position at 81.8° N, 223° E with $A_{95}=3.3^\circ$, $K=84$ and $N=23$ which is only slightly different from the present dipole position (see also Fig. 5).

The majority of the sites of the Upper Cretaceous Nubian sandstone have reversed polarity indicating an age younger than 78 m.y. B.P. (younger than the long normal polarity sequence in the Upper Cretaceous) (Schult and Guerreiro 1980).

Other Results

All other palaeomagnetic results are summarized in Table 3, together with the combined results from previous investigations. The pole for the Upper Eocene Baharia iron ores is in agreement with other Tertiary poles for Africa (Schult 1974). The Miocene basaltic flows of Baharia oasis and Oligocene basalts (Qatrani and other basaltic flows in the surroundings of Cairo (Hussain et al. 1976) have somewhat unusual palaeopoles for the Tertiary in relatively low latitudes. However, the number of sites is restricted at the present stage and it seems too early to draw further conclusions.

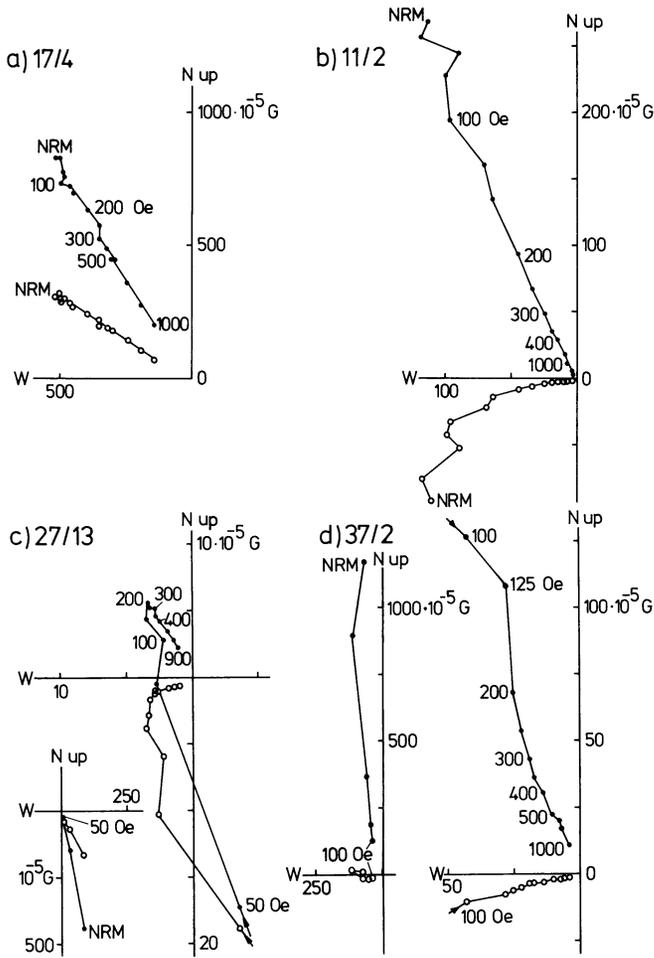


Fig. 2a-d. Vector diagrams showing the variation of the remanence vector during progressive AF demagnetization of samples from volcanic rocks. *Open* and *solid symbols* indicate components in the vertical EW and horizontal planes, respectively. The remanence is relatively stable in **a** and **b** and in some cases as in **c** and **d** it is relatively unstable. In **c** the horizontal component of CARM is overprinted by a reversed unstable component. $1 \text{ G} = 10^3 \text{ Am}^{-1}$, $1 \text{ Oe} \cong 10^{-4} \text{ T}$

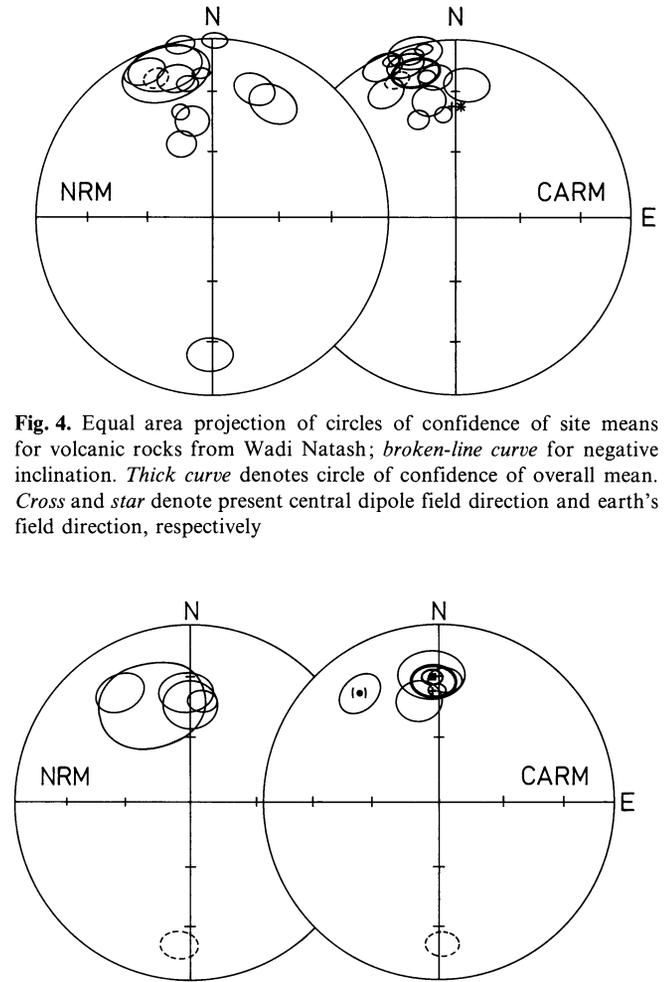


Fig. 4. Equal area projection of circles of confidence of site means for volcanic rocks from Wadi Natash; *broken-line curve* denotes circle of confidence of overall mean. *Thick curve* denotes circle of confidence of overall mean. *Cross* and *star* denote present central dipole field direction and earth's field direction, respectively

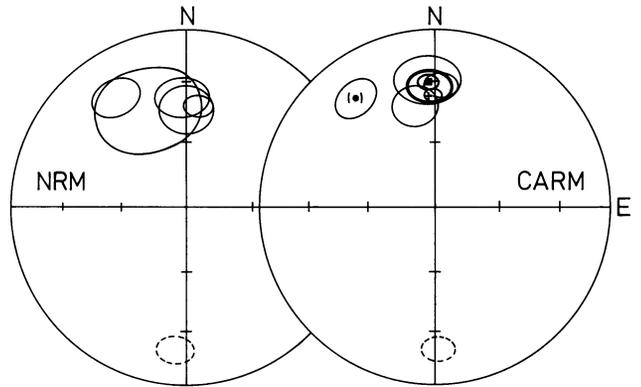


Fig. 5. Circles of confidence of site means for Nubian Sandstone from Wadi Natash. *Broken-line curve* denotes negative inclination, *thick curve* the overall mean omitting the mean in brackets. *Square* marks the overall mean with circle of confidence combining all data (see Table 2), calculated for a site location at Wadi Natash. *Cross* denotes present dipole field direction

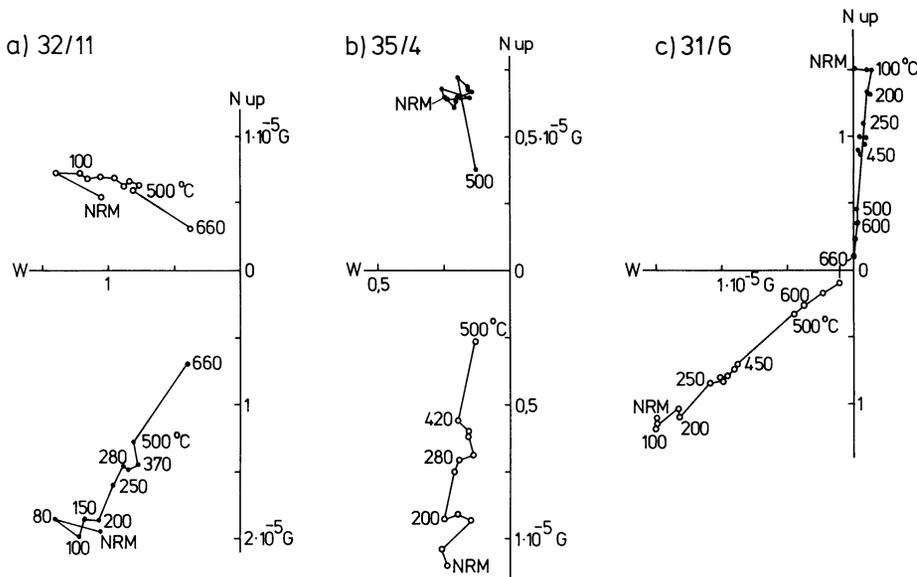


Fig. 3a-c. Vector diagrams showing the variation of the remanence vector during progressive thermal demagnetization of the Nubian sandstone. **a** and **b** *Open* and *solid dots* indicate components in the vertical EW and horizontal planes, respectively. **c** *Open dots* indicate components in the vertical NS plane with its Northern end pointing to the *left (W)*; *solid dots* indicate components in the horizontal plane

Table 1. Site mean palaeomagnetic results for Wadi Natash volcanic rocks (mean site location 24.4° N, 34.25° E)

Site	NRM					MDF(Oe)	AF(Oe)	CARM					VGP		
	<i>N</i>	<i>D</i>	<i>I</i>	α_{95}	<i>k</i>			<i>N</i>	<i>D</i>	<i>I</i>	α_{95}	<i>k</i>	°N	°E	
11	20	343	+16.1	16.1	5	150	250	19	346.6	+ 5.4	9.2	14.2	64.2	247	
12	16	18	+25.3	8.0	18.6	80	200	21	345.0	+ 7.2	4.9	41.8	64.7	251	
13	19	349	+44.3	7.4	21.0	100	150	21	347.7	+33.3	7.0	21.1	77.0	270	
16	16	No consistent result													
17	18	339	-17.6	4.9	50	500	200	19	337.5	-18.1	4.5	56	49.8	250	
18	15	28	+28.9	9.9	15.8	40	150	16	4.8	+26.2	8.5	19.7	78.5	190	
19	3	343	+12.7	13	51	300	200	4	334.8	+ 7.4	9.1	103	58.1	268	
20, 21	No consistent result														
22	20	344	+38.5	3.2	103	150	200	21	353.7	+41.7	4.1	61	84.2	302	
23	20	336	+11.2	7.3	20.7	20	200	21	342.4	+11.4	6.3	26	64.8	259	
24	22	356	+20.0	2.5	142	200	300	20	352.0	+17.1	2.0	269	72.6	242	
25	31	338	+54.0	6.3	17.2	100	200	34	339.8	+41.2	4.3	33.7	71.6	307	
27	15	181	+22.3	8.9	19.1	20	270	11	331.6	+21.9	7.4	39.1	60.1	284	
36	5	350	+ 2.3	4.9	297	200	200	5	338.3	+ 6.0	2.7	763	60.1	262	
37/1	8	1	+ 1.4	4.4	162	30	180	8	349.5	+ 4.5	3.4	275	65.6	241	
37/6	4	346	+20.8	6.9	179	150	500	4	352.1	+21.5	6.6	195	74.8	245	
38	24	350	+24.9	3.7	64	200	200	24	348.5	+20.4	3.6	68	72.3	255	
Mean of site means								15	345.4	+16.7	8.5	21.4	69.3	258.1	
												$A_{95}=5.8$	$K=44.9$		

N=number of samples, *D*=declination, *I*=inclination, α_{95} or A_{95} =radius of 95% confidence circle, *k* or *K*=precision parameter, MDF=medium destructive peak field necessary to erase half of NRM intensity by alternating field demagnetization, AF=peak value of alternating field used for demagnetization (10^4 Oe \approx 1 T)

Table 2. Site mean palaeomagnetic results for Nubian Sandstone from Wadi Natash (ca. 24.45° N, 34.0° E)

Site	NRM					Cleaning	CARM					VGP	
	<i>N</i>	<i>D</i>	<i>I</i>	α_{95}	<i>k</i> ° C		<i>N</i>	<i>D</i>	<i>I</i>	α_{95}	<i>k</i>	° N	° E
26	22	328	+28	9.3	12	450	23	(324.8	+24.5)	8.6	13.4	(54.8	291)
29	22	358	+38	10.2	10	450	21	3.4	+32.0	6.9	22.4	82.6	196
31	19	7	+42	5.6	38	400	19	0.2	+37.1	3.6	90	86.3	211
32	26	184	-20	6.9	18	350	29	178.6	-20.1	6.5	17.7	75.9	220
33	13	0	+44	11.1	15	300	13	350.1	+41.7	9.4	20.4	81.0	303
35	5	339	+40	22	13	500	8	357.2	+28.9	12.8	19.7	80.6	231
No consistent results for sites 14, 15, 28, 30, 34													
Mean of site means (4 normal, 1 reversed)							5	358.1	+32.0	8.7	78.4	82.9	231.4
											$A_{95}=6.4$	$K=143$	
Mean of site means (this paper plus Schult et al. 1978; 10 normal, 13 reversed)							23				81.8	222.7	
											$A_{95}=3.3$	$K=84$	

Legend see Table 1

Polar Wander Path for Africa in the Mesozoic

Recent attempts to define the polar wander path for Africa for the Late Mesozoic (Hargraves and Onstott 1980; Hussain et al. 1980) implied relatively great mobility of the pole whereas in earlier papers it was often believed that in the Mesozoic the African palaeopoles were closely grouped. In Table 4 Mesozoic poles for Africa and South America are compiled. The same labels have been used for poles with approximately the same ages. The poles of these two continents may be compared with the aid of reconstructions of continental drift. The African poles and the proposed polar wander path are shown in Fig. 6a confirming the results of Hargraves and Onstott (1980) and Hus-

sain et al. (1980), with some modifications. The "rotated" South American pole positions (together with the African polar wander path) are shown in Fig. 6b and listed with the rotation parameters in Table 4. It has been shown that for the Mesozoic a better coincidence of the South American and African poles can be achieved by not using the pre-drift reconstruction from Bullard et al. (1965) but by assuming a small separation of the continents from the so-called pre-drift position in the Southern Atlantic in Early and Middle Mesozoic time (Schult and Guerreiro 1979). This seems also to be valid for the Late Palaeozoic (Vilas and Valencio 1977). Therefore for the Triassic through Lower Cretaceous a reconstruction was chosen as proposed originally for 110 m.y. *B.P.*, about 15 m.y. after the assumed begin-

Table 3. Site mean palaeomagnetic results from Baharia Oasis (28.2° N, 28.9° E) and Qatrani basalt (29.8° N, 30.7° E)

Site	NRM					AF(Oe)	CARM					VGP	
	<i>N</i>	<i>D</i>	<i>I</i>	α_{95}	<i>k</i>		<i>N</i>	<i>D</i>	<i>I</i>	α_{95}	<i>k</i>	° N	° E
a) Baharia Oasis: basalts													
2	19	200.3	+61.1	12.6	7.9	200	18	174.0	- 3.1	9.1	15.1	62.9	222
3	25	213.5	+67.5	5.2	23.5	250	25	192.6	+ 5.2	6.4	21.4	56.6	185
4	No consistent result												
5	3	171.2	+63.2	20.7	36.6	250	4	176.1	- 7.9	5.9	244	65.3	216
6	11	212.6	+54.6	12.5	14.1	150	10	(210.0)	+41.1)	15.7	10.3	(30.6)	177)
Mean of site means							3	181.2	- 2.0	18.4	46	62.6	206.0
											$A_{95}=16.3$	$K=58$	
b) Baharia Oasis: iron ores													
7	31	335.6	+48.2	8.5	10.0	250	27	349.6	+48.9	6.0	22.4	80.8	310
8	36	192.5	-51.0	3.6	43							78.3	95
Mean of site means (this paper plus Schult et al. 1978; 3 normal, 6 reversed)							9	188.0	-43.6	6.4	64.8	83.5	138.6
											$A_{95}=7.0$	$K=55$	
c) Qatrani basalt													
1	38	201.4	-57.0	1.7	170	150	41	209.2	-60.2	1.0	472	63.8	87
Mean of site means (this paper plus Hussain et al. 1976; all reversed)							3	197.0	-59.1	8.7	202	72.7	81
											$A_{95}=12.7$	$K=95$	

Legend see Table 1

Table 4. Mesozoic palaeomagnetic poles for Africa and South America

Pole	Africa Formation	Age m.y.	Pole Position			South America Formation	Age m.y.	Pole Position			Rotated ^a	
			° N	° E	A_{95}			° N	° E	A_{95}	° N	° E
1	Triassic mean		69	263	4.9	Triassic mean		78	56	6.3	70	247
2	Stormberg, Karroo	154-190	65	262	11.7	Chon-Aike	157-173	87	17	6	63	261
3	Hoachanas	161-173	62	252	7	Maranhão	158±12	85	83	6.9	62	250
4	Mateke Hills	~168	59	260	8.3							
5	Moroccan volcanics	K1	44	251	10	R. de los Molinos	129-150	78	193	8	48	249
						Almafuerte	123±4	72	205	6	41	251
						C. Colorado	121±3	83	196	10	53	252
6	Kimberlite pipes	122-162	36	277	17							
7	Kaoko lavas	110-128	48	267	2	Serra Geral	115-130	78	234	5.7	47	261
						Serra Geral	119±5	85	295	3.7	57	266
						Maranhão	118±6	84	261	1.9	54	264
8	Mlanje Massif	116-128	60	262	12	Rumipalla	<121	88	326	9	60	261
9	Lupata volcanics	106-111	62	260	3.5	La Serena	~110	81	29	4.5	68	261
10	Wadi Natash volc.	77-100	69	258	5.8	Cabo de S. Agostinho	85-99	88	135	4.5	68	247 ^b
11	Kimberlite pipes	82-88	61	224	7.4							
12	Volcanics Sicily	71-81	62	229	3.8							
13	Moroccan sediments	K1-u	75	217	5.5							
14	Nubian sandstone	Ku	82	230	3.4	Poços de Caldas	63-80	81	53	10	85	211 ^c
15	Red Siltstone	Ku	79	208	6							
16	Tororo	Ku(?)	76	195	9							

Rotation parameters for South America from present with respect to Africa (Sclater et al. 1977):

^a 48.3° about a pole at 49.2° N, 31.8° W (originally estimated for a reconstruction for 110 m.y. *B.P.*, here tentatively taken as "pre-drift" position; see text)^b 40.1° about a pole at 56.6° N, 34.1° W, estimated for 95 m.y. *B.P.*^c 32.8° about a pole at 67.3° N, 39.5° W, estimated for 80 m.y. *B.P.*

References for African poles: 2 Average of three Lower Jurassic poles after McElhinny (1973); 6, 11 Hargraves and Onstott (1980); 10, 14 this paper; other poles see Schult and Guerreiro (1979). South American poles: 9 Palmer et al. (1980); 10 Schult and Guerreiro (1980); other poles see Schult and Guerreiro (1979)

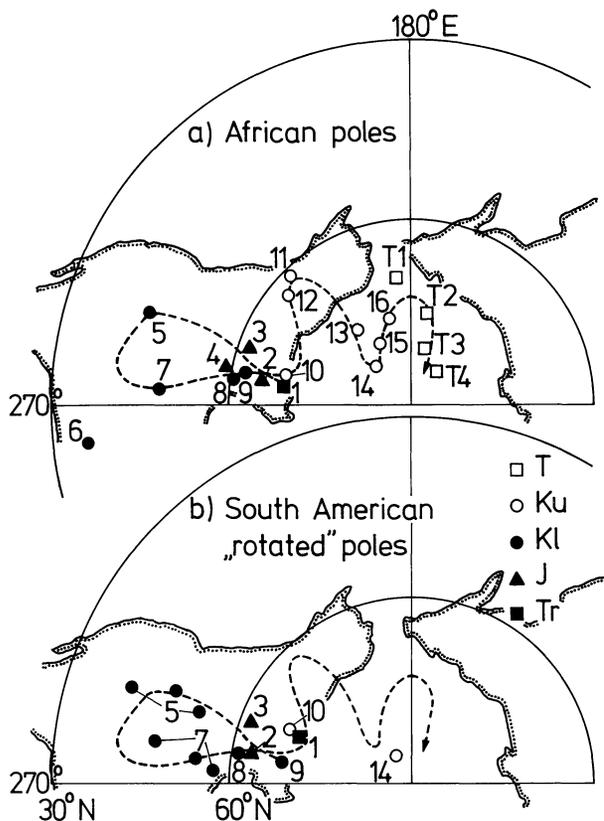


Fig. 6. **a** Triassic mean, Jurassic, Cretaceous and some Palaeocene poles for Africa and proposed apparent polar wander path. Actually the *dotted line* indicates the trend of pole positions and the "path" is of variable width depending on the α_{95} values of the poles. Labels 1–16 refer to Table 4. Tertiary poles: T1 Upper Eocene volcanics in Northern Egypt (Hussain et al. 1979); T2 and T3 Oligocene Ethiopian South-Eastern (Schult 1974) and Western (Brock et al. 1970) plateau basalts; T4 Eocene Baharia iron ores (Table 3). **b** "Rotated" South American poles using reconstructions of continental drift. Labels and "rotation" parameters see Table 4. Same labels in **a** and **b** refer to poles with approximately same ages. The indicated polar wander path is reproduced from **a**

ning of the rifting (Sclater et al. 1977). For the Upper Cretaceous the respective reconstructions for 95 and 80 m.y. *B.P.* after Sclater et al. (1977) were used.

Most of the African poles cluster around 65° N and 260° E but a few poles (5, 6, and 7 in Fig. 6a) indicate a loop of the polar wander path to lower latitudes in the Lower Cretaceous. Pole 7 (Kaoko lavas) seems to be reliable but unfortunately pole 6 (for Kimberlite pipes) was derived from only 4 sites with a relative large circle of confidence and pole 5 (Moroccan volcanics) is from rocks located in the tectonically active Atlas area. However, the loop to low latitudes is confirmed by several Lower Cretaceous "rotated" South American poles (Fig. 6b). There is also agreement of the pole positions of both continents at the beginning (Triassic/Jurassic) and at the end of the loop (early Late Cretaceous). Particularly the pole for the Wadi Natash volcanics (10) is near the "rotated" pole for Cabo de Santo Agostinho (10) of the same age (Schult and Guerreiro 1980), assuming a reconstruction for 95 m.y. *B.P.*

For late Upper Cretaceous times poles 11 and 12 imply another loop. Pole 12 is from South East Sicily which may be considered as part of the African plate (Schult 1973). For pole 11 (Kimberlite pipes) previous data (McFadden and Jones 1977)

were recently improved by Hargraves and Onstott (1980). This loop is not recorded by the presently available scarce South American data for the Upper Cretaceous. The pole for the Nubian sandstone (14) is consistent with other high latitude Upper Cretaceous poles (13–16) for Africa and also consistent with the "rotated" pole 14 from South America (Schult and Guerreiro 1980) for a reconstruction for 80 m.y. *B.P.*

The location of selected Eocene and Oligocene African poles (Fig. 6a) may be interpreted as another loop of the polar wander path before it approaches the present geographic pole in Late Tertiary. Relative low latitude pole positions (e.g. Table 3) indicate further excursions in Tertiary time. However no definitive conclusions can be drawn from the data presently available.

The proposed African polar wander curve for the Mesozoic is different in detail from that transferred from North American data (Irving 1977) by applying the rotation poles for the opening of the North Atlantic according to Sclater et al. (1977). Similar differences exist when comparing with other indirectly obtained polar wander curves (Van der Voo and French 1974). There is agreement only in Late Triassic and Jurassic. The North American polar wander path is much better defined than the African one and therefore further confirmations of the African Mesozoic data are necessary. On the other hand large differences also exist for the Late Palaeozoic (e.g. Van den Berg 1979) and therefore reconstructions of drift of continents adjacent to the North Atlantic other than those usually used may be considered.

Acknowledgements. We thank Professor Dr. G. Angenheister and Professor Dr. M. Fahim for their advice and help. We also thank H. Winhard for carrying out part of the measurements. The research has been done within the bilateral agreement between the Deutsche Forschungsgemeinschaft (DFG), the Gesellschaft für Technische Zusammenarbeit (GTZ), and the Academy of Scientific Research and Technology (ASRT) of Egypt. The financial support of the DFG and GTZ is gratefully acknowledged.

References

- Brock, A.: An experimental study of palaeosecular variation. *Geophys. J. R. Astron. Soc.* **24**, 303–317, 1971
- Brock, A., Gibson, I.L., Gacii, P.: The palaeomagnetism of the Ethiopian flood basalt succession near Addis Ababa. *Geophys. J. R. Astron. Soc.* **19**, 485–497, 1970
- Bullard, E.C., Everett, J.E., Smith, A.G.: The fit of the continents around the Atlantic. *Philos. Trans. R. Soc. A*, **258**, 41–75, 1965
- El Akkad, S., Issawi, B.: Geology and iron ore deposits of the Bahariya Oasis. *Geol. Surv., Cairo, Paper No. 18*, 1963
- El Ramly, M.F.: A new geological map for the basement rocks in the Eastern and south Western Deserts of Egypt, scale 1:1,000,000. *Ann. Geol. Surv., Cairo, vol. 2*, 1–18, 1972
- El Shazly, E.M.: The geology of the Egyptian region.: In: *The Ocean Basins and Margins, Vol. 4A*, A.E.M. Nairn, W.H. Kanes, and F.G. Stehli, eds., pp. 379–444. New York: Plenum Press 1977
- El Shazly, E.M., Krs, M.: Paleogeography and paleomagnetism of the Nubian sandstone, Eastern Desert of Egypt. *Geol. Rundsch.* **62**, 212–225, 1973
- Hargraves, R.B., Onstott, T.C.: Paleomagnetic results from some Southern African kimberlites, and their tectonic significance. *J. Geophys. Res.* **85**, 3587–3596, 1980
- Hussain, A.G., Schult, A., Soffel, H., Fahim, M.: Magnetization and palaeomagnetism of Abu Zaabal and Abu Rawash basalts (Egypt). *Helwan Obs. Bull.* **134**, 1–15, 1976
- Hussain, A.G., Schult, A., Soffel, H.: Palaeomagnetism of the basalts of Wadi Abu Tereifiya, Mandisha and dioritic dykes of Wadi Abu Shihah, Egypt. *Geophys. J. R. Astron. Soc.* **56**, 55–61, 1979
- Hussain, A.G., Schult, A., Soffel, H.: Preliminary polar wander path of Egypt between Cretaceous and Quaternary. *Berliner Geowissenschaftliche Abhandlungen A/19*, 93–94, 1980

- Irving, E.: Drift of the major continental blocks since the Devonian. *Nature* **270**, 304–309, 1977
- Lowrie, W., Channell, J.E.T., Alvarez, W.: A review of magnetic stratigraphy investigations in Cretaceous pelagic carbonate rocks. *J. Geophys. Res.* **85**, 3597–3605, 1980
- McElhinny, M.W.: *Palaeomagnetism and plate tectonics*. London: Cambridge University Press 1973
- McFadden, P.L., Jones, D.L.: The palaeomagnetism of some upper Cretaceous kimberlite occurrences in South Africa. *Earth Planet. Sci. Lett.* **34**, 125–135, 1977
- Palmer, H.C., Hayatsu, A., MacDonald, W.D.: Palaeomagnetic and K-Ar age studies of a 6 km-thick Cretaceous section from the Chilean Andes. *Geophys. J. R. Astron. Soc.* **62**, 133–153, 1980
- Sabet, A.H., Tsogoev, V.B., Bessonenko, V.V., Babourin, L.M.: Geological map of the central Eastern Desert of Egypt. Cairo: The Geological Survey of Egypt and Mining Authority 1973
- Said, R.: Explanatory notes to accompany the geological map of Egypt. The Geological Survey of Egypt, Paper No. 56, 1971
- Schult, A.: Palaeomagnetism of Upper Cretaceous volcanic rocks in Sicily. *Earth Planet. Sci. Lett.* **19**, 97–100, 1973
- Schult, A.: Palaeomagnetism of Tertiary volcanic rocks from the Ethiopian Southern Plateau and the Danakil block. *J. Geophys.* **40**, 203–212, 1974
- Schult, A., Soffel, H.C., Hussain, A.G.: Palaeomagnetism of Cretaceous Nubian sandstone. *J. Geophys.* **44**, 333–340, 1978
- Schult, A., Guerreiro, S.D.C.: Palaeomagnetism of Mesozoic igneous rocks from the Maranhao Basin, Brazil, and the time of opening of the South Atlantic. *Earth Planet. Sci. Lett.* **42**, 427–436, 1979
- Schult, A., Guerreiro, S.D.C.: Palaeomagnetism of Upper Cretaceous volcanic rocks from Cabo de Sto. Agostinho, Brazil. *Earth Planet. Sci. Lett.* **50**, 311–315, 1980
- Sclater, J.G., Hellinger, S., Tapscott, C.: The paleobathymetry of the Atlantic Ocean from the Jurassic to the present. *J. Geol.* **85**, 509–552, 1977
- Van den Berg, J.: Paleomagnetism and the changing configuration of the western Mediterranean area in the Mesozoic and Early Cenozoic eras. *Geol. Ultraiectina* No. **20**, 1979
- Van der Voo, R., French, R.B.: Apparent polar wandering for the Atlantic-bordering continents: Late Carboniferous to Eocene. *Earth Sci. Rev.* **10**, 99–119, 1974
- Vilas, J.F., Valencio, D.A.: Palaeomagnetism of South American rocks and the Gondwana continent. IV. International Gondwana Symp., Calcutta, Papers Vol. **II**, 923–930, 1977

Received May 18, 1981, Revised Version July 17, 1981

Accepted July 20, 1981

Geomagnetic Induction Studies in Scandinavia

II. Geomagnetic Depth Sounding, Induction Vectors and Coast-Effect

Alan G. Jones

Institut für Geophysik der Universität Münster,
Gievenbecker Weg 61, D-4400 Münster, Federal Republic of Germany

Abstract. In this paper an event of very favourable structure for induction purposes, which was observed by the Scandinavian magnetometer array, is discussed and analysed in detail. The responses derived, in both the time and frequency domains, display a large coast effect at all coastal stations, both on the sea coasts and on the Gulf of Bothnia. Two relatively large inland anomalies are also delineated. The more significant of the two, the Storavan anomaly, is apparent using all the analysis techniques employed, and may be associated with a remnant of the Svionian island arc system. The less dominant anomaly, in the vicinity of Mieron, has no obvious correlation with geology or tectonic formations, and appears to be polarisation sensitive – it is only energised by east-west magnetic fields.

First approximation modelling of the coast effect observed by the northwestern stations illustrated that the responses are well satisfied by the conductivity contrast between sea and land.

Finally, the validity of the derived induction vectors, and the possible effects of induction for source field studies, are discussed.

Key words: Magnetometer arrays – Geomagnetic induction studies in Scandinavia – Coast effect – Geomagnetic depth sounding

Introduction

In this work, the second of a series of papers (the first being Jones (1980), hereafter referred to as Paper I) concerned with various aspects of geomagnetic induction in Scandinavia as observed by the Münster IMS magnetometer array (Küppers et al. 1979), some qualitative aspects of the induced field, and a first approximation modelling of the coast effect, will be described.

Magnetometer array studies for induction purposes, using large numbers of variometers, mostly based on the original Gough-Reitzel design (Gough and Reitzel 1967), have been very successful in mapping the flow of anomalous electric current associated with lateral variations of electrical conductivity, either in the crust, or the upper mantle, or both. Reviews of some array studies, and techniques for analysing and presenting the data, are to be found in, for example, Gough (1973a, b), Frazer (1974) and Lilley (1975).

In almost all previous 2D magnetometer array studies, with the notable exception of that of Bannister and Gough (1977), the feature of prime interest was the Earth's internal

conductivity structure. Thus, the magnetometers were typically 10–20 km apart, both to detect small scale anomalies and to avoid spatial aliasing problems. For the Scandinavian array however, the main purpose was the detection and analysis of local magnetic disturbance fields and their interpretation in terms of possible ionospheric current structure. Hence, the station spacing was of a minimum of some 80 km, and was typically 120 km, in northern Scandinavia, and of 200–300 km in southern Scandinavia. Such a spacing restricts an induction study utilising the data to only the gross effects observable.

In this paper an event, exhibiting an unusually uniform horizontal magnetic field, is described, and the data are analysed qualitatively, by various techniques, both in the time and frequency domains. The resulting information details two rather strong inland conductivity anomalies, and describes enhancements of the vertical magnetic field both at the sea coast and at the Gulf of Bothnia. The coast effect, as observed by the northwestern stations, is modelled by a thin sheet approximation, and it is shown that the data do not require a lateral variation in the crust and mantle between the ocean and the continent, although such a variation cannot be excluded.

Data

The observations to be discussed in this paper were recorded by the Münster IMS Scandinavian Magnetometer Array (reported in Küppers et al. 1979) of modified Gough-Reitzel magnetometers (Gough and Reitzel 1967; Küppers and Post 1981) at the 24 locations shown in Fig. 1. All pertinent information regarding these stations is to be found in Table 1 of Küppers et al. (1979).

The particular interval chosen for this study was 13:00–16:00 UT on 6 July 1977, because of the somewhat unique characteristics of the magnetic field during this time. The three-component magnetic data from the 24 locations shown in Fig. 1 were digitised, reduced to variations relative to the quiet-time base line (taken as the field level during the interval 03:30–04:30 UT, also on 6 July 1977), and rotated into the Kiruna Cartesian coordinate system. This gave magnetic components, here labelled $x(t)$ (geomagnetic “north”), $y(t)$ (geomagnetic “east”) and $z(t)$ (vertical, positive downwards) (note: in other papers dealing with data from this array the labels A, B, Z have been employed to denote north, east, and vertical respectively), with a temporal resolution of 10 s, and a typical magnetic variation resolution of 2 nT (Küppers et al.

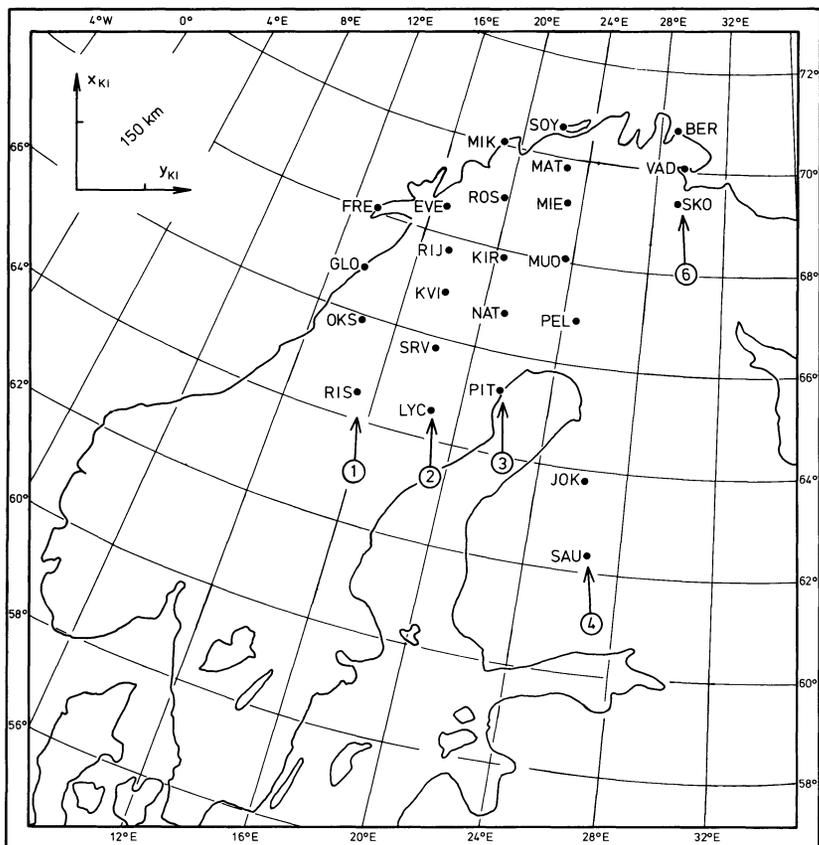


Fig. 1. Map showing the locations, in geographical co-ordinates, of the magnetometer stations which provided data for the event discussed in this study. In the upper left-hand corner are indicated the axes of the Kiruna cartesian co-ordinate system (see text). Profiles 1–4 and 6, which are perpendicular to lines of constant revised corrected geomagnetic latitude, are designated by the encircled numbers

1979, see also Jones, in press 1981a). The definitions of “north” and “east” for a particular location are given in Küppers et al. (1979). These data were chosen for analysis for induction purposes because of the large degree of horizontal field spatial uniformity that existed during the interval. The data are illustrated in Fig. 2 as profiles 1, 2, 3, 4, and 6 (Fig. 1) of the x , y , and z components. The approximate spatial uniformity exhibited, which is seen in the horizontal components, is a very uncommon feature of magnetic data observed at these high geomagnetic latitudes.

Information in the Time Domain

A simple inspection of the magnetograms recorded by an array of instruments often reveals the gross features of the lateral variations in conductivity. In particular, the component most indicative of internal conductivity inhomogeneities is the vertical magnetic field z . Large features are known to produce phase shifts, and even phase reversals, in z between two points spanning an inhomogeneity. An excellent example of this simple method for mapping a large 2D feature is given by Alabi et al.’s (1975) work on the North American Central Plains (NACP) anomaly (see their Fig. 3).

No such gross effect as the NACP anomaly is observed in northern Scandinavia, but the data illustrated in Fig. 2 display the following characteristics:

(i) very strong attenuation of the high-frequency components of the z -field with distance away from the coast (Fig. 2), particularly for profiles 3, 4, and 6,
 (ii) attenuation with decreasing latitude of the low-frequency components of the z -field,

(iii) re-emergence of high-frequency components in the z -field at stations SRV (profile 2) and PIT (profile 3, on the Gulf of Bothnia),

(iv) phase reversal in the high-frequency z -components observed along profile 3, with undetectably small high-frequency content at KIR,

(v) no phase shift observed over the whole array associated with the peak in the x -component at 14:00:00 UT,

(vi) strong increase in the y -component observed at OKS (profile 1) when compared with the neighbouring stations of GLO and RIS,

(vii) marked increase in the y -component at LYC (profile 2),

(viii) general attenuation in the y -component moving from the north-east (e.g., BER) towards the south-west,

(ix) attenuation in the high-frequency content in the y -component with distance away from station MUO (profile 4).

Because the data were reduced to variations relative to a quiet-time value (see above), and not to an arbitrary level, it is possible to derive the equivalent external current density for a particular instant of time. Figure 3a illustrates the approximate equivalent external current density on the ground, observed at 14:00:00 UT (corresponding to the peak in the x -component) over northern Scandinavia. The equivalent external current density, displayed as vectors at each station, was derived by taking the instantaneous magnetic disturbance vectors at 14:00:00 UT and estimating that the internal/external field ratio, given by $(1 - kC)/(1 + kC)$ (Schmucker 1970; 1973) with k (source wavenumber) $\approx 1/2000$ km (see below) and C (inductive response function) ≈ 130 km (see Paper I, Table 1, period of 890 s), was ≈ 0.9 . The thus estimated exter-

nal magnetic field on the ground was interpreted in terms of current density by assuming that a homogeneous current sheet of infinite extent flowed directly above the instrument, i.e., $(j_e)_y = 2(x_e)/\mu_0$. The magnetic field situation at 14:00:00 UT is representative of the general situation existing during the two "bursts" of activity, i.e., 13:40–14:20 UT and 15:15–15:45 UT. The high degree of spatial uniformity in the horizontal disturbance field is shown clearly. The corresponding vertical disturbance field, as contoured in Fig. 3b, as well as precisely describing a coast effect, also infers an inhomogeneity in the PIT-SRV-LYC region. It should be stressed here that any anomalous features observed in the vertical field must have their counterparts in the horizontal fields, hence *total* spatial uniformity of the horizontal fields is not possible. (Anomalies are however generally more difficult to detect in the horizontal components because, for a totally uniform field over a 1D flat earth, the internally induced and externally inducing fields are of equal magnitude, but are opposite in phase in the vertical component, and in-phase in the horizontal components, leading to $z_t = 0$, $x_t = 2x_e$ and $y_t = 2y_e$, where subscript *t* denotes total field components.) Latitude profiles in a co-ordinate system 20° anti-clockwise to the Kiruna system (corresponding to the direction shown by the KIR equivalent current vector), exhibit a strikingly longitudinally uniform N20°W magnetic field with a latitudinal attenuation of $-0.065 \text{ nT km}^{-1}$ (Fig. 4a). The horizontal magnetic component resolved in the direction N70°E exhibits a longitudinal attenuation of approximately -0.11 nT km^{-1} (Fig. 4b), but some non-uniform effects are apparent. These values of the spatial gradients of the horizontal fields, and the estimate of the inductive response function $C \approx 130 \text{ km}$, give an estimate of the normal vertical field, from

$$Z_n = C \left(\frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} \right)$$

(see, for example, Schmucker 1970; Jones 1980) of $z_n \approx -25 \text{ nT}$. The instantaneous vertical field exhibits this value in central northern Scandinavia, but at the northern coastal stations and towards the south-west and south strong non-1D earth effects are apparent.

It must be stressed that such a comparison between the horizontal and vertical fields at a particular instant of time may be used only in a qualitative manner and conclusions must be expressed tentatively. This is because the observed instantaneous fields are only related in a strict sense if the induction is purely in-phase. Out-of-phase components lead to time delays between the inducing and induced parts (Schmucker 1980; Jones in press 1981b).

The features described in this section can be split into two parts, those arising from the source-field, and those arising from the induced fields. Features (ii), (v), (viii) and Fig. 4a, b are all illustrative of the non-uniformity of the source field. The rest show effects of internal contributions which are certainly not insignificant.

Information in the Frequency Domain

The spectral content of the event is illustrated by the sonogram analysis of the data from stations KIR (Fig. 5) and BER (Fig. 6). The data were recursion filtered by a narrow band-pass, with a selectivity of 0.3 (Hermance 1973), at centre periods of B1=60 s, B2=100 s, B3=200 s, B4=300 s, B5=450 s, B6=600 s and B7=1,000 s. The data exhibit the features that (i) there is no information below 100 s, (ii) at 100 s

there are only data in the first interval of activity, i.e., 13:30–14:30, (iii) the polarisation characteristics of the first and second intervals of activity (i.e., 13:30–14:30 and 15:00–16:00) are different at all periods, and (iv) the approximate spatial uniformity of the horizontal field is observed over the whole frequency range. The latter two points are especially important when estimates of response functions are to be determined, as will be discussed later.

Estimates of the smoothed auto-spectral densities (or "power spectra") for each component, and of the cross-spectral densities between pairs of components, were derived by the usual techniques of statistical frequency analysis. The steps involved were:

- (i) extraction of all data in the interval 13:01:00–15:51:40 UT to give a reduced data length of 1.024 points per component,
- (ii) removal from each component of zero, first and second order polynomial trends,
- (iii) application of a cosine taper to the first and last 10% of each component data series,
- (iv) computation of the raw Fourier spectra, of 511 complex and 2 real harmonics, of each component,
- (v) first-order correction of the instrument response, as illustrated in Küppers and Post (1981), by multiplying each harmonic with the term

$$1 + i \frac{\omega_c}{\omega}$$

where ω_c is the -3 dB point of the component response curve. The response of a typical instrument fitted reasonably well to a first-order low-pass Butterworth filter with -3 dB points given by: *x*-component = 9.5 s; *y*-component = 13.0 s; and *z*-component = 5.5 s,

- (vi) computation of raw auto and cross spectra,
- (vii) smoothing of raw auto and cross-spectra by a constant $-Q$ box-car window, with $Q=0.3$ ($Q = \Delta\omega/\omega_0$),
- (viii) normalising the smoothed spectra by data set length, and by the factor 1/0.875 to correct for the application of the cosine taper (Bendat and Piersol 1971), to yield $\hat{S}_{ab}(\bar{\omega})$, the estimated cross spectra between components *a*(*t*) and *b*(*t*).

Fourier Spectra Maps

Contouring of Fourier spectra at a certain period has proved in past array studies to be a useful technique for delineating conductivity anomalies (see de Beer and Gough 1980 and references therein). Indeed, in a recent work Gough and de Beer (1980) consider that Fourier spectra maps are superior to induction vector information when the horizontal components are strongly interdependent (see discussion on this topic in later section).

For this event of 3 h duration, the longest period at which stable spectral estimates can be made is of the order of 1,000 s, due to spectral frequency smoothing considerations which are necessary in order to reduce the associated variance of the estimates. The maps chosen for illustration are those at 1,000 s, for reasons detailed below, but they are representative of the range 200 s–1,000 s.

It was found that the contoured maps of the real and imaginary parts of the smoothed Fourier spectra gave more detail than did the more usual maps of amplitude and phase. Figure 7a–f show the real and imaginary parts of the three

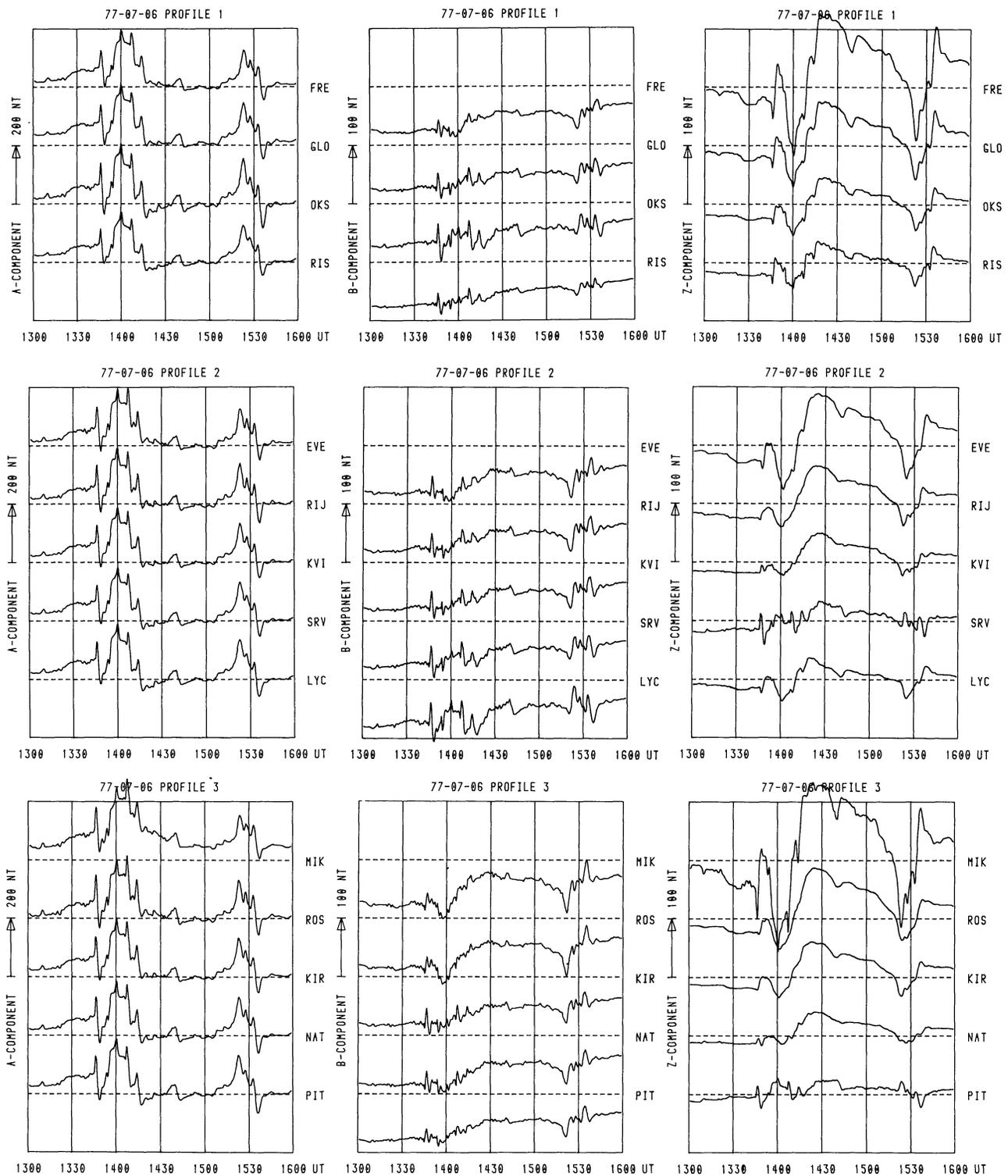
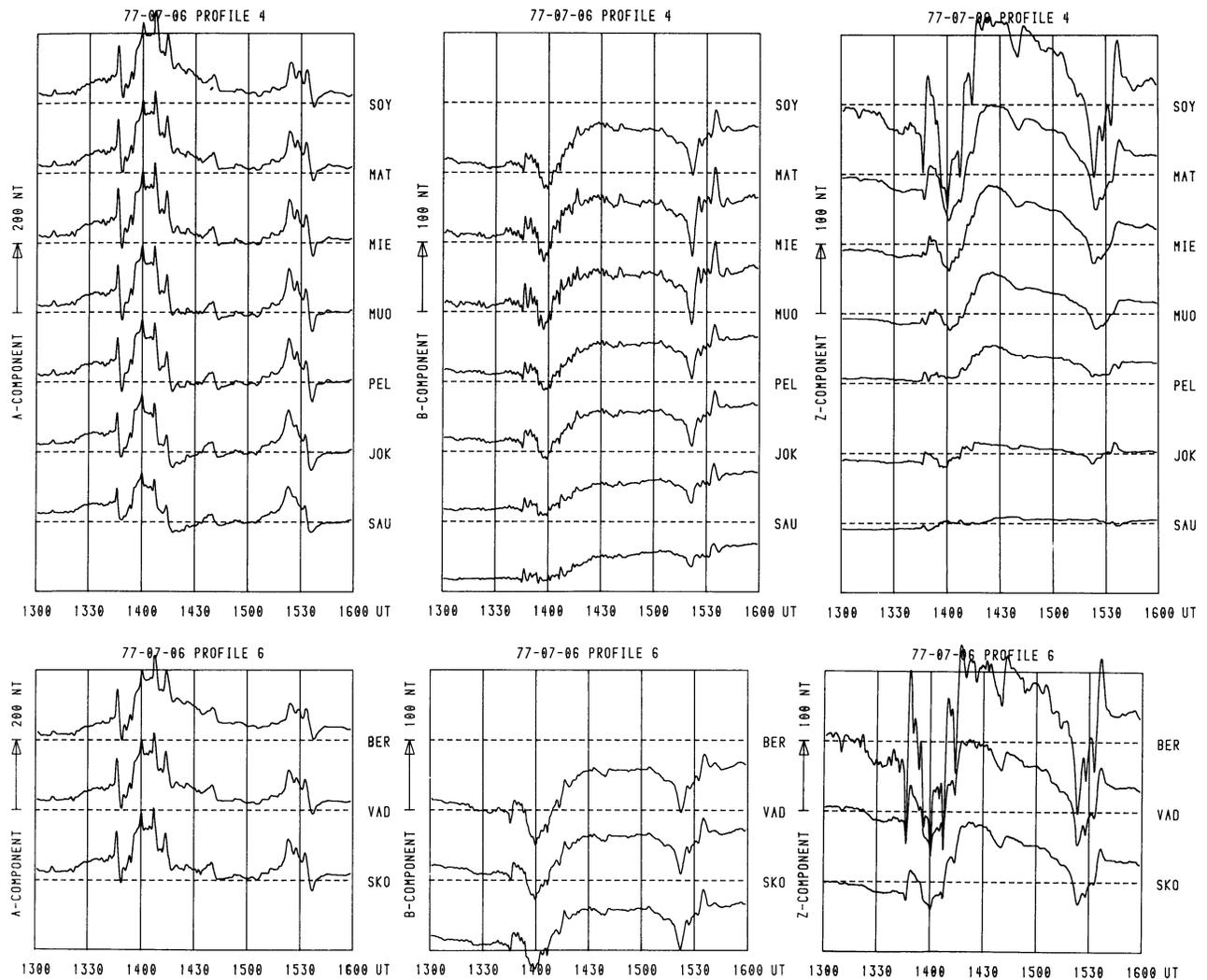


Fig. 2. Magnetic disturbances, as recorded by the stations illustrated in Fig. 1, in the interval 13:00–16:00 UT on 6 July 1977, compared with quiet-time baselines (*dotted lines*). A, B, and Z components refer to north, east, and vertical respectively in the Kiruna co-ordinate system (note: these components are termed $x(t)$, $y(t)$ and $z(t)$ in the text). The data are stacked along Profiles; *Profile 1* FRE→RIS; *Profile 2* EVE→LYC; *Profile 3* MIK→PIT; *Profile 4* SOY→SAU; *Profile 6* BER→SKO

components at 1,000 s period. The polarisation characteristics are shown in Fig. 9a and correspond to an average field over the event as observed at KIR with a major axis pointing due north, an ellipticity of 0.12, and a ratio of polarised power to total power of 0.98. That the event displays an almost linear

polarisation is regarded as advantageous for Fourier spectra mapping from the conclusions drawn by Gough and de Beer (1980).

Figure 7a–f confirm the major details already seen in the data as set out above, namely:



point (iii), the Storavan anomaly ($\text{Im}(Y)$, $\text{Re}(Z)$ and $\text{Im}(Z)$), point (ix), the Mieron anomaly ($\text{Re}(Y)$, $\text{Im}(Y)$ and $\text{Re}(Z)$), and point (i) a strong coast effect ($\text{Im}(Z)$).

Amplitude and phase maps (not shown here) confirm the comments by de Beer and Gough (1980) that the intracontinental anomalies are best seen in the Z phase, whilst the coast effect shows in Z amplitude.

Polarisation Characteristics of the Horizontal Field

Any two-component vector field may theoretically be decomposed into three parts: (i) totally polarised signal, (ii) totally unpolarised signal, and (iii) noise contributions (Jones 1979). However, without an *a priori* knowledge of the noise terms, such a decomposition is not possible and the expressions detailed in Fowler et al. (1967) may lead to biased estimates of the polarisation parameters (Jones 1979).

Figure 8a, b illustrate the auto-spectral densities of the three-magnetic components for stations KIR (Fig. 8a) and BER (Fig. 8b). The greatest noise contribution is believed to be the digitising error (Jones in press 1981b) of the order of 2 nT, the level of which is indicated by the broken line at $15 \text{ nT}^2 \text{ Hz}^{-1}$, i.e., the power spectral density level of a 1068 point random series, of variance 2.75 nT, extended by zeroes

to 2048 points. As described by Fig. 8a, b, the power spectra exhibit a very strong attenuation with decreasing period, with the result that the signal-to-noise ratio of the horizontal components only becomes insignificantly small at periods less than about 100 s. Hence the polarisation parameter estimates will not be significantly biased at periods greater than 100 s.

The polarisation parameters were derived from the smoothed auto- and cross-spectra by the expressions given in Born and Wolf (1964) (repeated in Fowler et al. 1967 and Jones 1979). The polarised parts of the fields observed at each location are illustrated as ellipses at four periods; 1,000 s, 450 s, 200 s and 100 s in Fig. 9a-d respectively. The ratio of polarised power to total power was mostly in the range 0.90–0.95 at all stations for periods greater than 100 s. In the polarisation maps, the strong uniformity of the horizontal field is again evident, but other features can also be seen. The general decrease in ellipticity at 1,000 s and 450 s going from the NE to SW is a function of the decreasing Y field in this direction, and hence is a source effect. However, the following are not explainable by source contributions:

(i) the increase in polarised power with distance away from the coast along profiles 2, 3, and 4 at all periods,

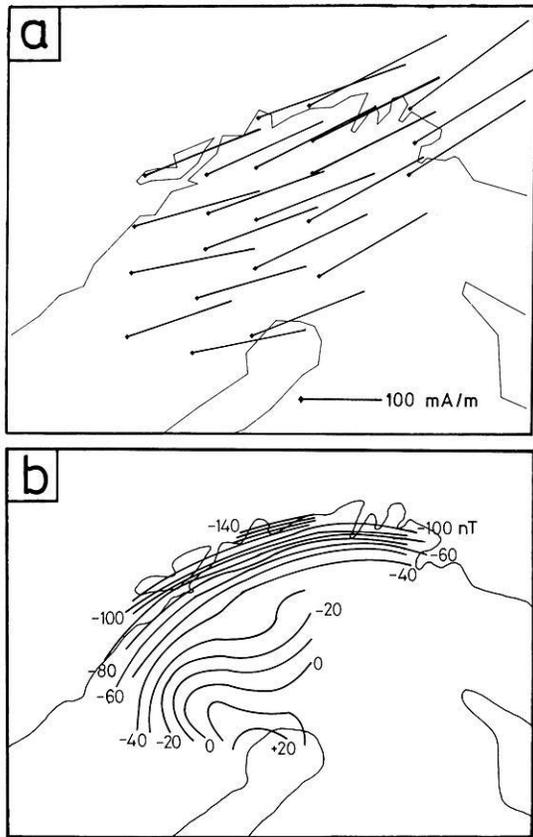


Fig. 3. **a** Upper panel: Equivalent external current density, as observed by the magnetometer array, at 14:00:00 UT on 7th July 1977; **b** Lower panel: the vertical magnetic disturbance field on the ground for the same time instance as Fig. 3a

- (ii) the tendency for the ellipse observed at LYC to be orientated more to the east than at SRV and KVI, also on profile 2,
- (iii) the small polarised power observed at RIS at all periods compared with neighbouring stations.

Induction Vector Information

The single-station transfer functions $[T_x, T_y]$ are defined as those functions that minimise, in some sense, the error of the fit to the equation

$$Z = T_x X + T_y Y + \zeta \quad (1)$$

(dependence on frequency assumed throughout), which relates the vertical magnetic field component to the horizontal magnetic field components at the same frequency. It is obvious that for any fields whatsoever transfer functions $[T_x, T_y]$ may be estimated. What is of paramount significance is whether the derived functions may be interpreted in terms of induction effects. The assumptions under which such an interpretation is valid are:

- (i) $Z \sim Z_n$, i.e., the vertical field contains approximately no normal part,
- (ii) $E[S_{z_n x}] = E[S_{z_n y}] = 0$ (where E denotes expectation value), i.e., the horizontal fields do not correlate with any normal vertical field,
- (iii) $X \sim X_n$, $Y \sim Y_n$, i.e., the horizontal fields can be approximated to be normal.

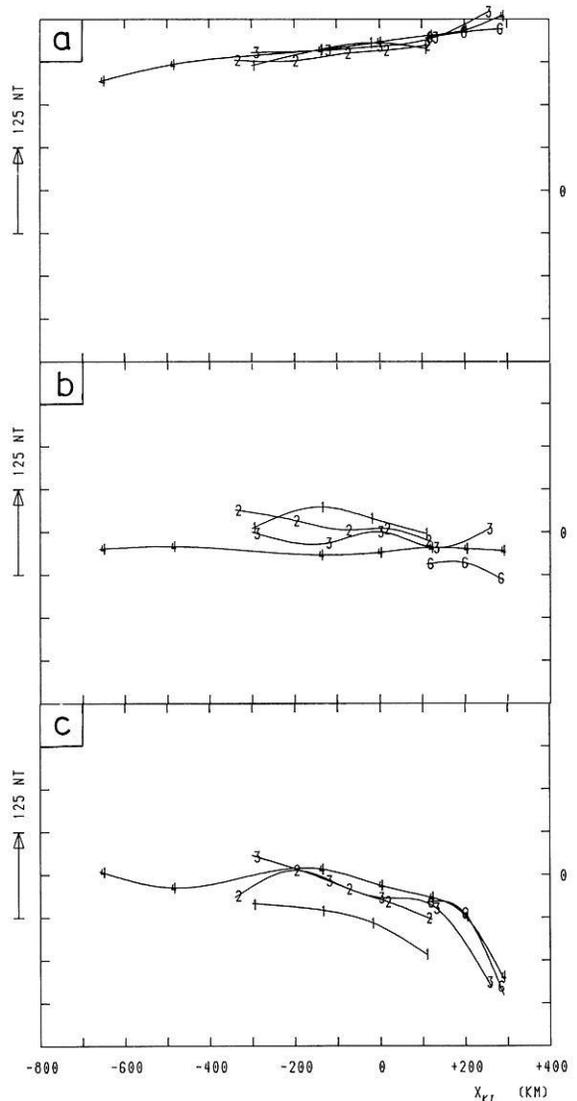


Fig. 4a-c. Latitudinal profiles at 14:00:00 UT on 7 July 1977 in the rotated co-ordinate system, given by a 20° anti-clockwise rotation of the Kiruna co-ordinate system, i.e., such that the magnetic disturbance field observed at Kiruna (KIR) (see Fig. 3a) has no eastward or westward directed component in the rotated system. **a** Upper panel; northward directed magnetic disturbance field along profiles 1 to 4 and 6, as denoted by the numbers; **b** Middle panel; eastward directed magnetic disturbance field; **c** Lower panel; vertical magnetic disturbance field

Furthermore, a 2D interpretation of the observed $[\hat{T}_x, \hat{T}_y]$ tacitly assumes that current channelling effects, due to 3D conductivity structure, are not significant.

The most widely used expressions for estimating $[T_x, T_y]$ are those which lead to estimates unbiased by noise contributions in X and Y (Jones in press 1981a). This estimate for, for example, the T_x transfer function is

$$\hat{T}_x = \frac{\hat{S}_{xz} \hat{S}_{yy} - \hat{S}_{xy} \hat{S}_{yz}}{\hat{S}_{xx} \hat{S}_{yy} - \hat{S}_{xy} \hat{S}_{yx}} \quad (2)$$

The denominator of Eq. (2), which may be written as $\hat{S}_{xx} \hat{S}_{yy} (1 - \hat{\gamma}_{xy}^2)$, where $\hat{\gamma}_{xy}^2$ is the estimated coherence between components $x(t)$ and $y(t)$, indicates the well-known requirement that $x(t)$ and $y(t)$ must contain some independent information, i.e., that the horizontal field is not totally polarised.

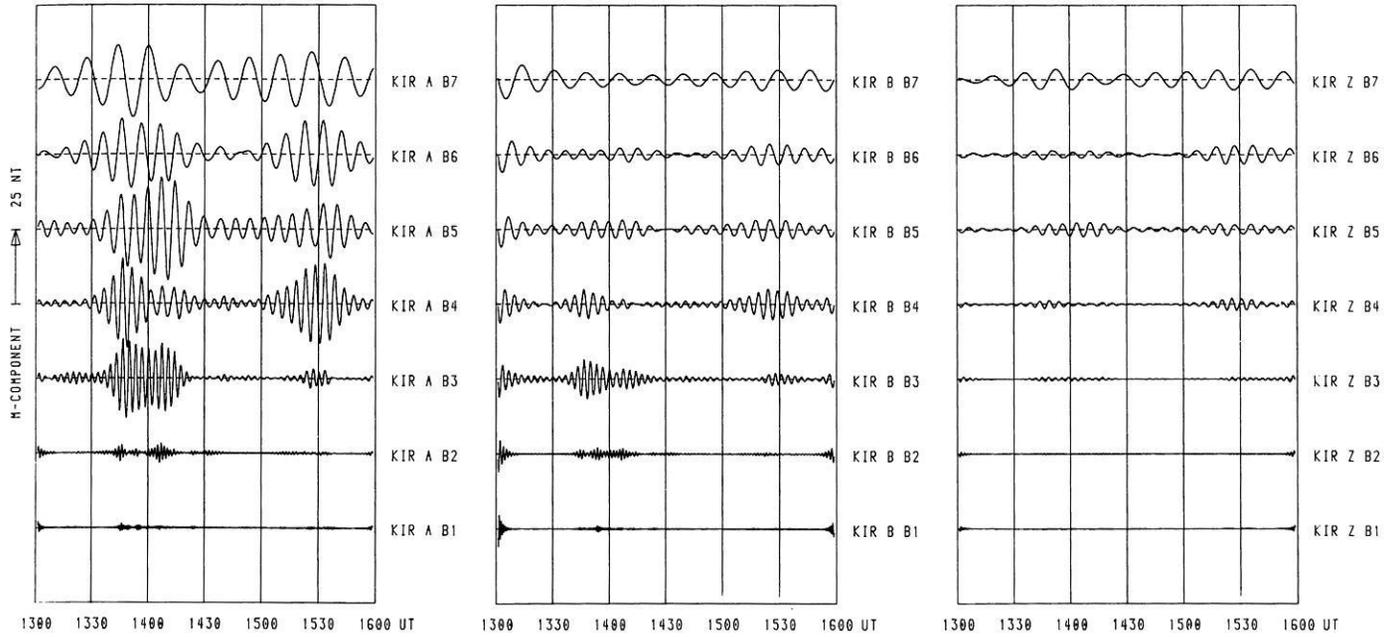


Fig. 5. Sonogram analysis of the magnetic disturbance recorded at Kiruna (KIR - see Fig. 1). The filters had a selectivity of 0.3, and a centre period of: B1 60 s; B2 100 s; B3 200 s; B4 300 s; B5 450 s; B6 600 s; B7 1,000 s (A, B, Z as in Fig. 2)

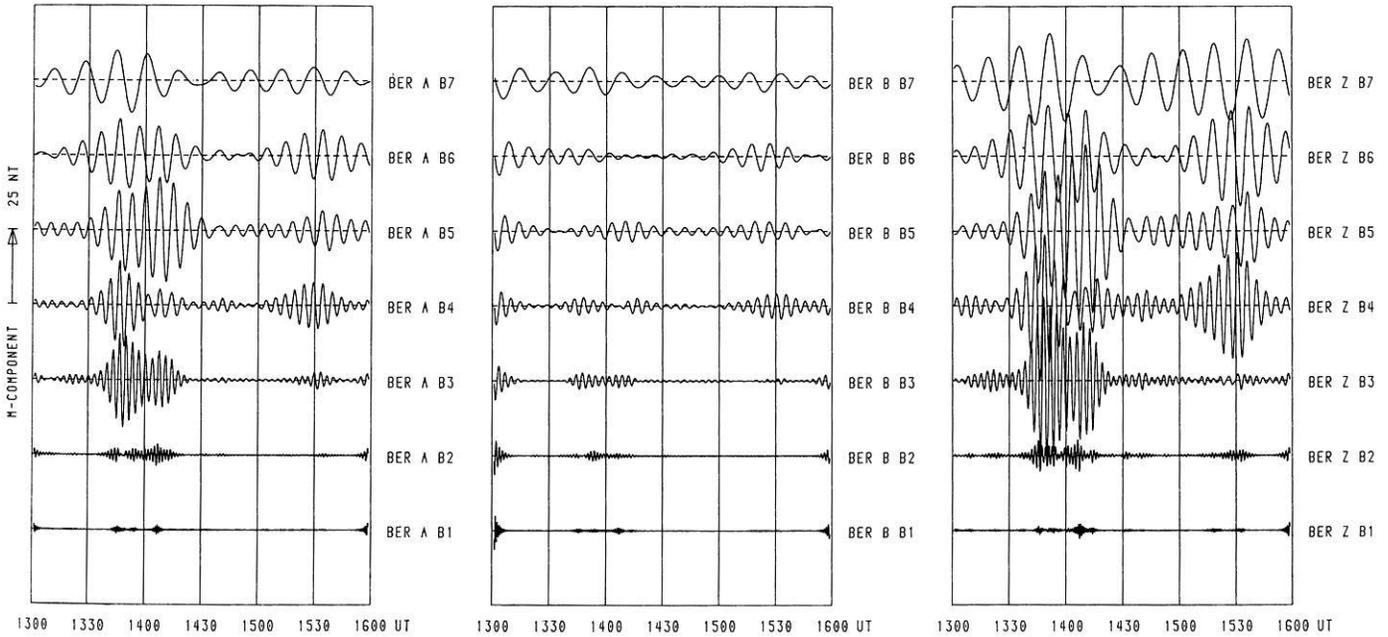


Fig. 6. As Fig. 5 but for the data recorded at Berlevag (BER - see Fig. 1)

A quantitative measure of how well the estimation process has been undertaken is given by the estimates of the circles of confidence, at a certain probability level, applicable to the estimates $[\hat{T}_x, \hat{T}_y]$. Assuming that the error, ξ , in Eq. 1 is a random variable with a normal probability distribution function and which is totally uncorrelated to $x(t)$, $y(t)$, or $z(t)$, the estimates of the radii of the circles of confidence are given by

$$\hat{r}_x^2 = \frac{4}{v-4} F_{4:v-4;\alpha} \frac{(1-\hat{\gamma}_{zxy}^2) \hat{S}_{zz}}{(1-\hat{\gamma}_{xy}^2) \hat{S}_{xx}} \quad (3)$$

(similarly for \hat{r}_y^2 by replacing \hat{S}_{xx} with \hat{S}_{yy} .)

where

v = number of degrees of freedom associated with the estimate, given by twice the number of raw estimates averaged over ($v=10$ at 1,000 s, and $v \approx 100$ at 100 s)

$F_{4:v-4;\alpha}$ = 100 α percentage point of the $F_{4:v-4}$ distribution function,

and

$\hat{\gamma}_{zxy}^2$ = estimate of the multiple coherence between $z(t)$, and $x(t)$ and $y(t)$ as inputs, which is equal to $(1-S_{\xi\xi}/S_{zz})$ (Goodman 1965, repeated in Bendat and Piersol 1971).

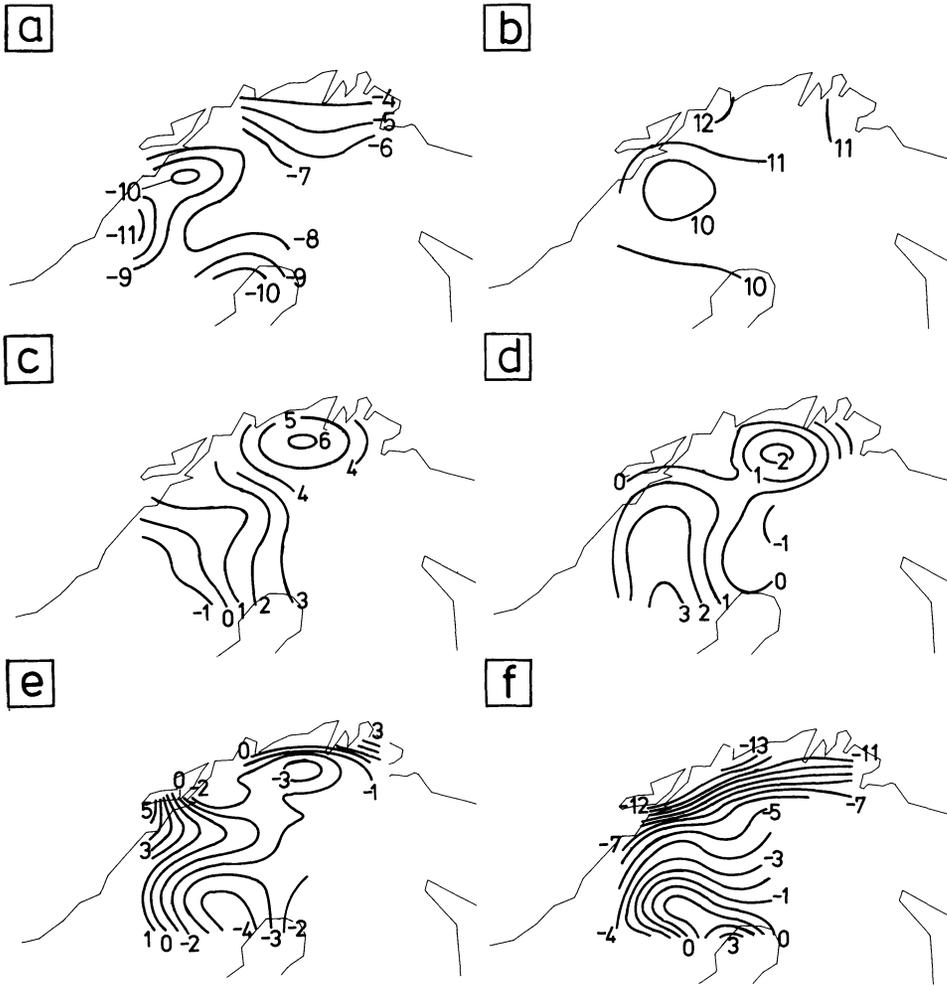


Fig. 7a-f. Smoothed Fourier spectral density maps for a central period of 1,000 s for the data illustrated in Fig. 2. **a** Upper left, real part of $X(f)$; **b** Upper right, imaginary part of $X(f)$; **c** Middle left, real part of $Y(f)$; **d** Middle right, imaginary part of $Y(f)$; **e** Lower left, real part of $Z(f)$; **f** Lower right, imaginary part of $Z(f)$ (arbitrary units)

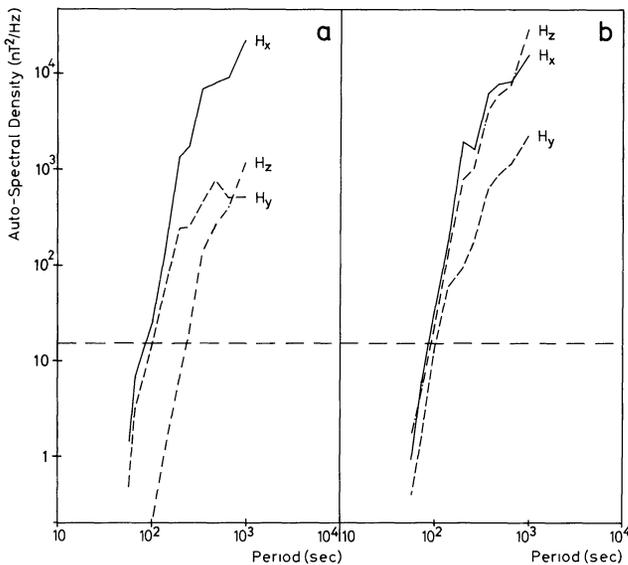


Fig. 8a, b. Smoothed Auto-Spectral Densities, in nT^2/Hz , for the data recorded at **a** KIR and **b** BER. H_x , H_y and H_z refer to the north, east, and vertical components respectively in the Kiruna system. The dotted line indicates the assessed resolution level (see text)

Denoting unit vectors \mathbf{i} and \mathbf{j} as pointing along axes x_{KI} and y_{KI} respectively, the estimated real and imaginary induction vectors are given from the estimates $[\hat{T}_x, \hat{T}_y]$ by

$$\hat{\mathbf{V}}_r = -\text{Re}(\hat{T}_x) \cdot \mathbf{i} - \text{Re}(\hat{T}_y) \cdot \mathbf{j} \quad (4a)$$

$$\hat{\mathbf{V}}_i = \text{Im}(\hat{T}_x) \cdot \mathbf{i} + \text{Im}(\hat{T}_y) \cdot \mathbf{j} \quad (4b)$$

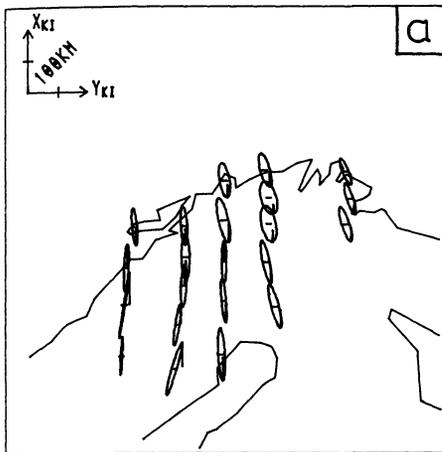
(note that the real vector only has been reversed to point towards internal current), and the confidence intervals of $\hat{\mathbf{V}}_r$ and $\hat{\mathbf{V}}_i$ are described by an ellipse with axes \hat{r}_x and \hat{r}_y .

The single-station induction vectors, estimated by Eqs. (2), (4a) and (4b) for the event, are illustrated in Fig. 10 (real and 11 (imaginary) at the four periods 100 s (10a, 11a), 200 s (10b, 11b), 450 s (10c, 11c) and 1,000 s (10d, 11c). The 68% probability level confidence ellipses for the end points of each vector are also illustrated in the figures by their major (here \hat{r}_y because of the smaller magnitude of \hat{S}_{yy} compared with \hat{S}_{xx} ; Fig. 8) and minor (\hat{r}_x) axes.

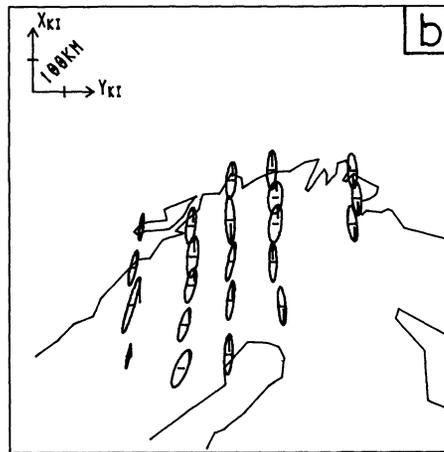
The coast effect is a very dominant feature of the derived in-phase induction vectors, but intracontinental effects are also noticeable:

- (i) the small induction vector in the neighbourhood of KIR at 100 s–450 s,

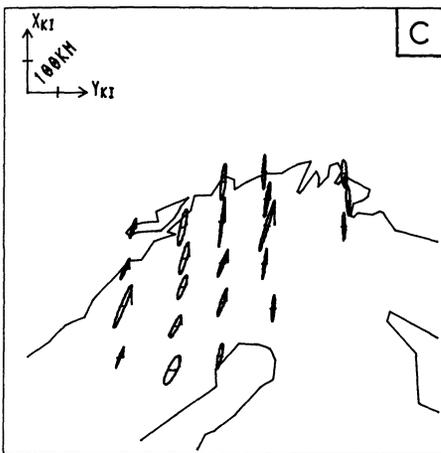
77-07-06 1000S



77-07-06 450S



77-07-06 200S



77-07-06 100S

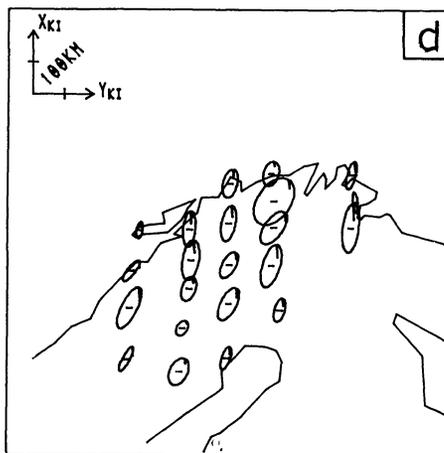


Fig. 9a-d. Ellipses of the polarised power of the horizontal magnetic field, as observed at a 1,000 s, b 450 s, c 200 s and d 100 s

- (ii) the presence of a large anomaly in the OKS-RIS-SRV region,
- (iii) the effect on the vectors at LYC and PIT due to the Gulf of Bothnia.

The large predominantly northward-pointing in-phase vectors observed at 1,000 s may be due to source-field influence. If KIR can be assumed to be over a horizontally layered earth, as concluded in Paper I, then the activity in the z -component at long periods (>500 s), as seen in Figs. 5 and 8, is an indication that the external vertical magnetic field is not being cancelled by the internally induced vertical magnetic field. Hence at long periods the source field of this event is not totally uniform. Work on 3D source fields over a 1D earth (Mareschal 1980) showed that data recorded further than 6° to 8° south of a 3D electrojet's southern boarder, the earth being a half space of $1,000 \Omega\text{m}$ resistivity, will not be significantly affected for periods less than 60 min. However the very noticeable three-dimensionality of Scandinavia and the coastal and nearby oceanic waters make inferences of the induction effects of non-uniform source fields very difficult, but it may be concluded that even if source fields do affect the northern coastal stations, the effect is not significant for stations at geomagnetic latitude around that of OKS-PEL.

Coast Effect

The coast around Scandinavia is, as illustrated in the bathymetry map (Fig. 12), of highly complex form. The Nor-

wegian Sea is of deep ocean depths (2-3 km) and the continental edge is close to the Lofoten station of FRE, but the Barents Sea is shallow, less than 500 m, for the most part less than 300 m, over its whole extent.

In order to examine the coast effect exhibited along a line perpendicular to the continental edge (AA' in Fig. 12), the estimated single-station transfer functions $[\hat{T}_x, \hat{T}_y]$ were rotated into a co-ordinate system with axes parallel and perpendicular to the coast, to yield $[\hat{T}_\parallel, \hat{T}_\perp]$. For the E -polarisation case, i.e., electrical current flowing SW/NE in the ocean, the ratio of the vertical to horizontal magnetic field is given by the transfer function \hat{T}_\perp .

The estimated real parts of the \hat{T}_\perp transfer function at two periods, 200 s and 1,000 s, for stations FRE, GLO, EVE, ROS, RIJ, KVI, KIR, NAT, together with their associated 68% probability confidence intervals (i.e., one standard deviation of the mean) resolved into a direction parallel to AA', are shown in Fig. 13. The station EVE exhibits a negative real part at 200 s, which may be an indication of current perturbation in Ofotfjord, a 60 km long fjord south of EVE (Evanes) leading to the port of Narvik. The negative $\text{Re}(\hat{T}_\perp)$ observed at NAT is interpreted as due to current perturbation in the Baltic and around the Gulf of Bothnia (see next section).

A first approximation to the expected attenuation of the vertical/perpendicular-horizontal magnetic field ratio is given by the E -polarisation results of the somewhat drastic model

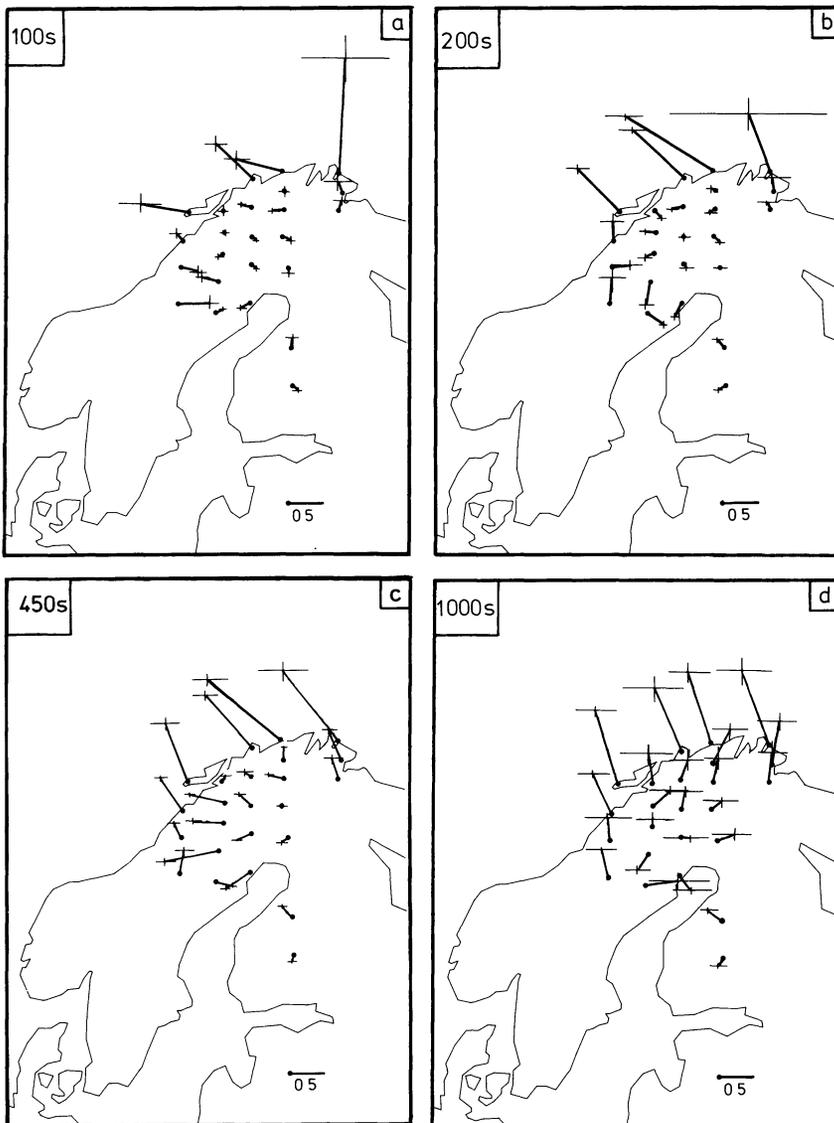


Fig. 10a-d. Reversed real induction vectors, and their 68 % intervals, as observed for periods of a 100 s, b 200 s, c 450 s, and d 1,000 s

of Fischer et al. (1978, reviewed in Fischer 1979) of a perfectly conducting ocean lying on a medium of uniform resistivity ρ . With the values of $\rho=230\ \Omega\text{m}$ at 200 s, and $\rho=130\ \Omega\text{m}$ at 1,000 s (taken from Paper I), the attenuation of $\text{Re}(T_{\perp})$ with distance from the ocean-land boundary was determined from Fig. 13 of Fischer (1979). This theoretical attenuation is displayed in Fig. 13, at the two periods, and a reasonably good fit to the data is achieved.

The effect of a laterally-varying integrated conductivity thin-sheet model, shown in Fig. 14, which describes the bathymetry (shown as a dashed line in Fig. 14) observed along profile AA', with a subsurface of a $125\ \Omega\text{m}$, 110 km thick layer overlying a $3\ \Omega\text{m}$ half-space (taken from Paper I), was calculated using the programme of Schmucker (1971). Two models were employed, one with, and one without, a "tau" representation of Vestfjord which lies between the Lofoten Islands and the mainland.

The derived $\text{Re}(H_z/H_{\perp})$ for both models are also illustrated in Fig. 13a, b, where a very good fit is apparent between the data and the model. At the shorter period GLO fits to the model without a Vestfjord, as should be expected as Glomfjord (GLO) lies directly on the coast, but it is clear that

the data from RIJ and ROS are better explained by including an inductive effect in Vestfjord.

A very important conclusion to be drawn from this model fitting is that the data resolved along profile AA' are almost totally explainable by induction at a coast-land boundary. No lateral variation in mantle conductivity along AA' is required, although such a variation cannot be excluded.

The data along profile BB', however, do not fit a much simpler τ model of

$$\tau = \begin{cases} 1,200\ \text{S} & \text{for } 150\ \text{km} < x \\ 1,000\ \text{S} & \text{for } 0 < x < 150\ \text{km} \\ 0.1\ \text{S} & \text{for } x < 0, \end{cases}$$

where station BER is located at $x=0$, to represent the shallow Barents Sea, over a three-layer substructure with parameters $\rho_1=10^4\ \Omega\text{m}$, $d_1=30\ \text{km}$, $\rho_2=125\ \Omega\text{m}$, $d_2=140\ \text{km}$ and $\rho_3=3\ \Omega\text{m}$ (model taken from Paper I). The values of $\text{Re}(\hat{T}_x)$ are all a factor of 2-4 too large, which is interpreted as the influence of one, or both, of two possible effects:

(i) two-dimensional current flow around the northern coast of Scandinavia, in particular around the Varangerhalvöya penin-

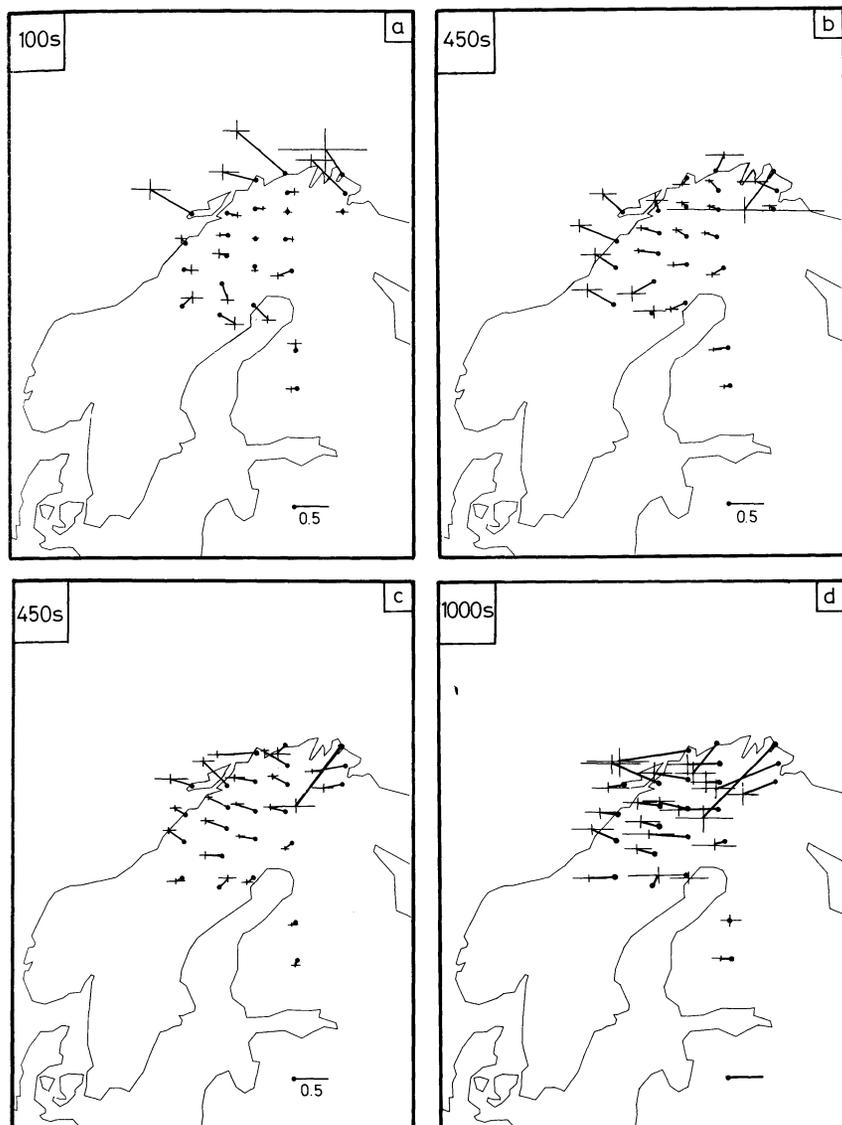


Fig. 11 a-d. Unreversed imaginary induction vectors, and their 68 % confidence intervals, as observed for periods of **a** 100 s, **b** 200 s, **c** 450 s and **d** 1,000 s

sula on which Berlevag (BER) and Vadsö (VAD) lie, (ii) distant source contributions, as discussed above.

The effect cannot be due to a more conducting mantle under the Barents Sea because, as shown by Dosso (1973), a conducting mantle under an ocean reduces the H_z/H_x ratio.

Discussion and Effects on Source-Field Studies

This paper has concentrated on an analysis, by many different methods, of one single event which was very suitable for induction purposes. The striking horizontal component uniformity, unusual for auroral and sub-auroral zones, ensured that particularly the variations in the vertical component, $z(t)$, reflected contributions from induced fields.

Two inland anomalies were identified in almost all of techniques employed on the data. The more dominant of the two, the Storavan anomaly, is delineated approximately by the stations PIT-LYC-OKS-SRV. This anomaly, which lies between the well-known Svecokarelian fault to the north (which in Finland is a zone of extensive mineralisation) and the Bothnian basin to the south, coincides almost exactly

with a postulated remnant of the Svionian island arc system (Meissner 1979). The other inland anomaly, the Mieron anomaly, which appears to be polarisation sensitive, does not correlate with any known geological or tectonic formation.

With respect to induction vectors and their determination, Gough and de Beer (1980) recently cast aspersions regarding the validity of interpreting induction vectors estimated from events where the horizontal components are clearly well correlated. Their study illustrated that vectors derived from one set of data – in this case their 1971 array data – do not agree with those from another set of data – their 1977 array data. They explained the discrepancy as being due to an incorrect choice of data for induction vector estimation purposes in the latter case, and suggested two “practical precautions”; (a) “events should have dissimilar X and Y magnetograms”, and (b) “cross-correlations should be computed” (between the X and Y data) ... (and) “Events should be included whose cross-correlations peak at time shifts distributed between large positive and negative values”. These constraints can however be easily deduced from inspection of Eq.(3), which gives the expression for estimating the confidence intervals of estimated transfer functions, and are already part of the theory of

statistical frequency analysis. For a two-input/single-output linear system, as described by Eq. (1), the confidence intervals are reduced by: (i) increasing ν , the number of degrees of freedom, (ii) reducing $\hat{\gamma}_{xy}^2$, the estimate of the coherence between the two input series, i.e., $x(t)$ and $y(t)$, and (iii) increasing $\hat{\gamma}_{zxy}^2$, the estimate of the multiple coherence between the output, i.e., $z(t)$, and the two inputs. Point (i) may be accomplished either by ensemble averaging, i.e., over different realisations, or events, of the data, or by frequency smoothing, i.e., averaging together, in some fashion, a certain number of neighbouring raw Fourier harmonic estimates, or by both.

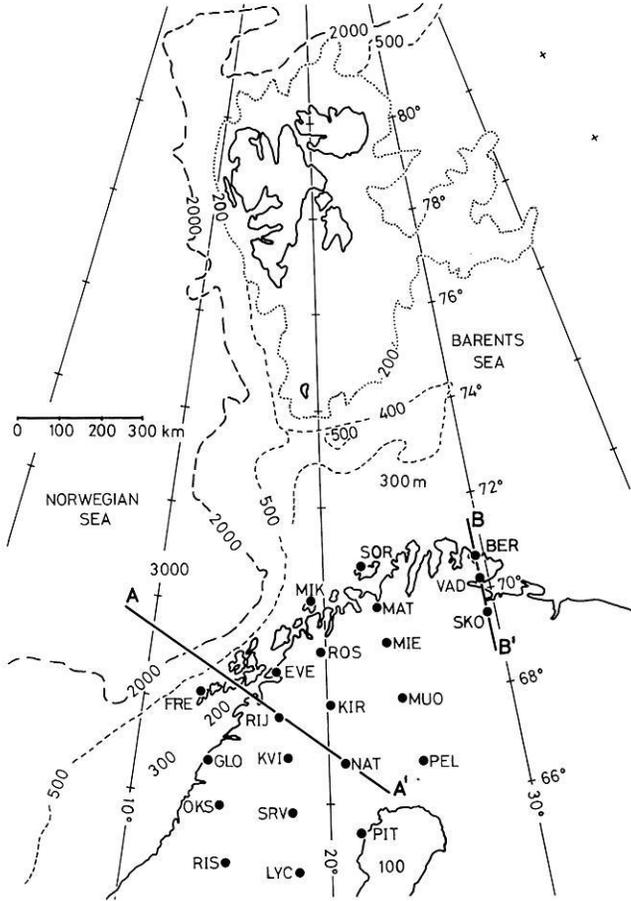


Fig. 12. Bathymetry map of the coast around northern Scandinavia, showing water depth (in meters), the station locations, and profiles AA' and BB'

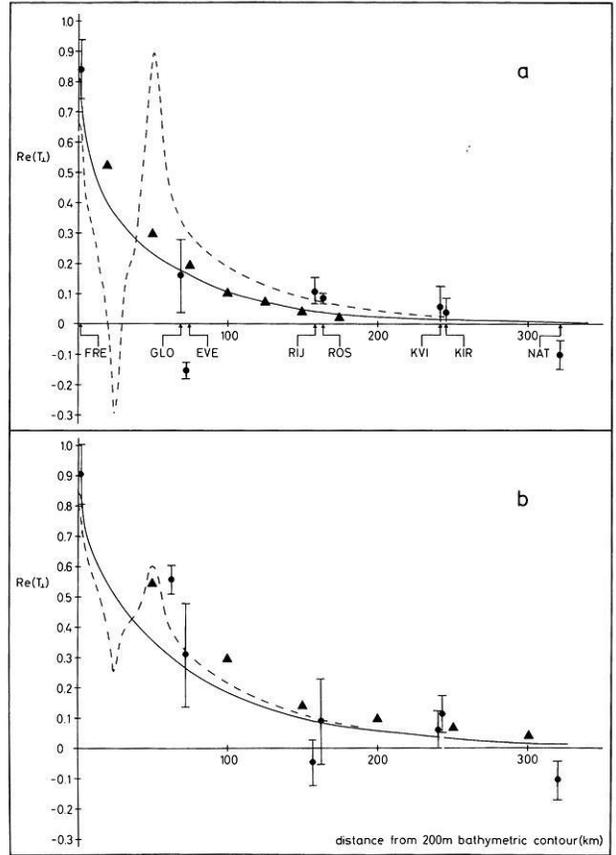


Fig. 13a, b. Dots indicate the magnitude of the real part of the inductive transfer function at **a** 200 s and **b** 1,000 s in the direction parallel to AA', i.e., perpendicular to the coast, or E-polarisation, for the stations FRE, GLO, EVE, RIJ, ROS, KVI, KIR, and NAT, as compared to the distance from the 200 m bathymetric contour west of FRE (see Fig. 12). The ranges are the 68 % confidence intervals of the rotated functions. Triangles indicate the Fischer first-approximation of a perfectly conducting thin sheet overlying a half space at **a** 200 s and **b** 1,000 s for half-spaces of 230 Ωm and 130 Ωm respectively. Full lines are the Schmucker thin-sheet modelling results for the model shown in Fig. 14 without a Vestfjord representation. Dashed lines are the Schmucker thin-sheet modelling results for the model shown in Fig. 14 with a tau representation of Vestfjord (see text)

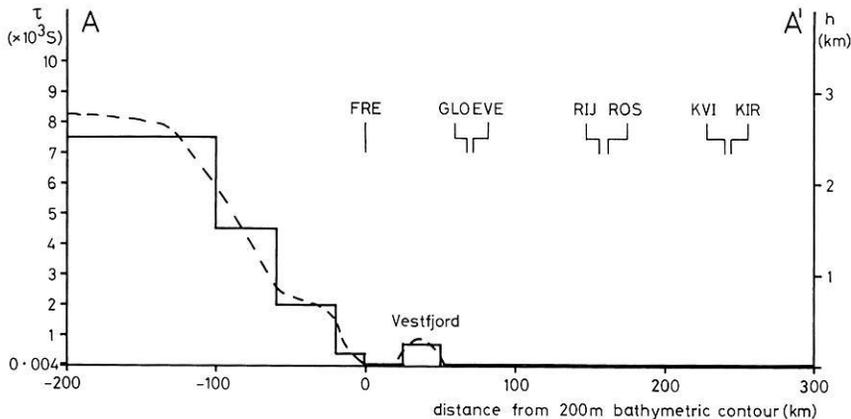


Fig. 14. Full line: thin-sheet model approximation of the lateral variation in crustal conductivity along Profile AA'. The continental crust has an integrated conductivity of 4S. The locations of the stations, as resolved along AA', are also illustrated. The thin-sheet surface overlies a 110 km thick layer of 125 Ωm , followed by a half-space of 3 Ωm . Dashed line: the variation in bathymetry along profile AA'. The position of Vestfjord, which lies between the Lofoten Islands and the mainland, i.e., between FRE and GLO, is also indicated

When performing such an operation however, it must be without doubt that ν is actually being increased, i.e., that the different individual estimates to be averaged contain independent information. For geomagnetic data, this requires that the source field structure be sufficiently different for each realisation, which is clearly not the case for 7 of the 10 events from the 1977 array data discussed by Gough and de Beer (1980), as shown by their Fig. 7 (lower part). Point (ii) above, the requirement of a low coherence between the two horizontal components, is a statistical description of point (b) of Gough and de Beer. The necessity of a high coherence, or correlation in the time domain, between $z(t)$ and the horizontal components, $x(t)$ and $y(t)$, is not mentioned by Gough and de Beer, but is obviously of equal importance to their points (a) and (b). For the 10 events from their 1977 array data, the ensemble average had a low ν (possibly only 3 significantly different polarisations from the 10 employed, see their Fig. 7; lower part) and a high $\hat{\nu}_{xy}^2$, which both add to produce large values of the estimates of the confidence intervals of $[\hat{T}_x, \hat{T}_y]$, i.e., \hat{r}_x^2 and \hat{r}_y^2 . It is therefore highly probable that, contrary to the statements of Gough and de Beer (1980), the vectors derived from the 1977 array data are totally in accord with those estimated from the 1971 array data.

For the event analysed here, the differing polarisations for the two "bursts" of activity, and at differing frequencies, ensured that the estimates would contain sufficient information. Also, by calculating the confidence intervals it is possible to be certain that, at the 68% confidence level, the directions and amplitudes are significant.

A first approximation modelling of the response function, observed along a line which is normal to the continental shelf structure, inferred that the data from stations FRE, GLO, EVE, RIJ, ROS, KVI, and KIR may be adequately explained by the conductivity contrast between the continent and the ocean. No lateral variation in lower crust or mantle conductivity was required. However, this does not exclude such a variation, it would be necessary to consider longer period variations, i.e., >1 h, to gain more insight into the likely variation of the Moho or asthenosphere at the continent – ocean boundary.

For the original purpose of the array, it is obviously of paramount importance if the coast effect and the two inland anomalies detailed herein have a significant effect on the total magnetic fields observed, as this would make untenable the rather direct interpretation of the magnetic disturbance vectors, as equivalent current vectors, usually undertaken by the Münster IMS group. As 3D Earth-conductivity modelling with typical ionospheric source structure has not yet been accomplished, the only recourse is to inspect work, either published or in press, for indications of effects that could be attributable to anomalous conductivity structures. The large positive time-difference Z fields observed by PIT and MIK at the time of the abrupt "switch-on" of an additional current system, as detailed in Untiedt et al. (1978), Fig. 14, is possibly due to a phase delay of short period Z variations, as inferred by the large imaginary induction vectors at short periods (see Fig. 11). The large negative Z fields observed by stations MAT and MIE at 17:42 UT in Fig. 3 of Mersmann et al. (1979), when compared with the neighbouring stations of ROS and TR (Tromsø magnetic observatory), are possibly artifacts of the polarisation sensitive anomaly centered on those stations. In the same figure, the equivalent current arrows at PIT and OUL (Oulu, not shown here, see Fig. 2 of Küppers et al. 1979) are anomalous, and are due to induction

in the Gulf of Bothnia region. This effect is also seen in Fig. 11 of Baumjohann et al. (1980) and Fig. 4 of Baumjohann and Kamide (in press 1981).

In the majority of cases, however, the influence of the induced parts is difficult to recognise or assess. It may not be insignificant that the position of the differential Z -field minimum shown in Fig. 9 of Opgenoorth et al. (1980) corresponds exactly with the location of the Storavan anomaly, or that the maximum of the Y -field at 04:00–04:20 UT of the pulsation event of Głaßmeier (1980), Fig. 5b, is centered on stations MAT and MIE. Indeed, Głaßmeier noted that the polarisation characteristics of station MIE were ambiguous (his Fig. 7, 04:10 UT).

That the Mieron anomaly was not detected in the 2D field separations along profile 4 undertaken by Mersmann et al. (1979) and Küppers et al. (1979) is due to the polarisation properties of the anomaly. Both studies analysed data from 2D eastward electrojets, and, as has been shown, the anomaly becomes apparent only when there is a significant north/south component of equivalent current flow.

For source-field studies, the area of central northern Scandinavia is particularly free of inhomogeneities, hence permitting a 1D Earth approximation. A relevant model for this area is given in Jones (in press 1981c), and comprises a top layer of $10^4 \Omega\text{m}$ and 7 km thick (the sialic layer) overlying a moderately resistive lower crust of $360 \Omega\text{m}$ down to the Moho at 46 km, with a mantle of $80 \Omega\text{m}$ and an asthenospheric layer between 170–230 km depth of $5 \Omega\text{m}$. For this model, at a frequency of 10^4 s for a source field of wavelength 1,200 km, the ratio of the external field to the total field is 0.93, with a phase lead of the former of 140 s. Thus, the assumption made in Baumjohann and Kamide (in press 1981) of neglecting internal contributions for their study is justifiable for central northern Scandinavia to within 10%, which is the accuracy of the modelling involved (Baumjohann, private communication). For events of longer wavelength, or shorter period, the internal contributions become more significant. These comments are, it must be stressed, for a 1D Earth Contributions from a real earth, with especial regard to the coast effect, are expected to be larger. For example, over the Norwegian Sea (modelled by assuming a 2 km deep sea, of $\rho = 0.33 \Omega\text{m}$, overlying a 5 km thick sialic layer, of $\rho = 10^4 \Omega\text{m}$, over the same lower crust and upper mantle as above), the ratio of the external field to the total field at 10^4 s period for a wavelength of 1,200 km is 0.76, with a phase lead of the former over the latter of 470 s. Over the Barents Sea (300 m deep sea-water), these values are 0.91 and 225 s respectively. For the shorter period of 1,200 s and a characteristic wavelength of 1,000 km, the external/total field ratio over central northern Scandinavia is 0.86 with a phase lead of 22 s. In the study of Gustafsson et al. (1981), a perfect conductor at a depth of 125 km was used to describe internal contributions in an attempt to model a very strong Ps 6 event of half-width approximately 500 km. Such an approximation gave an external/total field ratio of 0.89 with, obviously, a phase difference of 0 s. Hence, for events which exhibit a dominant period, a suitable choice of the depth of a perfect conductor is satisfactory to within the accuracies of the model-fit to data observed over central northern Scandinavia. This simple approximation is certainly not valid however if the event to be modelled has a broad power spectrum or if the activity occurs on, or near to, the coasts.

In conclusion, there is certainly a large coast-effect at all the coastal stations, including those on the Gulf of Bothnia

(PIT and OUL), and there is unequivocal evidence for a large conductivity anomaly in the region of Storavan (SRV). This anomaly may be attributable to a remnant of the Svionian island arc system. A second inland anomaly, in the vicinity of stations MAT and MIE, is tentatively proposed here, it displays a polarisation sensitivity, being apparent only for northward or southward directed equivalent current flow, i.e., for a Y-field. A similar anomaly has previously been located in Australia (Gough et al. 1972).

Acknowledgements. The author is deeply grateful to Professors U. Schmucker and J. Untiedt for critically reviewing an earlier version of this manuscript and suggesting many improvements. The author also wishes to express his gratitude to the Deutsche Forschungsgemeinschaft for financial support.

References

- Alabi, A.O., Camfield, P.A., Gough, D.I.: The north American central plains conductivity anomaly. *Geophys. J.R. Astron. Soc.* **43**, 815–834, 1975
- Bannister, J.R., Gough, D.I.: Development of a polar magnetic substorm: A two-dimensional magnetometer array study. *Geophys. J.R. Astron. Soc.* **51**, 75–90, 1977
- Baumjohann, W., Kamide, Y.: Joint two-dimensional observations of ground magnetic and ionospheric electric fields associated with auroral zone currents. 2. Three-dimensional current flow in the morning sector during substorm recovery. *J. Geomagn. Geoelectr.*, in press 1981
- Baumjohann, W., Untiedt, J., Greenwald, R.A.: Joint two-dimensional observations of ground magnetic and ionospheric electric fields associated with auroral zone currents. 1. Three-dimensional current flows associated with a substorm-intensified eastward electrojet. *J. Geophys. Res.* **85**, 1963–1978, 1980
- Beer, J.H. de, Gough, D.I.: Conductive structures in southernmost Africa: a magnetometer array study. *Geophys. J.R. Astron. Soc.* **63**, 479–495, 1980
- Bendat, J.S., Piersol, A.G.: *Random data: analysis and measurement procedures*. New York: Wiley-Interscience 1971
- Born, J., Wolf, E.: *Principles of Optics*. New York: MacMillan 1964
- Dosso, H.W.: A review of analogue model studies of the coast effect. *Phys. Earth Planet. Inter.* **7**, 294–302, 1973
- Fischer, G.: Electromagnetic induction effects at an ocean coast. *Proc. IEEE* **67**, 1050–1060, 1979
- Fischer, G., Schnegg, P.-A., Usadel, K.D.: Electromagnetic response of an ocean coast model to E-polarisation induction. *Geophys. J.R. Astron. Soc.* **53**, 599–616, 1978
- Fowler, R.A., Kotick, B.J., Elliot, R.D.: Polarisation analysis of natural and artificially induced geomagnetic micropulsations. *J. Geophys. Res.* **72**, 2871–2883, 1967
- Frazer, M.C.: Geomagnetic sounding with arrays of magnetometers. *Rev. Geophys. Space Phys.* **12**, 401–420, 1974
- Glabbeier, K.-H.: Magnetometer array observations of a giant pulsation event. *J. Geophys.* **48**, 127–138, 1980
- Goodman, N.R.: Measurement of matrix frequency response functions and multiple coherence functions. AFFDL TR 65-56, Air Force Flight Dynamics Laboratory, Wright-Patterson AFB, Ohio, February 1965
- Gough, D.I.: The interpretation of magnetometer array studies. *Geophys. J.R. Astron. Soc.* **35**, 85–98, 1973a
- Gough, D.I.: The geophysical significance of geomagnetic variation anomalies. *Phys. Earth Planet. Inter.* **7**, 379–388, 1973b
- Gough, D.I., Beer, J.H. de: Source-field bias in geomagnetic transfer functions: A case history. *J. Geomagn. Geoelectr.* **32**, 471–482, 1980
- Gough, D.I., Lilley, F.E.M., McElhinny, M.W.: A polarisation-sensitive magnetic variation anomaly in south Australia. *Nature* **239**, 88–91, 1972
- Gough, D.I., Reitzel, J.S.: A portable three component magnetic variometer. *J. Geomagn. Geoelectr.* **19**, 203–215, 1967
- Gustafsson, G., Baumjohann, W., Iversen, I.: Multi-method observations and modelling of the three-dimensional currents associated with a very strong Ps 6 event. *J. Geophys.* **49**, 138–145, 1981
- Hermance, J.F.: Processing of magnetotelluric data. *Phys. Earth Planet. Inter.* **7**, 349–364, 1973
- Jones, A.G.: On the difference between polarisation and coherence. *J. Geophys.* **45**, 223–229, 1979
- Jones, A.G.: Geomagnetic Induction Studies in Scandinavia-I. Determination of the inductive response function from the magnetometer array data. *J. Geophys.* **48**, 181–194, 1980
- Jones, A.G.: On the reduction of bias in response function estimation for hand-digitised data. *Geophys. J.R. Astron. Soc.*, in press 1981a
- Jones, A.G.: Comment on “Geomagnetic Depth Sounding by Induction Arrow Representation: A Review” by G.P. Gregori and L.J. Lanzerotti, *Rev. Geophys. Space Phys.*, in press 1981b
- Jones, A.G.: On the electrical crust mantle structure in Fennoscandia: No Moho, and the Asthenosphere revealed? *Geophys. J.R. Astron. Soc.*, in press 1981c
- Küppers, F., Post, H.: A second generation Gough-Reitzel magnetometer. *J. Geomagn. Geoelectr.* **33**, 225–237, 1981
- Küppers, F., Untiedt, J., Baumjohann, W., Lange, K., Jones, A.G.: A two-dimensional magnetometer array for ground-based observations of auroral zone electric currents during the International Magnetospheric Study (IMS). *J. Geophys.* **46**, 429–450, 1979
- Lilley, F.E.M.: Magnetometer array studies: A review of the interpretation of observed fields. *Phys. Earth Planet. Inter.* **10**, 231–240, 1975
- Mareschal, M.: A modelling of some effects of the source field on the Z/H ratio evaluated at sub-auroral latitudes in the period range 15 to 60 min. (Abstract) EOS, *Trans. Am. Geophys. Union* **61**, 1980
- Meissner, R.: Fennoscandia – A short outline of its geodynamic development. *GeoJournal* **3.3**, 227–233, 1979
- Mersmann, U., Baumjohann, W., Küppers, F., Lange, K.: Analysis of an eastward electrojet by means of upward continuation of ground-based magnetometer data. *J. Geophys.* **45**, 281–298, 1979
- Opgenoorth, H.J., Pellinen, R.J., Maurer, H., Küppers, F., Heikkilä, W.J., Kaila, K.U., Tanskanen, P.: Ground-based observations of an onset of localised field-aligned currents during auroral breakup around magnetic midnight. *J. Geophys.* **48**, 101–115, 1980
- Schmucker, U.: Anomalies of geomagnetic variations in the southwestern United States. *Bull. Scripps Inst. Oceanogr., Univ. Calif. Press* **13**, 1970
- Schmucker, U.: Interpretation of induction anomalies above nonuniform surface layers. *Geophysics* **36**, 156–165, 1971
- Schmucker, U.: Regional induction studies: a review of methods and results. *Phys. Earth Planet. Inter.* **7**, 365–378, 1973
- Schmucker, U.: Diskussionsbeitrag zu „Über die Unterschiede zwischen verschiedenen Definitionen der Induktionspfeile“, In: Protokoll über das Kolloquium „Elektromagnetische Tiefenforchung“ in Berlin, 31.3.–3.4. 1980
- Untiedt, J., Pellinen, R., Küppers, F., Opgenoorth, H.J., Pelster, W.D., Baumjohann, W., Ranta, H., Kangas, J., Czechowsky, P., Heikkilä, W.J.: Observations of the initial development of an auroral and magnetic substorm at magnetic midnight. *J. Geophys.* **45**, 41–65, 1978

Received April 6, 1981; Revised version June 14, 1981

Accepted June 25, 1981

Magnetotelluric Sounding in the Transitional Zone Between the Eastern Alps and Pannonian Basin

A. Ádám¹, F. Márcz¹, J. Verö¹, Á. Wallner¹, G. Duma², and R. Gutdeutsch²

¹ Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences H-9401 Sopron, POB 5, Hungary

² Institut für Meteorologie und Geophysik, Universität Wien, Währingerstr. 17, A-1090 Wien, Austria

Abstract. Magnetotelluric sounding has detected a correspondence between some seismically active narrow fracture zones at the boundary of the Eastern Alps and an extreme conductivity increase at a depth of about 7 km. The shallow depth (30–40 km) of a conductivity anomaly detected in the deep crust may result from the increased heat flow in the transitional zone between the Pannonian Basin and the Alps. A close relation between the electric and seismic crustal models seems to exist at depths around 7 km.

Key words: Eastern Alps – Pannonian Basin – Magnetotelluric sounding – Electric conductivity anomaly – Tectonics

Introduction

In the years 1978 and 1979, 6 magnetotelluric deep soundings were carried out along the Sects. 04 and 05 of the Alpine Longitudinal Profile (Alpine Explosion Seismology Group 1976) (Fig. 1). These measurements were performed as a cooperative project between the Geodetic and Geophysical Research Institute of the Hungarian Academy of Sciences at Sopron and the Meteorological and Geophysical Institute of Vienna University.

From seismic experiments it is known that the relatively thin crust of the Pannonian Basin (25 km) becomes thicker towards the West, from the margin of the Eastern Alps, up to a thickness of about 50 km. Furthermore, according to other seismic measurements, the crust is not uniform, velocity inversions having been found at depths of 7–10 km and 20–30 km (Gutdeutsch and Arič, 1977). The eastern margin of the Eastern Alps consists of seismically active fracture zones.

Seismological information may be supplemented by information from geoelectric methods, as they reflect physical parameters other than the velocity of elastic waves. The most important factors which determine the electrical resistivity are the fluid content and the temperature of the rocks. The seismic velocity is less sensitive to these factors than the electrical resistivity. These considerations have led us to the idea of performing geoelectric measurements along the Alpine Longitudinal Profile.

Measurements

The distribution of electrical resistivity may be investigated by magnetotelluric deep sounding (Cagniard 1953). This method uses variations of the natural horizontal electromagnetic field for the determination of the electric impedance Z and the resistivity ρ as functions of the period T .

The equipment consisted of telluric and magnetic sensors and an analogue recording instrument. Chart speeds used were 2 cm/min for daytime and 0.6 cm/min for nighttime. The maximum scale value for electric potentials was 10 $\mu\text{V}/\text{mm}$. Magnetic variations were transformed into electric signals by Hungarian-made MTV-2 variometers with a maximum scale value of 0.01 nT/mm (Ádám and Major 1967). The noise level in the measurement area was rather high. It was possible to more or less avoid regional noise caused by industry, mines, electric railways etc., but noise caused by local consumers, mainly by electrical installations on farms was unavoidable. As the low resistivity sedimentary cover is thin, local noise had high energy. In the evening hours the so-called “TV-noise” appeared (including also noise of other origin such as public lighting). Due to these types of noise the long period variations (substorms) could be used only from 11 p.m. local time till the early morning, while for shorter period variations (pulsations) the most advantageous time of the day was the early morning till about 7 a.m..

Data Processing

As a first step an appropriate number of recording intervals with low noise level and different frequency content were chosen at each measuring point. The length of the intervals was determined by the acceptable signal/noise ratio, being generally between 10 min and several hours. From each of these 300–1,000 digital values were obtained.

The digital data were processed on the HP2100 computer of the Geodetic and Geophysical Research Institute in Sopron, using the program described by Verö (1972). This program calculates the impedance elements by filtering in the time-domain. The level of acceptance for every sequence of 10 points is that the coherencies $\text{Coh}(E_x H_y)$ and $\text{Coh}(E_y H_x)$ should be greater than 0.9, with the exception of the station Lassnitz where the level had to be decreased to 0.6 because of the extremely low resistivity, i.e., low E -amplitudes. The output of the program consists of impedance polar diagrams, resistivities in different directions, directions of the extrema, ρ_{\max} and ρ_{\min} , and phase characteristics of the impedance.

Results

The results of the data processing are: Sounding curves ρ_{\max} and ρ_{\min} (Fig. 2), phase curves φ_{\max} and φ_{\min} in the directions of ρ_{\max} and ρ_{\min} (Fig. 3), and polar diagrams of the main and secondary impedances.

The sounding curves and the phase curves are drawn as

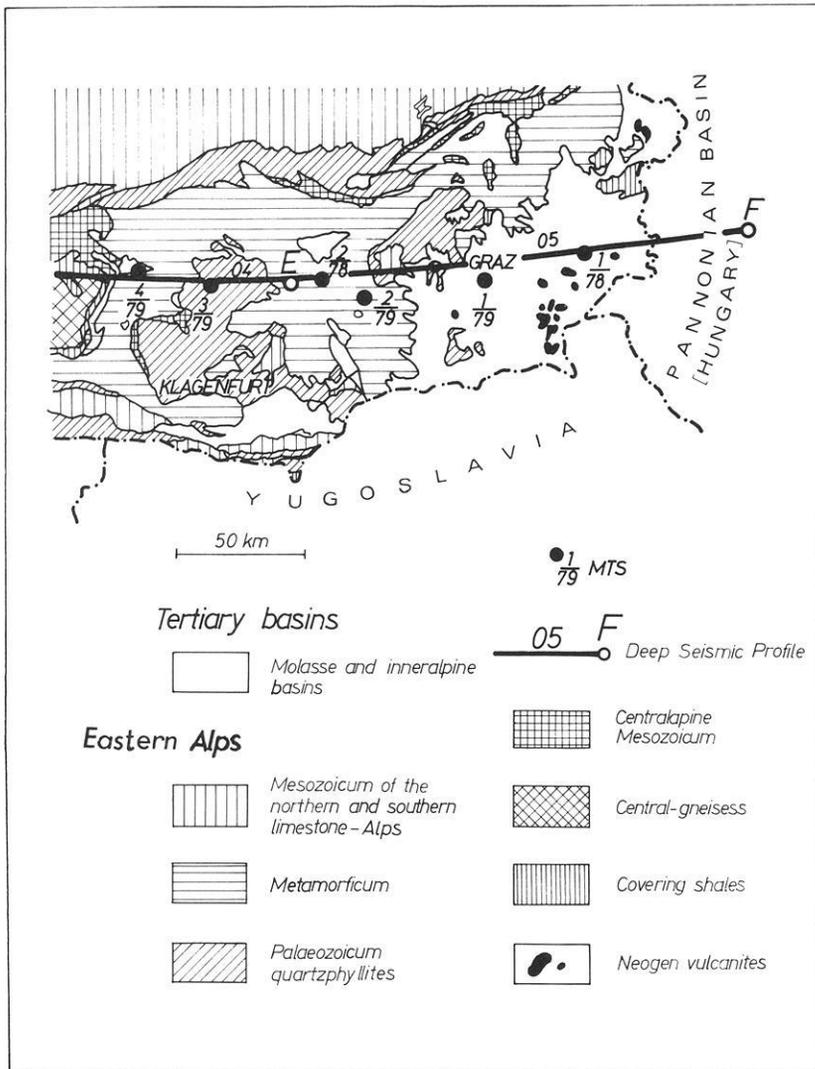


Fig. 1. Geologic map of Southeastern Austria (after P. Beck-Manyetta, Geologische Bundesanstalt, 1964) with magnetotelluric sounding points (black dots) and the Alpine Seismic Crustal Profile

near as possible to the median values of the corresponding quantities in period ranges containing a sufficient number of data; error bars denote mean square errors of the medians determined from the scatter of the individual values.

The sounding curves were approximated by 1D models, separately for ρ_{max} and ρ_{min} curves. Depth and resistivity values result from these calculations, errors of the depths are deduced from families of theoretical curves. A complete agreement between observed and theoretical curves (model) was not always reached, e.g., sharp peaks and valleys at the station Oberpreitenegg could not be approximated. These 1D models may only be regarded as first approximations. Following the tectonic model and the theory of field distortion (for details see next section), we aimed at a close fit of the decreasing parts of the E -polarization curves. In Fig. 2 the dashed curves represent the best approximations of the ρ_{min} curves. From the parameters of these models, the depths to the top of conductive zones are given in Table 1. The phase curves computed on the basis of the models differ from the observed ones in the majority of cases (see the dashed curves in Fig. 3 and also Berktold et al. (1976); Berdichevsky and Dmitriev (1976) also do not use phase curves in the description of 2D distortions). The digitization method could be responsible for a minor part of the deviations. These errors, however, will be at least partly randomized because of the rather high number of sections used.

Table 1. Depths of conductive zones (1D-models) derived from ρ_{min} - curves in the period range of the measurements

Station	Depth (km)	
Rehgraben (1/78)	26 ± 6	162 ± 15
Breitenbuch (1/79)	9 ± 1	37 ± 2.5
Oberpreitenegg (2/79)	7.5 ± 0.6	35 ± 3.5
St. Georgen (2/78)	32 ± 8	
Lassnitz (3/79)	5.5 ± 1.2	
Lasaberg (4/79)	8.5 ± 1	

Interpretation of Data

Distortion Effects

Since the curves ρ_{min} and ρ_{max} as well as the models determined on the basis of these curves differ from each other, the problem is to decide which sounding curve is less influenced by near-surface distortion of the electric field. In the measuring area field distortions could be caused by:

- the mountain frame (edge effect);
- fracture tectonics;
- the varying thickness of surface sediments (S -effect);
- a combination of these effects.

In the following these possibilities will be investigated.

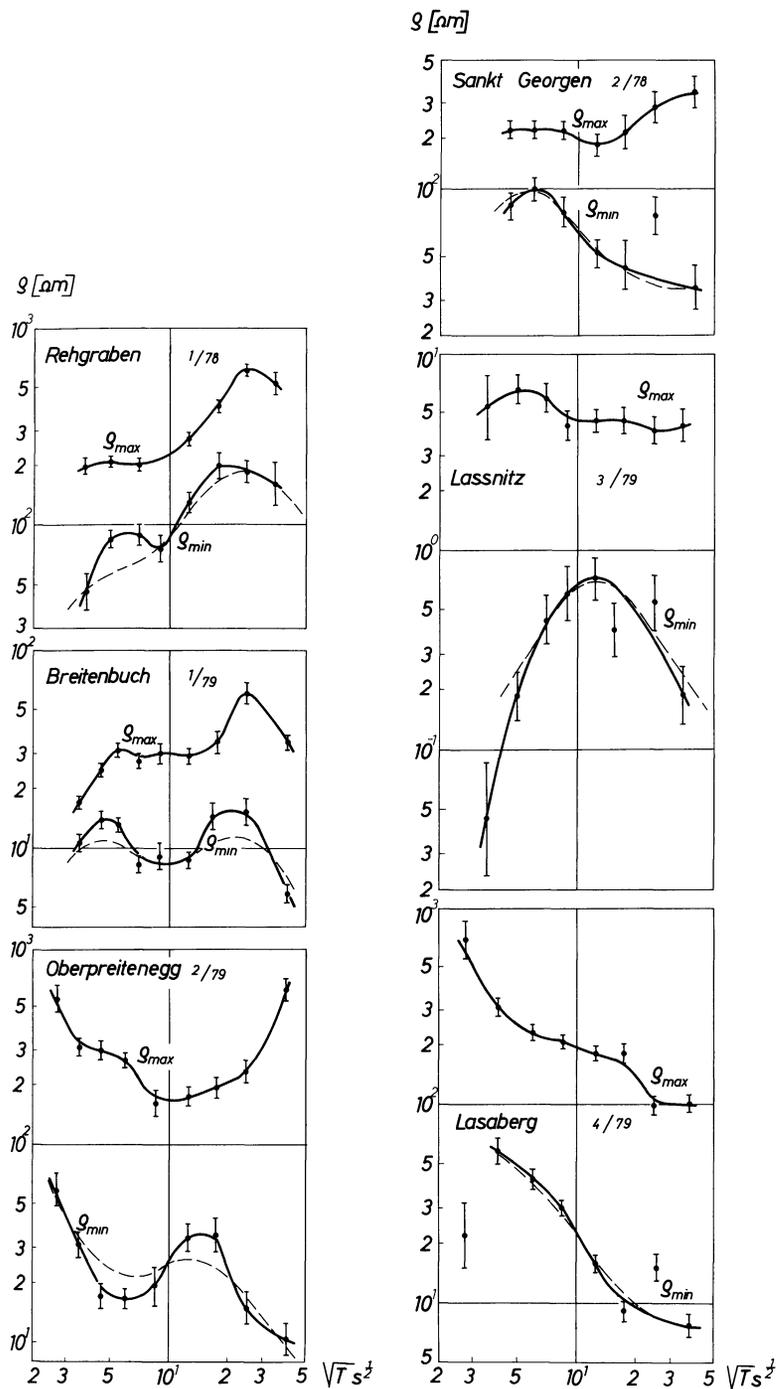


Fig. 2a, b. ρ_{min} and ρ_{max} sounding curves measured at the stations shown in Fig. 1

(a) Near to elongated high mountains bordering deep sedimentary basins with a sharp transition between, in the H -polarization case, the sounding curves obtained in the basin are distorted into ρ_{min} curves (i.e., the ρ_{min} direction is perpendicular to the boundary of the basin) and indicate apparent conductive layers (e.g., on the Kara-Kum platform in the vicinity of the Kopet Dagh mountains). This is the "edge effect" (Berdichevsky and Dmitriev 1976).

In our area an edge effect could be suspected first of all in the Graz Basin (Fig. 4). At point 1/78 the direction of Z_{xymin} (and of ρ_{min}) is, however, nearly parallel to the nearest part of the basin edge. So the edge effect can be excluded here.

West of Graz in the vicinity of point 1/79 the basin edge

turns by nearly 90° . Accordingly, in this area there are two directions of the basin edge, therefore the directions of ρ_{max} and ρ_{min} cannot be predicted on the basis of an edge effect. The other four measuring points lie in mountainous areas with thin sedimentary cover where the preconditions for the appearance of the edge effect are not valid.

(b) In Fig. 4 a correlation between fracture tectonics and the directions of Z_{xymin} may be inferred. The Z_{xymin} axes are perpendicular to well-known tectonic lines such as the Lavanttal-zone at point 2/79 and the Metnitz-Strassburg-St. Veit fracture zone, which is seismically active (Drimmel 1979), at point 3/79. At the site Lassnitz the resistivity values are extremely small, $0.03\text{--}5\ \Omega\text{m}$ for 20 s pulsations. A similar correlation may be

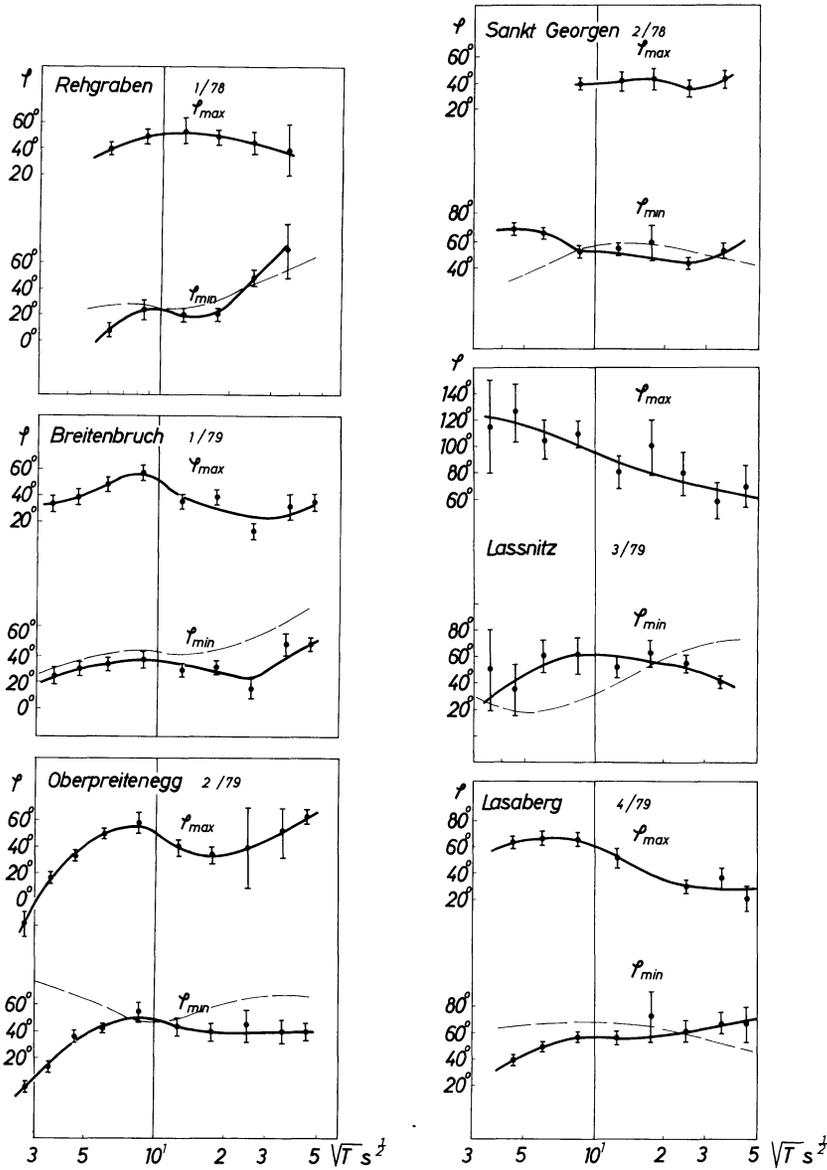


Fig. 3a, b. φ_{min} and φ_{max} phase curves measured at the stations shown in Fig. 1

inferred from the direction $Z_{xy_{max}}$ with the “Raab-Störung” in the Graz Basin (point 1/78) as well as some NNW-directed faults in the vicinity of point 1/79, the latter being nearly parallel to the Metnitz- and Lavanttal-fractures. In this area the heat flow reaches values as high as 100 mWm^{-2} possibly due to convective heat conduction through the fracture zones (see heat flow map by Čermák and Hurtig (1979)). The situation is similar to the geothermal conditions in the Pannonian Basin. The above-mentioned fracture zones may be considered as 2D structures.

A correlation between narrow fracture zones and magnetotelluric $Z_{xy_{max}}$ directions has already been observed in the Pannonian Basin (Ádám 1969; Stegena et al. 1971) and especially in its western part, i.e., in the area of the Transdanubian conductivity anomaly (Ádám 1976a, b, 1977) where more than 50 magnetotelluric soundings have been carried out. Figure 5 shows the tectonic structures and the directions of $Z_{xy_{max}}$ around Lake Balaton. There the directions are also perpendicular to the nearest fractures, and the $Z_{xy_{max}}$ axes are nowhere directed parallel to the strike direction of the fractures which are narrow in this area, whereas in the case of broad fracture zones such a parallelism is observed, as in the case of the 40 km broad Rhine-

Graben covered by thick sediments (Haak 1970; Reitmayer 1971).

2D model-calculations for the Transdanubian anomaly (Tátrallyay 1977) proved that in the case of a narrow ($\leq 10 \text{ km}$) two-dimensional conductive body embedded into a highly resistive medium, the direction of $Z_{xy_{max}}$ is not directed parallel to the direction of strike above the body. Consequently both outside and above the body the curve ρ_{min} corresponds to the case of E -polarization which best approximates the 1D layered model above the centre of the structure. Moving away from the centre, the depth to the top of the conducting body deduced from the ρ_{min} curves increases and so differs from the real one. Therefore it is called “apparent depth” (see one of Tátrallyay’s models in Fig. 6).

(c) If the “narrow fracture” model (Tátrallyay 1977) is covered by sediments of varying thickness, a galvanic distortion appears. This is the so-called S -effect. For its detailed description see Berdichevsky and Dmitriev (1976). A consequence of this effect is that the descending branch of the ρ_a curves in the case of H -polarization is shifted along the ρ axis. The extrema of the curves are found at the correct periods but at different appar-

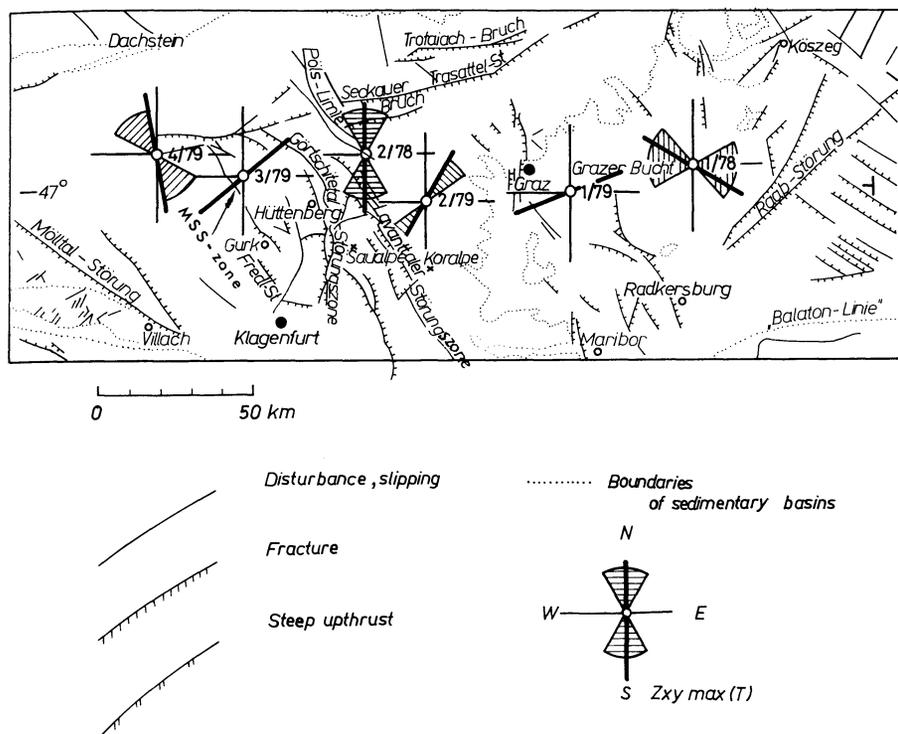


Fig. 4. Tectonic map of the southeastern part of Austria (after Tollmann 1970) with the major axes of the Z_{xy} polar diagrams, $Z_{xy\max}$ and the angular range of their variations with the period T (MSS-zone: Metnitz-Strassburg-St. Veit zone)

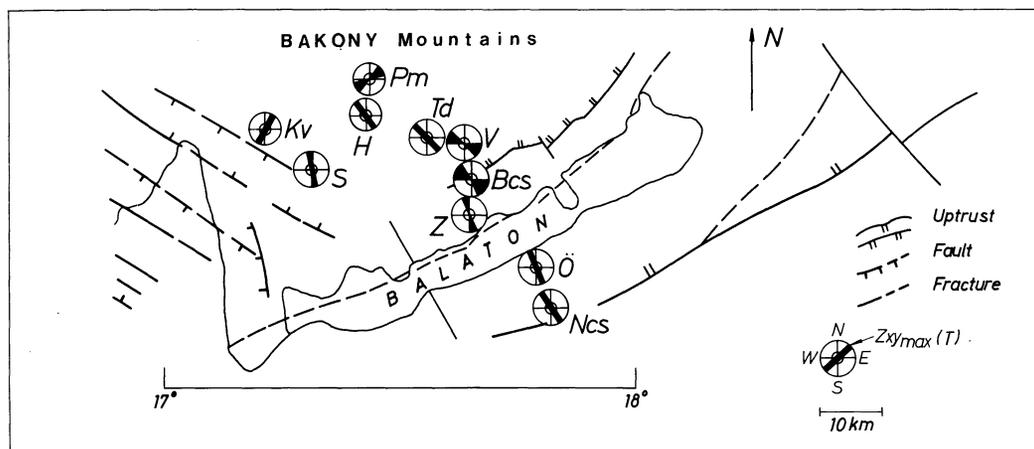


Fig. 5. Tectonic map of Transdanubia around Lake Balaton with the major axes of the Z_{xy} polar diagrams, $Z_{xy\max}$ and the angular range of their variations with the period T

ent resistivity levels. The parameters of the conductive layer calculated from the H -polarized ρ_a curves are in close correlation with the horizontal conductivity $S_1 = h_1/\rho_1$ of the sediment while the data (depth, horizontal conductivity) of the conductor derived from E -polarized ρ_a curves do not depend on S_1 values and surface geo-electric conditions. Accordingly, more reliable information about the depth of the conductive formation can be inferred from the E -polarized ρ_a curves.

On the basis of the above criterion the equivalence of the E -polarization and ρ_{\min} curves has been supported statistically for the Transdanubian MT-data (Ádám 1980) assuming that the sedimentary basin is preformed by the same fractures which contain the conductor. Some relevant results are summarized in Fig. 7a-c.

Figure 7a shows a plot of the apparent depths h of the conductive body from the curves ρ_{\min} and ρ_{\max} (1D models) against

the horizontal conductivity $S_1 = h_1/\rho_1$ of the uppermost layer. Depth values $h_{\rho_{\min}}$ calculated from the ρ_{\min} curves depend considerably less on S_1 than depth values $h_{\rho_{\max}}$ determined from ρ_{\max} curves. This is a good criterion for the E -polarization. The scatter of the depths $h_{\rho_{\min}}$ is also considerably smaller than scatter of $h_{\rho_{\max}}$.

Figure 7b shows the occurrence frequencies of depth values calculated for the same period range from ρ_{\min} and ρ_{\max} . The distribution of $h_{\rho_{\max}}$ has several smaller maxima in correspondence with surface geological conditions (thin sediments, Triassic or Permian rocks). This agrees with the conclusions drawn from Fig. 7a. The $h_{\rho_{\min}}$ values have two separate maxima at depths of about 7 km and 13 km, in accordance with the anomalous zones in the northern and southern Bakony Mountains.

Figure 7c shows the dependence of horizontal conductivity S of the conductive body on the depth of its top. S values

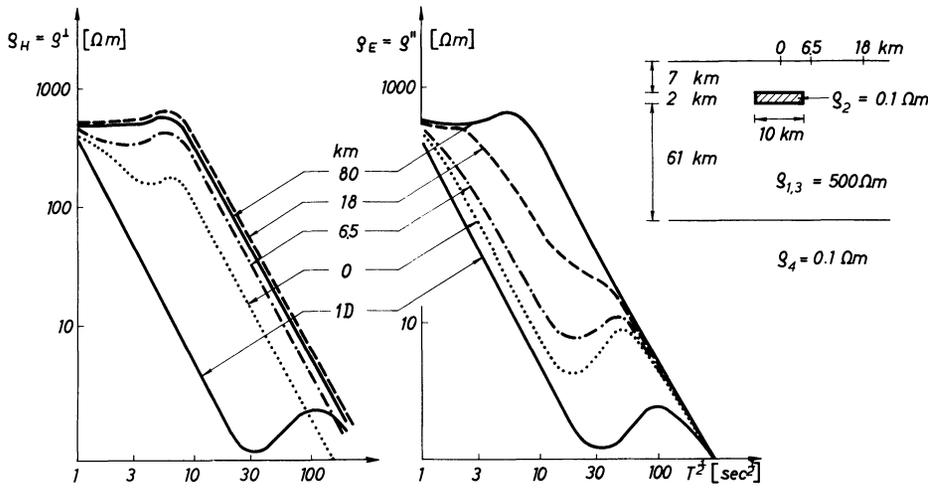
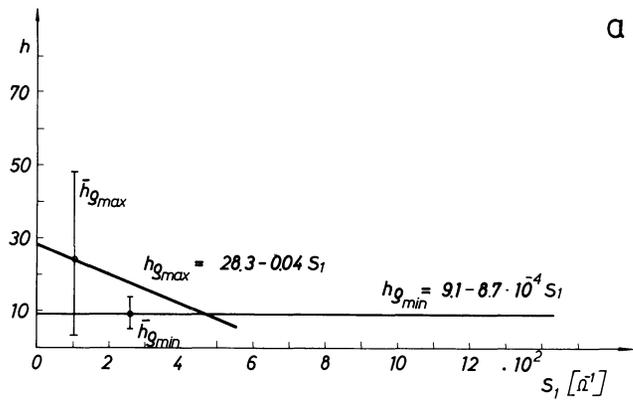
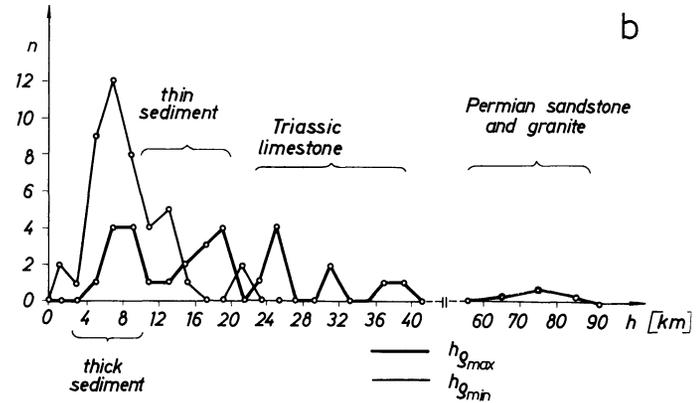


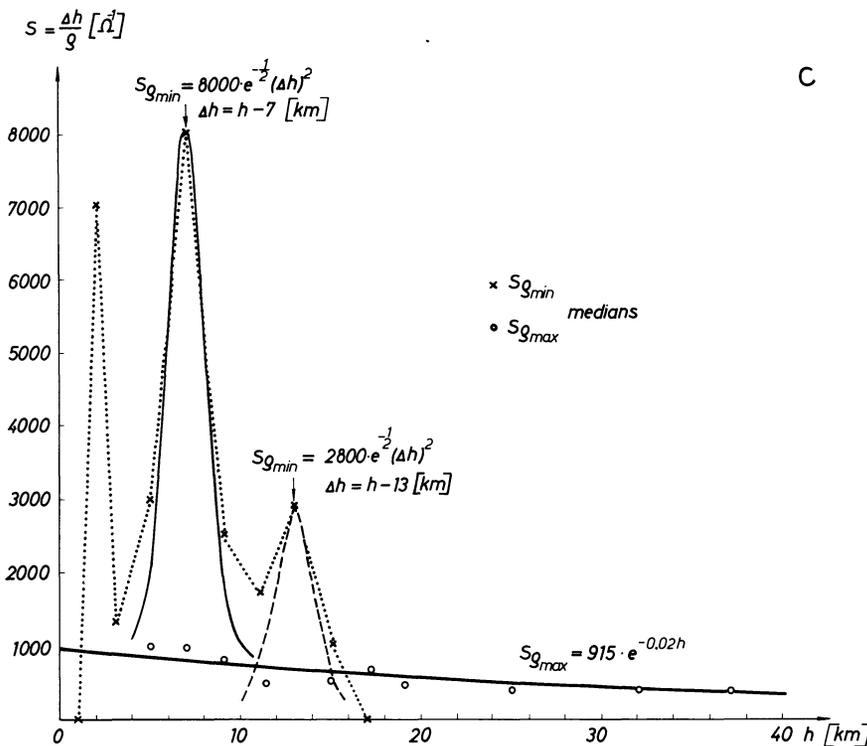
Fig. 6. Synthetic-magnetotelluric sounding curves over a model representing a narrow fracture zone with a conducting body at a depth of 7 km. 6.5, 18, 80 km are the distances of the “measuring point” from point 0. The 1D model shown for comparison has the same conductor depth and thickness (after Tatrallyay 1977)



a



b



c

Fig. 7 a-c. Statistical relationships between the geoelectric parameters of the Transdanubian conductivity anomaly (after (Adám 1980): **a** The depths of the conducting body calculated from ρ_{min} curves ($h_{\rho_{min}}$) and ρ_{max} curves ($h_{\rho_{max}}$) vs the horizontal conductivity of the uppermost layer, $S_1 = h_1/\rho_1$. **b** Occurrence frequencies of $h_{\rho_{min}}$ and $h_{\rho_{max}}$ values. For the peaks of $h_{\rho_{max}}$ the surface geologic formations are also given. **c** Horizontal conductivity S of the conducting layer vs depth of its top calculated on the basis of ρ_{min} curves ($S_{\rho_{min}}$) and ρ_{max} curves ($S_{\rho_{max}}$)

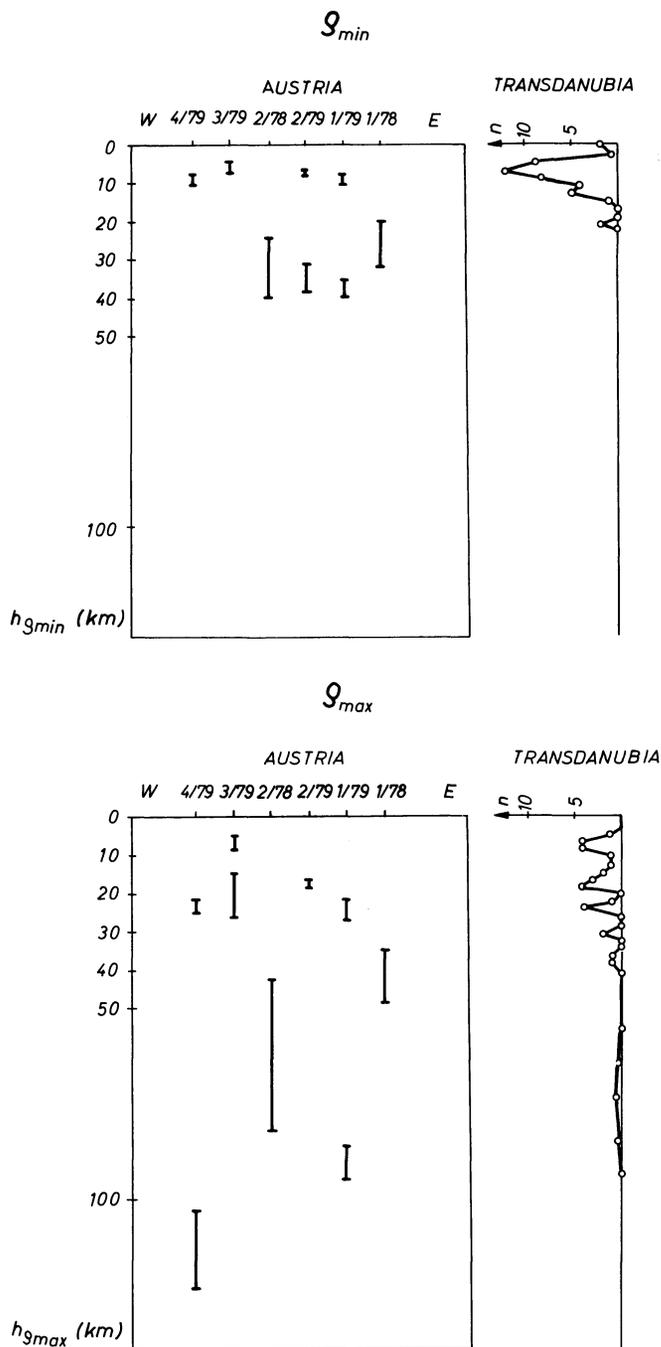


Fig. 8a, b. Depths of conducting bodies in the Styrian Alps with their error limits calculated on the basis of ρ_{min} and ρ_{max} curves. For comparison, the occurrence frequencies of $h_{\rho_{min}}$ and $h_{\rho_{max}}$ are also shown for Transdanubia

were determined by 1D model fitting. The values of $S_{\rho_{max}}$ decrease continuously (nearly exponentially) with the depth, while the distribution of the values $S_{\rho_{min}}$ has two peaks at depths of 7 km or 13 km corresponding to the real conducting zones of Fig. 7b. The peaks can be approximated by two functions of the form $S_0 e^{-\frac{1}{2}(\Delta h)^2}$. If the depth values do not coincide with the position of the peaks at 7 km and 13 km, the point is not above the centre of the anomalous body but above its flank according to Fig. 6. In summary we find again that ρ_{min} curves bear the real information about the conducting bodies in the fracture zones.

Interpretation

As the correlation between the fracture tectonics and the direction of $Z_{xy_{max}}$ is similar in the two regions compared above – Transdanubia and Eastern Alps – primary attention should be paid to the curves ρ_{min} . Although only six pairs of sounding curves were measured in Austria, the distribution of the depths of the conductive bodies is very similar to those found in the region of the Transdanubian anomaly (Fig. 8). Depth values calculated from ρ_{max} show considerably greater scatter than those calculated from ρ_{min} , especially in the second depth range below 30 km. The shift of $h_{\rho_{max}}$ towards greater depths corresponds to the S-effect.

According to the ρ_{min} curves the first depth range lies at 6–9 km. This agrees with the occurrence frequency peak of $h_{\rho_{min}}$ values in the region of the Transdanubian anomaly (Fig. 8). The curves ρ_{min} indicate a second conductive anomaly at depths of 30–40 km (Table 1).

Conclusions

The two regions, Transdanubia and the Eastern Alps, have several similar geophysical and tectonic features. The fractures in Transdanubia which divide the Hungarian Central Mountains into different parts are nearly parallel to the north-northwest directed fractures of Lavanttal and Metnitz. The fractures in Transdanubia are seismically active. In these zones the conductive formations are shallower, e.g., at a depth of 2–3 km at Bakonybél (Ádám 1976b). The heat flow in both regions is much greater than on the average. The earth's crust becomes thicker at the margin of the Eastern Alps – supposedly along the fractures – and also under the Bakony Mountains, i.e., under the central part of the Transdanubian conductivity anomaly. The earth's crust under the Bakony Mountains is thinner than 30 km whereas it amounts to 50 km in the Eastern Alps, the region of the present measurements. This may be why, in the Graz Basin and west of it (Lavanttal), an additional conductive zone exists in the earth's crust at 30–40 km depth as a result of the great heat flow, whereas in the Pannonian Basin the second conductive zone is indicated in the upper mantle at depths greater than 40 km (Ádám 1978).

On the basis of laboratory investigations, Duda et al. (1978) and Duda (personal communication 1980) confirmed the present interpretation of the Transdanubian anomaly: "The σ -anomaly is most likely caused by a geothermal fluid in the fractured region". This may also be valid for the conductive zones at about 7 km depth in the Eastern Alps.

Deep seismic investigations have found a low velocity zone in the same depth range (7–10 km), but only in the fracture areas. The deeper low-velocity zone appears between 20 km and 30 km (Gutdeutsch and Arič 1977). Accordingly, a close relation between the electric and seismic crustal models seems to exist only in the first depth range.

The present study is a first attempt to determine the conductivity structure of the transition zone between the Eastern Alps and the Pannonian Basin. It shows that a rough estimate of the depth of highly conductive layers can already be made. More detailed results may be reached in the future by additional MT soundings along the present profile and on profiles perpendicular to it.

Acknowledgements. The present magnetotelluric investigation is one of the projects of the "Österreichisch-Ungarische Gemischte Kommission für Wissenschaft und Technik". It is supported financially

by the Hungarian Academy of Sciences, the Bundesministerium für Wissenschaft und Forschung, Wien, and by the Austrian National Committee of the Geodynamics Project. We are gratefully indebted to all supporting institutions.

References

- Alpine Explosion Seismology Group: A lithospheric seismic profile along the axis of the Alps, 1975. I: First Results. *Pure Appl Geophys* **114**, 1109–1130, 1976
- Ádám, A., Major, L.: Stabilized high-sensitivity immersion magnetic variometer for magnetotelluric investigations (Type MTV-2). *Acta Geod. Geophys. Mont. Hung.* **2**, 211–215, 1967
- Ádám, A.: Appearance of the electrical inhomogeneity and anisotropy in the results of complex electrical exploration of the Carpathian Basin. *Acta Geod. Geophys. Mont. Hung.* **4**, 187–197, 1969
- Ádám, A.: Results of deep electromagnetic investigation in the Pannonian Basin. In: *Geoelectric and Geothermal Studies*, A. Ádám, ed.: pp 547–561. Budapest: Akadémiai Kiadó 1976a
- Ádám, A.: Distribution of the electrical conductivity in seismic (deep) fractures in Transdanubia. *Acta Geod. Geophys. Mont. Hung.* **11**, 277–285, 1976b
- Ádám, A.: The Transdanubian crustal conductivity anomaly. In: *Special Issue based on the 3rd Workshop on Electromagnetic Induction in the Earth*, held in Sopron, July 1976 (Contributed papers), A. Ádám, U. Schmucker, eds.: *Acta Geod. Geophys. Mont. Hung.* **12**, 73–79, 1977
- Ádám, A.: Geothermal effects in the formation of electrically conducting zones and temperature distribution in the Earth. *Phys. Earth Planet. Int.* **17**, 21–28, 1978
- Ádám, A.: Statistische Zusammenhänge zwischen elektrischer Leitfähigkeitsverteilung und Bruchtektonik in Transdanubien (Westungarn). *Acta Geod. Geophys. Mont. Hung.* **15**, 1980
- Berdichevsky, M.N., Dmitriev, V.I.: Distortion of magnetic and electric fields by near-surface lateral inhomogeneities. *Acta Geod. Geophys. Mont. Hung.* **11**, 447–483, 1976
- Berkthold A., Beblo, M., Kemmerle, K.: On the distribution of the electrical conductivity below the Eastern Alps. *Geol. Rundsch.* **65**, 715–732, 1976
- Cagniard, L.: Basic theory of the magnetotelluric method of geophysical prospecting. *Geophysics* **18**, 605–635, 1953
- Čermák, V., Hurting, E.: Heat Flow Map of Europe. Enclosure for *Terrestrial Heat Flow in Europe*. V. Čermák, L. Rybach eds. Berlin, Heidelberg: Springer-Verlag 1979
- Drimmel, J.: On the assessment of maximum earthquakes in the Alps and adjacent areas. *Tectonophysics* **55**, 1–5, 1979
- Duba, A., Piwinskii, A.J., Santor, M., Weed, H.C.: The electrical conductivity of sandstone, limestone and granite. *Geophys. J. R. Astron. Soc.* **53**, 583–597, 1978
- Gutdeutsch, R., Arič, K.: Eine Diskussion geophysikalischer Modelle des Grenzbereiches zwischen Ostalpen und Pannonischem Becken. *Acta Geologica Hung.* **XXI**, **4**, 287–296, 1977
- Haak, V.: Das zeitlich sich ändernde, erdelektrische Feld, beobachtet auf einem Profil über den Rheingraben: eine hiervon abgeleitete Methode der Auswertung mit dem Ziel die elektrische Leitfähigkeit im Untergrund zu bestimmen. Dissertation, 140 pp. 1970
- Reitmayr, G.: Räumliche Verteilung und Periodenabhängigkeit der Vorzugsrichtung des erdelektrischen Feldes und erste Ergebnisse der Auswertung von magnetotellurischen Messungen auf einem Profil über den Rheingraben. Protokoll über das Kolloquium „Erdmagnetische Tiefensondierung“ in Rothenberge/Westf. vom 14–16. September 1971 P. Weidelt, ed.: pp. 155–177, 1971
- Stegena, L., Horváth, F., Ádám, A.: Spreading Tectonics Investigated by Magnetotelluric Anisotropy. *Nature* **231**, 442–443, 1971
- Tátrallyay, M.: On the interpretation of EM sounding curves by numerical modelling using the S.O.R. method. In: *Special Issue based on the 3rd Workshop on Electromagnetic Induction in the Earth*, held in Sopron, July, 1976. (Contributed papers) A. Adam, v. Schmucker, eds. *Acta Geod. Geophys. Mont. Hung.* **12**, 279–285, 1977
- Tollmann, A.: Die bruchtektonische Zyklenordnung im Orogen am Beispiel der Ostalpen. *Geotekt. Forsch.* **34**, I–II, 1–90, 1970
- Verö, J.: On the determination of magnetotelluric impedance tensor. *Acta Geod. Geophys. Mont. Hung.* **7**, 333–351, 1972

Received December 10, 1980; Revised version June 15, 1981
Accepted June 19, 1981

Low Energy Solar Particles Observed During the Pre-Phase of Large Particle Events*

E. Kirsch, N. Martinic, and E. Keppler

Max-Planck-Institut für Aeronomie, D-3411 Katlenburg-Lindau 3, Federal Republic of Germany

Abstract. Solar proton and electron measurements ($E_p \geq 80$ keV, $E_e \geq 20$ keV) of the Helios 1 and 2 spacecraft are used together with additional information published in the Solar Geophysical Data to study the pre-phase particle emission occurring $\lesssim 5$ days before the large particle events in the same McMath region. Emphasis is laid on the question of whether both suggested particle acceleration steps in the pre-phase actually take place. It was found that about 60% of all pre-phase events are associated with Type II/IV radio-bursts, the general proton indicator. The remaining 40% of the pre-phase events were associated only with small flares and a group of Type III bursts. However the proton/electron ratio, which is ≥ 10 and ≥ 100 , at ~ 0.1 MeV and ~ 1.0 MeV particle energy, respectively, as well as the maximum particle energies reached indicate that the second (Fermi-type) acceleration process has taken place also for these events. The first and second acceleration process were not separable and it must be concluded that the Type III associated events are independent and are not a precursor of the main event.

Key words: Solar particle events – Radiobursts – Acceleration process

Introduction

Large cosmic ray events are generated in highly active solar regions which release about 10^{32} erg in 10^3 s when one or more optical flares occur, together with microwave and X-ray emission. It has been suggested that the acceleration of solar particles up to relativistic energies takes place in two different acceleration steps (de Jager 1969; Svestka and Fritzova-Svetskova 1974; Lin 1974; Svestka 1976; Ramaty 1979; and others). The first step, with time scales of 10–100 s, also called the flash or impulsive phase, would primarily produce electrons in the energy range 10–100 keV; effects of these accelerated electrons include the accompanying bremsstrahlung photons.

The second stage then is thought to generate relativistic protons and electrons and is accompanied by Type II/IV radio bursts. Particle acceleration in the second stage has been explained in terms of second order Fermi mechanism (Ramaty 1979), second order transit time damping (Fisk 1976), and the first order Fermi mechanism of shock acceleration (Axford et al. 1977; Fisk et al. 1980). The former acceleration processes are related to the formation of the Alfvén (Fermi) and the magneto-

sonic (transit time) wave spectrum that scatters particles in momentum space. Shock acceleration occurs at the shock discontinuity when “seed” particles perform multiple crossings at the shock front.

Type II, IV radio bursts are produced by shock waves interacting with both the solar atmosphere and the interplanetary plasma and generally indicate the emission of protons from the sun (Warwick 1962; Croom 1971; Akin'yan 1977). The Type III radio emission is associated with nearly relativistic electron streams interacting with the corona and interplanetary plasmas (Fainberg and Stone 1974), the electrons possibly being accompanied by protons (Svestka 1976). Type III radio bursts often occur without visible manifestation of flares.

A different acceleration process may be responsible for Type III bursts and the steady (hours to days) production of low energy (< 1 MeV) ions (Rust and Emslie 1979).

The purpose of the present study is to examine the pre-phase of large particle events, whenever enhanced, and to investigate the physical configuration of the solar emission compared to the main event. We use low energy ($E_p > 80$ keV, $E_e > 20$ keV) ion and electron measurements on the Helios spacecrafts within

Table 1. Solar particle events

Main Event	Pre-emission	Pre-emission with Type II/IV	Pre-emission with Type III
28 March 76	1	1	0
5 Sept. 77	1	1	0
22 Nov. 77	3	1	2
27 Dec. 77	3	3	0
1 Jan. 78	1	1	0
13 Feb. 78	4	2	2
25 Feb. 78	2	2	0
6 March 78	1	0	1
8 April 78	2	1	1
11 April 78	—	—	—
7 May 78	3	2	1
11 May 78	1	1	0
10 Nov. 78	3	1	2
11 Dec. 78	1	1	0
24 Jan. 79	1	0	1
16 Feb. 79	0	0	0
1 March 79	—	—	—
9 March 79	—	—	—
3 April 79	2	0	2
13 April 79	—	—	—
Sum: 20	29 (100%)	17 (~60%)	12 (~40%)

* A preliminary version has been presented at the 7th European Cosmic Ray Symposium, Leningrad, Sept. 1980

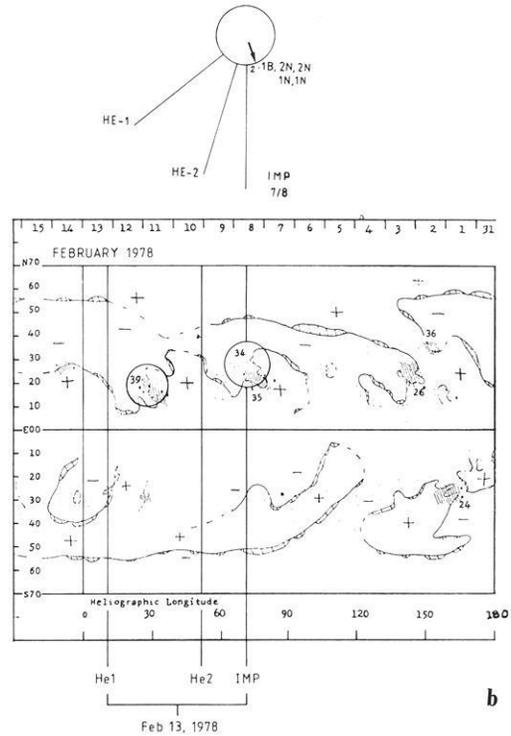
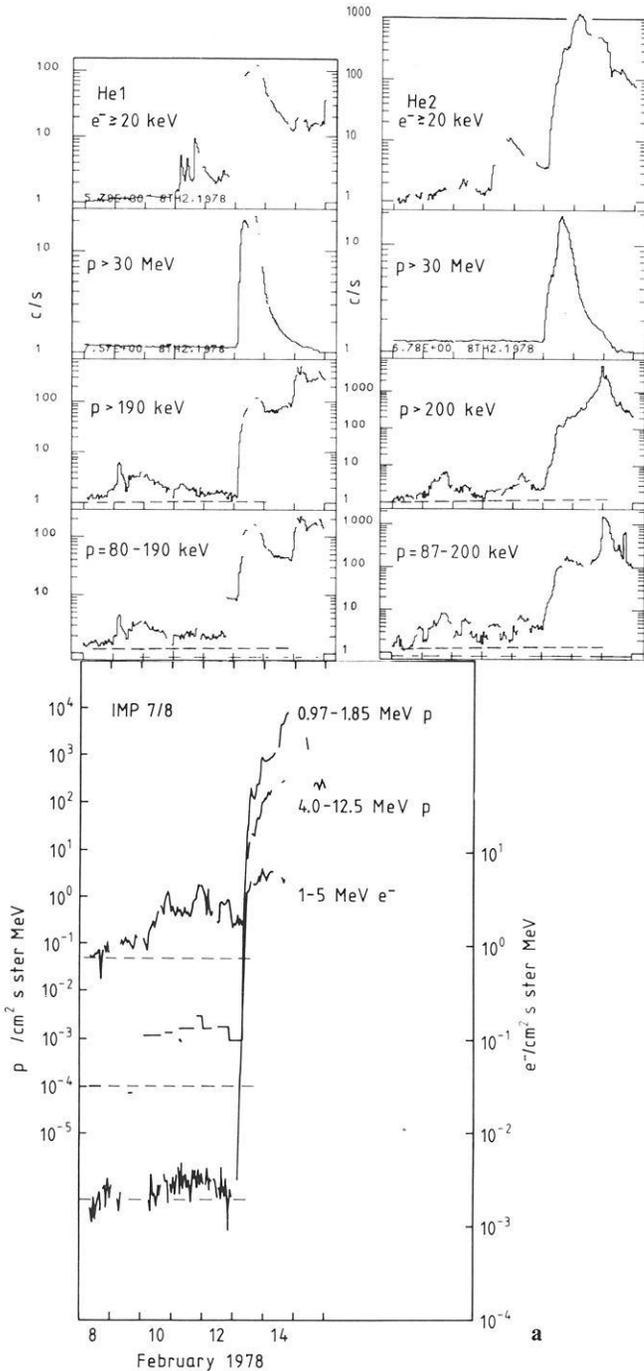


Fig. 1a-c. Data for February 1978. **a** Proton-electron measurements of Helios 1/2 (*upper part*) and of IMP 7/8 (*lower part*) from Feb. 8–15; **b** Positions of Helios 1/2 and of IMP 7/8 with respect to the Sun for Feb. 13 (*upper part*) and the H_{α} synoptic chart (*lower part*). The *vertical lines* indicate the foot points of the field lines which are connected to the three spacecrafts. **c** SMS-GOES X-ray measurements from February 8–13. The *horizontal bars* in the *upper part* represent Type III, II, IV – radio bursts and the flares in the McMath regions 15139, 15134

1 AU solar distance, as well as the IMP 7/8 satellites ($E_{p,e} \gtrsim 1$ MeV) particle data, X-ray and radio observations as published in the Solar Geophysical Data.

The preflare enhancements before solar cosmic ray events have been studied earlier by Kuzhevskiy and Chupova (1977) and Block and Kuzhevskiy (1979, 1980).

The physical quantities of interest are: association with Type II, III and IV radio bursts, X-ray emission, the proton to electron number ratio, energy spectrum, and the maximum energy of the accelerated particles. The criteria used to select the pre-emission events are as follows:

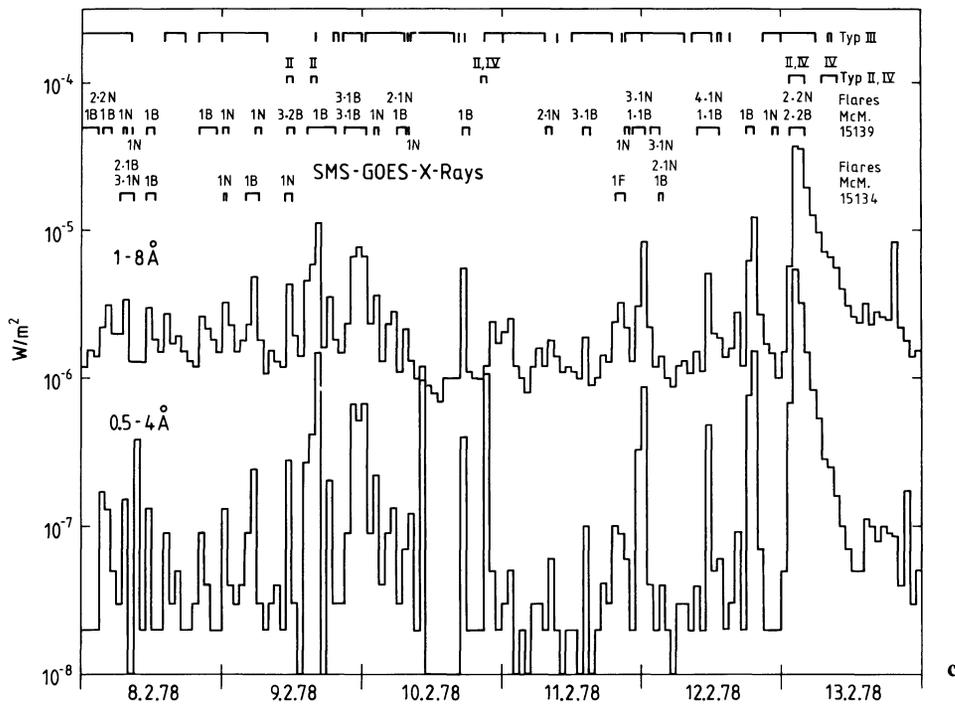
a) Identification of the flare in the same McMath plage region as the main event.

b) The thresholds used for the pre-enhancements are: (p for protons, e^- for electrons) Helios 1/2: ≥ 10 p/cm² s sr ($E_p \geq$

80 keV), ≥ 5 p/cm² s sr ($E_p > 190$ keV), ≥ 5 e⁻/cm² s sr ($E_e \geq 20$ keV) IMP 7/8: $\geq 5 \cdot 10^{-3}$ p/cm² s sr MeV ($E_p = 0.97$ –1.85 MeV), $\geq 1.0 \cdot 10^{-3}$ e⁻/cm² s sr MeV ($E_e = 1$ –5 MeV) or generally $\sim 20\%$ above the background.

c) The reported pre-events are “impulsive”, i.e. the particle data exhibit a sudden increase, a maximum and decay phase. (‘Non impulsive’ events not associated with flares have been discussed by Zwickl and Roelof, in press 1981). Both the main event and these impulsive ones take place in the same active region. The east-west solar scans (Solar Geophysical Data) at 3 and 10.7 cm wave lengths help to identify the heliographic longitude of the source region, discriminating from other solar active regions.

d) The particle injections take place under negligible coronal propagation. The foot points of the magnetic field lines are



connected to the 'fast propagation region' (Reinhard and Wibberenz 1974).

e) The pre-phase period should not be disturbed by shocks in the interplanetary space produced by other flares.

f) For the interplanetary transport process of the low energy protons and electrons we assume the classical models: diffusion parallel to the smooth interplanetary magnetic field with fewer scattering processes taking place between onset and maximum phase of the particle events and isotropisation in the late phase of the events (compare Fisk and Axford 1968; Völk 1975; Earl 1976; Gombosi and Owens 1980 and others).

Experiment Description

The charged particle spectrometer flown on Helios 1 and 2 uses an inhomogeneous magnetic field to separate ions and electrons, three single semiconductor surface barrier detectors for electron detection and a telescope consisting of two similar detectors for ions. The latter have a geometric factor of $1.5 \cdot 10^{-2} \text{ cm}^2 \text{ sr}$ for ions and the former $< 5 \cdot 10^{-2} \text{ cm}^2 \text{ sr}$ for the electrons. The experiment measures ions and electrons in 16 energy channels and 16 sectors in the ecliptic plane. The energy range covered is 20 keV–2 MeV for electrons and 80 keV–>750 keV for ions if they are interpreted as protons. For a detailed description of the instrument see Keppler et al. (1977).

The earth orbiting satellites IMP 7/8 are able to measure particles with $E > 0.2 \text{ MeV}$ using solid state detector telescopes as described in Solar Geophysical Data.

Observations

In Table I are listed 20 solar particle events from 1976–1979. They are characterized by the presence of $> 30 \text{ MeV}$ protons. The period of our observations correspond to the quiet period up to the maximum of the solar activity of the cycle 21 as measured by the sunspot numbers. About 75% of these events were associated with pre-enhancements, observed by at least one of the three spacecraft described above. This fraction is

consistent with the value obtained by Block and Kuzhevskiy (1979) during other periods and events. The 16 proton events exhibit 29 individual pre-emission cases; of which 17, or $\sim 60\%$, are associated with Type II/IV radio bursts. The rest, i.e. $\sim 40\%$, are associated with the Type III bursts.

We concentrate on the 40% of cases of enhancements associated with Type III radio bursts, that generally accompany small flares. The other events, i.e. 60% of the pre-phase enhancements associated with Type II/IV radio bursts, are being considered as the classical two-step acceleration mechanism and likely analog to the main particle event. We point out that Type II/IV associated events show often Type III radio bursts in addition.

Particle measurements from Helios and IMP probes during 8–15 February 1978 are presented in Fig. 1a. The main event starts on 13 February. In this illustration the time resolution is taken equal to 1 h. The proton-electron injections during the pre-phase of the main event can be seen above the (dashed line) background fluxes. No shocks are observed with the plasma experiment of the Helios 1, 2 during these pre-phase events (Dr. R. Schwenn, private communication). During the same interval no sudden commencements are reported (Solar Geophysical Data).

The maximum energy of the pre-emission events reach up to 20 MeV for protons and up to $> 1 \text{ MeV}$ for electrons. In Fig. 1b (the upper panel) are sketched the locations of the three spacecraft at the onset of the main event and the flares of the sun. The IMP satellites are located at 1 AU helioradial distance, the Helios spacecraft within 1 AU. The lower part of Fig. 1b illustrates the H_α -Synoptic chart with the coronal connecting longitudes of the three probes, calculated according to the mapping technique (Nolte and Roelof 1973), for 13 February. Also shown are the two McMath regions 14139 and 15134; their activity was triggered between 8 February and 13 February. The east-west solar scan reveals that it is quite probable that the main particle event as well as the pre-emission originated in McMath region 15139.

Figure 1c shows the X-ray measurements of the SMS-GOES satellite (0.5–4 and 1–8 Å) during 8–13 February. The upper

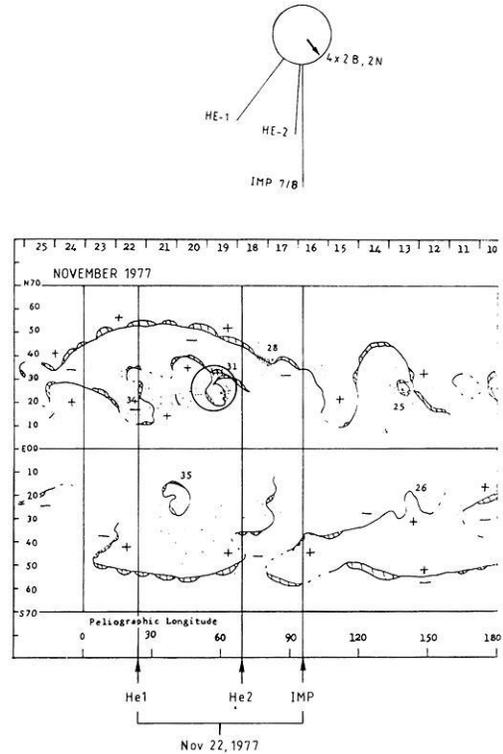
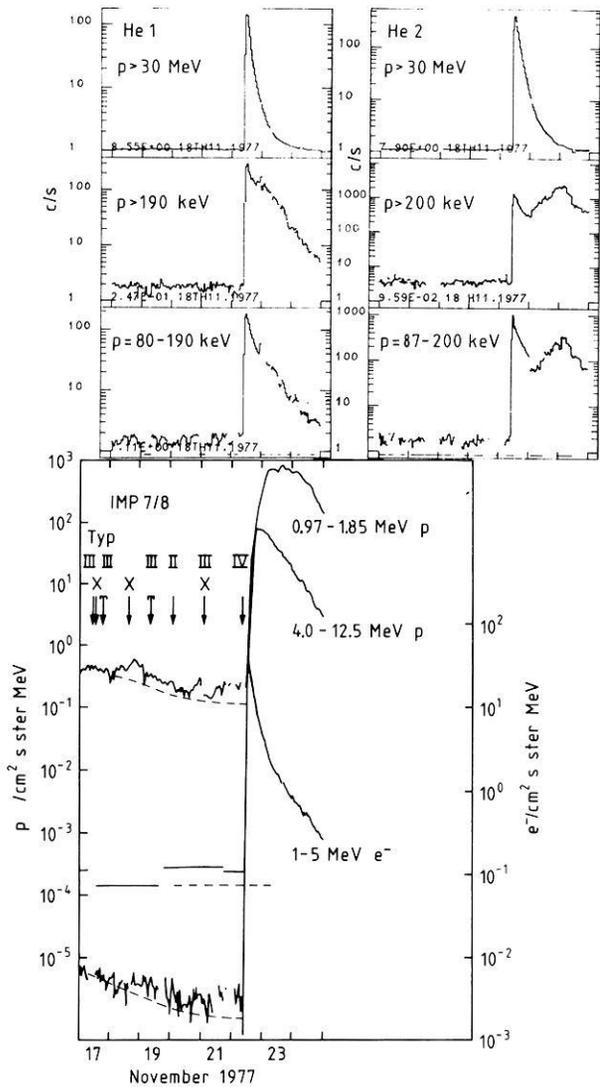


Fig. 2a, b. Data for November 1977. **a** Proton measurements of Helios 1/2 (upper part) and proton-electron measurements of IMP 7/8 (lower part) from Nov. 17–24; **b** As Fig. 1b but for Nov. 22, 1977

part, indicated by horizontal bars, exhibits the observed Type III, II and IV radio bursts and the type of the flares that have taken place in the active regions 15139 and 15134. It can be seen that not all Type III bursts are associated with distinct X-ray bursts. Only those particle increases in Fig. 1a which are associated with Type III bursts will be considered further. We note that the electron flux increases on 11–12 February shown in Fig. 1a are associated with Type III bursts. The IMP satellites detected small fluxes of >1 MeV electrons. Also, a proton to electron ratio at ~ 1 MeV energy of 500–1,000, for the pre-phase of these events could be determined. The Helios 1, 2 observations show that the proton to electron ratio at ~ 100 keV is already >10 , described by different energy channels of the experiment. The maximum energies of the accelerated protons and electrons during this pre-phase are ~ 20 MeV and >1 MeV respectively, seen by the IMP measurements. The fitted differential power law spectrum obtained from the Helios output is of the order of $\gamma \sim -3$ for the exponent, with small fluctuations during the interval we discuss (compare Block and Kuzhevskiy 1979; Fisk and Lee 1980, who expect exponential spectra from the transit time acceleration as well as from the shock acceleration mechanisms).

A second illustration is shown in Fig. 2a where the 22 November 1977 event is the main reference. The pre-emission of low energy protons and electrons is small (electrons are not

shown in the Figure) as measured by Helios 1, 2. The IMP satellites however reveal particle enhancements. The same event has been studied by Block and Kuzhevskiy (1979, 1980). They found proton and electron pre-emission as well as an indication of a hardening of the spectrum. The 22 November event occurred at the beginning of the present solar cycle.

The arrows in Fig. 2a (lower part) show the detected bursts of the X-, Type II, III and IV radiation. During the pre-phase of this event the X-ray emission was mostly below the level of detection. With the exception of 20 November, when a Type II radio burst occurred, most of the pre-emission events were associated with Type III bursts. Figure 2b illustrates the positions of the spacecraft, the flares at the Sun on 22 November and the H_{α} -synoptic chart with the calculated foot points of the field lines at the corona which are connected to the satellites. It can be seen that only one McMath active region could produce the pre-phase and the main particle emission. A special feature during 22 November is the hard X- and γ -ray emission which indicate that nuclear reactions have taken place in the solar photosphere (Chambon et al. 1978). Composition measurements for the main event (on 22 November 1977, as well as on 13 February 1978) have been published by McGuire et al. (1979). No evidence was found of low energy ^3He enrichment or of ^2H , or ^3H .

The maximum energies reached by protons and electrons

during the pre-phase of 22 November event amount to < 13 MeV and ~ 1 MeV, respectively. The proton-to-electron ratio at ~ 1 MeV particle energy is ~ 100 as obtained from the IMP measurements.

We summarise our observations noting that the illustrative examples given above can be seen as typical of the so-called pre-phase events.

1) About 60% of the particle enhancements preceding large solar particle events are associated with small flares occurring in the same McMath activity region as the main event and Type II, IV radio bursts. The remaining 40% of the pre-phase events, examined in the present study, are associated with small flares and a group of Type III radio bursts.

2) Such pre-phase proton events are associated with X-rays and quasi relativistic electrons (sometimes up to > 1 MeV).

3) The proton-to-electron number ratio for ~ 0.1 MeV particles amounts to ~ 10 , reaching > 100 at ~ 1 MeV energies.

4) The maximum energies of accelerated protons and electrons are < 20 MeV and > 1 MeV, respectively.

5) The shock waves possibly associated with such pre-phase events are below the level of detection. Sometimes the equatorial Dst-value of the geomagnetic field shows small disturbances which could result from flares producing the Type III bursts.

Discussion

The presented study has shown that small flares occurring in the pre-phase (< 5 days) of large particle events can obviously accelerate electrons and protons up to energies > 1 MeV and < 20 MeV, respectively in association with Type III bursts and X-rays without, however, detectable Type II, IV bursts and also without shocks. It has been suggested that solar particles are accelerated in two stages to high energies, the first one leads to ~ 1 MeV energy, the second one up to relativistic energies (de Jager 1969; Lin 1974; Svestka 1976; Lin and Hudson 1976; Bai and Ramaty 1976, compare also Emslie and Rust 1980). The time resolution (1 h) of the X-ray and particle data (the higher time resolution of the Helios experiment could not be used due to the small particle fluxes) give no direct evidence for the second acceleration stage. However, the maximum particle energies reached and the proton-to-electron ratio in particular indicate that the second acceleration process has taken place. In a recent paper Ramaty (1979) has treated analytically the second order Fermi-type acceleration process and found that more protons than electrons are accelerated by the second acceleration stage. Because a Type II/IV radio burst and a direct shock wave in the solar wind was not detectable during the events presented here, the role of an accompanying shock wave is still unclear. Since individual small flares in the pre-phase of larger particle events can accelerate protons and electrons to moderate energies we consider all pre-phase events as independent events and not as the precursor of the main event. The fact that the particle energy increases the nearer the main event comes in time (Block and Kuzhevskiy 1979, 1980) could be explained by larger and more efficient flares occurring in the same McMath region.

Kahler (1979) studied the pre-flare characteristic < 20 min before small flares in soft X-rays and concluded that coronal pre-flare heating lasts not longer than 2 min indicating again that all flares are independent events. The pre-flare events seem to have no influence on the composition of the main events as can be seen from the measurements published by McGuire et al. (1979).

Since Type III bursts are more common than proton events

and not always related to observable flares it can be concluded that electrons are temporarily stored in closed coronal field configurations (corresponding to a height of 4–6 solar radii) and are released by a still unknown mechanism. McGuire et al. (1975) concluded that a rest matter of $5\text{--}10 \mu\text{g}/\text{cm}^2$ allows a storage of electrons for > 5 days in the corona, a value which should also be valid for the pre-phase events. Stored protons seem to lose their energy rapidly due to their stronger ionisation effects with the rest matter in the acceleration region. Krimigis (1972) calculated storage times of < 1 day for low energy protons ($E = 0.1\text{--}0.3$ MeV) and coronal densities of $4 \cdot 10^5/\text{cm}^3\text{--}3 \cdot 10^4/\text{cm}^3$ (3–6 solar radii). It is noted here that McGuire et al. (1975) have suggested the continual acceleration of low energy protons and electrons in “active solar regions” which may be different from the acceleration process operating in the pre-phase of large particle events. Considering the energy spectrum, the maximum energies reached and the association with Type III bursts, we find that the pre-emission events can best be compared with the so-called micro events studied by McDonald and van Hollebeke (1972). They found that $\sim 20\%$ of the micro events are associated with Type II and IV radio bursts and $\sim 80\%$ with a group of Type III bursts only.

As a side result we note that the empirical formulas relating duration and intensity of the micro-wave burst and the expected proton flux near the Earth derived by Croom (1971) and Akin'yan et al. (1977) cannot be applied to the pre-phase events because they predict too high fluxes. The relation between X-ray intensity and ~ 1 MeV protons (Gold and Roelof 1977) also predicts too high fluxes.

Acknowledgements. This work has been supported partly by the Max Planck Gesellschaft zur Förderung der Wissenschaften e.V. and by the Bundesminister für Forschung und Technologie under Contracts RS-12, WRS 10/7 and WRS 10/8.

References

- Akin'yan, S.T., Fomichev, V.V., Chertok, I.M.: Determination of the parameters of solar protons in the neighborhood of the Earth from radio bursts. 1. Intensity function. *Geomagn. Aeron.* **17**, 5–8, 1977
- Axford, W.I., Leer, E., Skadron, G.: The acceleration of cosmic rays by shock waves. 15th International Cosmic Ray Conference, Vol. **5**, 132–137, Plovdiv, Bulgaria: 1977
- Bai, T., Ramaty, R.: Gamma ray and micro wave evidence for two phases of acceleration in solar flares. *Solar Phys.* **49**, 343–358, 1976
- Block, G.M., Kuzhevskiy, B.M.: Preflare enhancements of low energy proton intensity in the solar cosmic ray events. 16th International Cosmic Ray Conference, Vol. **5**, 1–5, Kyoto, Japan: 1979
- Block, G.M., Kuzhevskiy, B.M.: Electrons in the preflare rise phase of solar cosmic rays. *Geomagn. Aeron.* (in Russian) **3**, 538–541, 1980
- Chambon, G., Hurley, K., Niel, M., Talon, R., Vedrenne, G., Likine, O.B., Kouznetsov, A.V., Estouline, J.V.: A hard X-ray and gamma ray observation of the 22 November 1977 solar flare. In: NASA Technical Memorandum 79619, Gamma ray spectroscopy in Astrophysics, T.L. Cline and R. Ramaty, eds.: pp. 70–75, 1978
- Croom, D.L.: Forecasting the intensity of solar proton events from the time characteristics of solar micro wave bursts. *Solar Phys.* **19**, 171–185, 1971
- Earl, J.A.: Non diffusive propagation of cosmic rays in the solar system and in extragalactic radio sources. *Astrophys. J.* **206**, 301–311, 1976
- Emslie, A.G., Rust, D.M.: Energy release in solar flares. *Solar Phys.* **65**, 271–281, 1980

- Fainberg, J., Stone, R.G.: Satellite observations of Type III solar radio bursts at low frequencies. *Space Sci. Rev.* **16**, 145–188, 1974
- Fisk, L.A., Axford, W.I.: Effect of energy changes on solar cosmic rays. *J. Geophys. Res.* **73**, 4396–4399, 1968
- Fisk, L.A.: The acceleration of energetic particles in the interplanetary medium by transit time damping. *J. Geophys. Res.* **81**, 4633–4640, 1976
- Fisk, L.A., Lee, M.A.: Shock acceleration of energetic particles in corotating interaction regions in the solar wind. *Astrophys. J.* **237**, 620–626, 1980
- Gold, R.E., Roelof, E.C.: A prediction technique for low energy solar proton fluxes near 1 AU. In: *COSPAR Space Research XVI*, M.J. Rycroft, ed., pp 791–796. Berlin: Akademie-Verlag 1976
- Gombosi, T.I., Owens, A.J.: The interplanetary transport of solar cosmic rays. *Astrophys. J.* **241**, L129–L132, 1980
- De Jager, C.: Solar flares, properties and problems, 1–15, In: “Solar flares and Space Research”, C. De Jager and Z. Svestka, eds. Amsterdam: North-Holland Publishing Company 1969
- Kahler S.W.: Preflare characteristics of active regions observed in soft X-rays. *Solar Phys.* **62**, 347–357, 1979
- Keppler, E., Richter, A.K., Richter, K., Umlauf, G., Wilken, B., Williams, D.J.: A survey on measurements of medium energy protons and electrons obtained with the particle spectrometer E8 on board of Helios. *J. Geophys.* **42**, 633–644, 1977
- Krimigis, S.M.: A lower limit to the altitude of coronal particle storage regions deduced from solar proton energy spectra. In: *High energy phenomena on the sun*, NASA SP 342, 478–485, 1972
- Kuzhevskiy, B.M., Chupova, L.M.: Relationship between preflare enhancements of solar cosmic-ray intensity and the X-ray emission from the sun. *Geomagn. Aeron.* **17**, 223–224, 1977
- Lin, R.P.: Non-relativistic solar electrons. *Space Sci. Rev.* **16**, 189–256, 1974
- Lin, R.P., Hudson, H.S.: Non thermal processes in large solar flares. *Solar Phys.* **50**, 153–178, 1976
- McDonald, F.B., van Hollebeke, M.A.: Solar cosmic ray micro-events in High energy phenomena on the sun. *NASA SP 342*, 404–417, 1972
- McGuire, R.E., Lin, R.P., Anderson, K.A.: Observation of low energy electrons and protons from solar active regions. 14th International Cosmic Ray Conference, Vol. **5**, 1785–1790, Munich, Germany, 1975
- McGuire, R.E., von Roseninge, T.T., McDonald, F.B.: A survey of solar cosmic ray composition 1974–1978. 16th International Cosmic Ray Conference, Vol. **5**, 61–66, Kyoto, Japan, 1979
- Nolte, J.T., Roelof, E.C.: Large-scale structure of the interplanetary medium. I. High coronal longitude of the quiet-time solar wind. *Solar Phys.* **33**, 241–257, 1973
- Ramaty, R.: Energetic particles in solar flares. In: *Particle Acceleration Mechanisms*, J. Aarons, C. Max, C. McKee, eds.: pp 135–154. American Institute of Physics 1979
- Reinhard, R., Wibberenz, G.: Propagation of flare protons in the solar atmosphere. *Solar Phys.* **36**, 473–494, 1974
- Rust, D.M., Emslie, A.G.: World Data Center A for Solar Terrestrial Physics, Energy release in solar flares. Proceedings of the Workshop on Energy Release in Flares, 26 February – 1 March 1979, Cambridge, Massachusetts, U.S.A., Report UAG-72, 11–15, 1979
- Svestka, Z., Fritzoza-Svestkova, L.: Type II radio bursts and particle acceleration. *Solar Phys.* **36**, 417–431, 1974
- Svestka, Z. In: *Solar flares*. B.M. McCormac, ed.: pp 238–349. Dordrecht, Holland: Reidel Publ. Company 1976
- Völk, H.J.: Cosmic ray propagation in interplanetary space. *Rev. Geophys. Space Phys.* **13**, 547–565, 1975
- Warwick, C.S., Haurwitz, M.W.: A study of solar activity associated with polar-cap absorption. *J. Geophys. Res.* **67**, 1317–1332, 1962
- Zwickl, R.D., Roelof, E.C.: Interplanetary propagation of <1 MeV protons in non-impulsive energetic particle events. *J. Geophys. Res.*, in press 1981

Received March 5, 1981; Revised Version July 22, 1981

Accepted July 22, 1981

Characteristics of Westward Travelling Surges During Magnetospheric Substorms

W.G. Tighe and G. Rostoker

Institute of Earth and Planetary Physics and Department of Physics, University of Alberta, Edmonton, Alberta, Canada T6G 2J1

Abstract. Data from arrays of magnetometers along lines of constant magnetic latitude and longitude supplemented by all-sky camera and riometer data are used to infer the characteristics of the temporal development and the typical scale size of westward travelling surges which occur during magnetospheric substorms. It is found that the motion of the head of the surge can be quite irregular, and that in extreme cases the surge form may grow and decay in a confined longitudinal sector without suffering any significant westward displacement. The positive *D*-component perturbation, known to be the characteristic signature of a surge, is generally confined within a longitude range of $\sim 6\text{--}10^\circ$ at $\sim 70^\circ$ N and is thought to be generated by a filamentary southward ionospheric current flowing at the head of the surge. A comprehensive model three-dimensional current system involving this equatorward current and northwestward current flow in the region to the east of the head of the surge is presented through a detailed comparison of model and observed latitude and longitude profiles of the magnetic disturbance. It is found that best agreement is obtained when the entire electrojet system flows from southeast to northwest relative to the lines of constant magnetic latitude.

Key words: Auroral zone magnetic fields – Westward travelling surge – Three-dimensional current system – Auroral electrojet – Substorm

Introduction

The magnetospheric substorm is a large scale episode of enhanced energy dissipation in the ionosphere reflected by marked increases in auroral luminosity, acceleration of electron and ions and significant joule heating due to increased current flow in the auroral zone electrojets. The concept of the substorm was developed by Akasofu (1964) for the auroral signatures, and has subsequently been extended to encompass associated variations in particles, current and electromagnetic noise in various portions of the frequency spectrum (Akasofu 1968). The substorm has only just recently been given an operational definition (Rostoker et al. 1980) in which multiple surges and associated current intensifications are permitted to occur inside the time frame of a single substorm. In this paper we will be addressing the problem of the development of the auroral and electric current features during individual intensifications within the body of a substorm.

The westward travelling surge is a well recognized signature of the magnetospheric substorm which is thought to represent

the western edge of the substorm disturbed region. The surge moves erratically in a northwest direction during the course of a substorm (Wiens and Rostoker 1975) although individual surge features may expand smoothly westward and surge forms will sometime develop to the east of the head of the westward electrojet (Pytte et al. 1976). There have been several detailed studies of the auroral and magnetic field signatures of surges (Akasofu et al. 1966; Meng 1965; Kisabeth and Rostoker 1973; Chen and Rostoker 1974; Rostoker and Hughes 1979; Baumjohann 1979) which have provided an important base of knowledge regarding the phenomenology of this substorm feature. Kisabeth and Rostoker (1973) have shown that the positive perturbation of the *D* (east-west) component of the magnetic field is a primary feature of the disturbance region and Rostoker and Hughes (1979) have reached the conclusion that this feature is due to an equatorward ionospheric current flowing at the head of the surge. However, lack of adequate ground based magnetometer coverage has, till now, prevented a detailed study of the westward propagation of the surge from being carried out. Over the past 5 years the University of Alberta has intermittently operated an east-west line of between three and four magnetometers along a line of constant geomagnetic latitude ($\sim 67.3^\circ$ N) over an east-west extent of $\sim 12^\circ$ of longitude. For some of this time an all-sky camera was operated at Fort Smith (SMIT) so that concurrent information on auroral luminosity was available for some events. For other events, riometer data were available which allowed the region of energetic electron precipitation to be studied. In this paper we use the above-mentioned data base to study the development and propagation of westward travelling surges, and we further present model three-dimensional current systems which we believe to be capable of producing the surge magnetic field perturbations in terms of both spatial motion and temporal intensity variations.

Presentation of the Data

In our study we have explored the development of twelve substorm events in which surge forms were detected in the Alberta sector. Tighe (1979) has analysed three of these events and in our paper we shall present detailed analyses of two of these events. For the first event on Day 214, 1974 the University of Alberta magnetometer array involved the stations of Uranium City, Fort Smith and Hay River spaced at intervals of $\sim 5^\circ$ along a line of constant geomagnetic latitude at $\sim 67.3^\circ$ N. Supplementary data available from standard observatories operated by Dept. of Energy, Mines and Resources (Earth Physics Branch), were used in the study, along with all-sky camera data

Table 1. Locations of stations used in this study

Station name	Code name	Geographic coordinates		Geomagnetic coordinates	
		Lat.	Long.	Lat.	Long.
Uranium City	URAN	59.6	251.5	67.4	304.3
Forth Smith	SMIT	60.0	248.0	67.3	299.7
Hay River	HAYR	60.8	244.2	67.3	294.1
Fort Providence	PROV	61.3	242.4	67.5	292.0
Fort Chipewyan	CHIP	58.8	248.9	66.3	302.1
Fort McMurray	MCMU	56.8	248.8	64.2	303.2
Leduc	LEDU	53.3	246.5	60.6	302.9
Cambridge Bay	CAMB	69.1	255.0	76.7	294.0
Yellowknife	YKNF	62.5	245.5	69.1	292.6
Meanook	MEAN	54.6	246.7	61.8	301.0
Fort Churchill	CHUR	58.8	265.8	68.8	322.5
Whiteshell		49.8	264.8	59.9	325.9
Great Whale River		55.3	282.2	66.8	347.2
Victoria		48.5	236.8	54.3	292.7
College		64.9	212.2	64.6	256.5
Sitka		57.1	224.7	60.0	275.4
Newport		48.3	242.9	55.1	300.0
Boulder		40.1	254.8	49.0	316.5
Honolulu		21.3	202.0	21.1	266.5

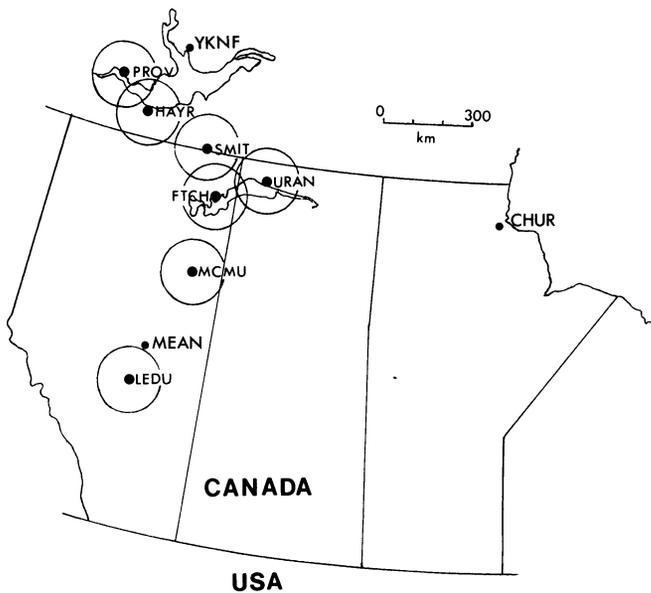


Fig. 1. Station locations of the University of Alberta magnetometers and riometers and other selected stations in the Alberta sector. The circles indicate the field of view of the riometers. The station code names and geomagnetic coordinates are given in Table 1

from Fort Smith. For the second event on Day 307, 1976 the IMS array was operative, with data being available from an east-west line along $\sim 67.3^\circ$ N stretching over $\sim 12^\circ$ from Uranium City to Fort Providence and along a meridian $\sim 300^\circ$ E from Leduc in the south to Cambridge Bay in the north. Riometer data from Fort Providence, Hay River, Fort Smith and Fort McMurray were available as well for this event. Station coordinates and code names used in this study are shown in Table 1 and the station locations are shown in Fig. 1.

The data are presented in the following formats: (i) The magnetometer data are shown in their classical magnetogram

format, stacked in each figure so as to facilitate easy comparison among the traces and to portray the evolution in time of the event. A coordinate transformation of the data results in the traces being representative of the north-south (H'), east-west (D') and vertical components Z in the centered dipole coordinate system. The magnetometer data are also portrayed in the profile format for a given instant in time. *Latitude profiles* involve perturbations from stations which lie along a line of constant geomagnetic longitude, while *longitude profiles* involve perturbations from stations which lie along lines of constant geomagnetic latitude. Both these profiles permit one to evaluate the spatial character of the substorm current flow at a given instant of time. The baseline for substorm events is chosen in the minute or so prior to onset, and is changed for each subsequent intensification to the minute prior to the intensification – such profiles have been termed differential profiles (Kisabeth 1972). (ii) The riometer data are shown as stacked time series, which indicate the level of the signal in terms of dB. (iii) The all-sky camera data are shown as linear mappings of the auroral images recorded at 30 s intervals using alternate 8 and 16 s exposure times.

The Substorm Surge Events

(i) Day 214 (2 August, 1974)

The event in question appears to have its onset at ~ 0626 (all times in this paper are given in universal time) after a sustained period of quietness. The stacked magnetograms for the H' , D' , and Z components are shown in Fig. 2a–c respectively. There is some indication of a slow buildup of electrojet strength starting around 0620, however the distinctive arc brightening identifying the substorm onset (Akasofu 1964) does not occur until 0626. The auroral mappings and matching longitude profiles of the magnetic data are shown in Fig. 3a, b. The arrows a, b, c, ... in Fig. 2 pertain to the auroral mappings A, B, C, ... in Fig. 3.

The substorm onset features a negative H' perturbation at the eastern stations of URAN and CHUR. CHUR shows a marked $+D'$ perturbation indicating it is near the head of the surge, while the Alberta array registers a small $-D'$ disturbance indicating it is not in the region of the surge. It is interesting to note the appearance of the surge form ~ 200 km to the east of URAN does not produce the familiar $+D'$ signature at that station suggesting that the $+D'$ surge signature may be quite localized spatially. In fact all three stations in the Alberta array show rather small magnetic perturbations in this phase of the substorm.

The appearance of the surge at the eastern edge of the Alberta array at ~ 0628 is accompanied by a sharp $+D'$ perturbation at URAN. This signature is slightly delayed at SMIT (implying some propagation delay) and is completely absent at HAYR (indicating that the effects of the surge are not felt that far west). The all-sky data in Fig. 3 indicate that the main surge form has moved to about 100 km east of URAN and that the arc structure has developed a perturbed character between URAN and SMIT. The buildup of $-H'$ at URAN as indicated by the longitude profile in Fig. 3a indicates that the western edge of the westward electrojet has moved closer to URAN in the two minutes since onset. By ~ 0630 the western edge of the surge has moved to between URAN and SMIT. Here we really can see the localized nature of the $+D'$ disturbance for profile C. Initially URAN fails to respond, and then drops negatively while SMIT registers the $+D'$ spike and HAYR still fails to respond. The longitude profile at 0629:57 is typical for that associated with a Birkeland current loop involving down-

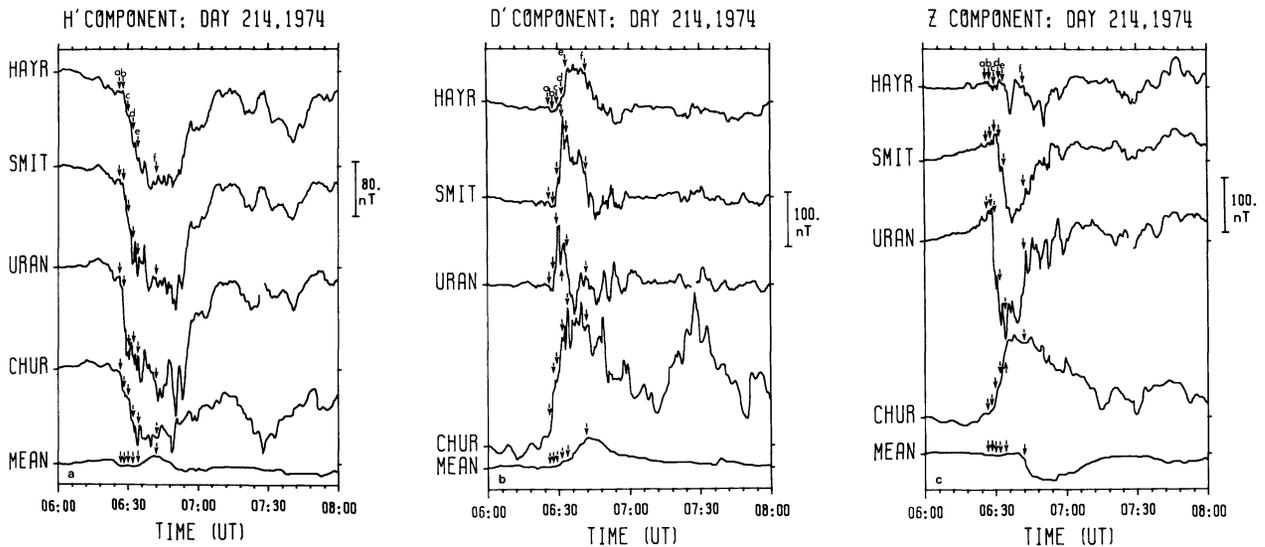


Fig. 2a-c. Stackplots of magnetograms from central and western Canadian stations covering the substorm event of day 214, 1974. The arrows labelled *a*, *b*, ... indicate specific instants during the substorm which are referred to in the text and in Fig. 3. *a* for the *H'*-component, *b* for the *D'*-component, *c* for the *Z*-component

ward field-aligned current at the north linked to upward field-aligned current to the south by equatorward ionospheric current (see Fig. 3 of Kawasaki and Rostoker (1979a)). The negative *H'* in the profile indicates further westward expansion (and possibly intensification) of the westward electrojet. By ~ 0632 the surge is over SMITH with its western edge reaching to within ~ 50 km from HAYR. At this time the *D'* component goes sharply positive at HAYR, SMIT and URAN showing all three stations to be under the influence of the surge. By ~ 0634 the active arcs have moved to the west of the Alberta array leaving the stations under the influence of a westward electrojet whose intensity does not seem to vary much with longitude (as seen from profiles *D* and *E* in Fig. 3).

This event features a surge which developed to the east of the Alberta array and which moved between ~ 0626 – 0631 across the array with a relatively smooth westward motion. Between ~ 0631 – 0633 the surge form became stationary, but at 0633 it jumped suddenly westward. This motion is inferred by studying the position of the leading edge of the surge as a function of time as shown in Fig. 4. Since we associate the $+D'$ magnetic disturbance with the surge form, it was useful to study the motion of a well defined feature in the *D'*-component longitude profile, this being the position of the polarity reversal in *D'* which can be seen in frame B of Fig. 3. The position of the polarity reversal was studied at 10 second intervals and is shown in Fig. 4 for comparison with the position of the head of the surge as inferred from the auroral luminosity. Clearly, over the period during which the surge was over the Alberta array, the slope of the two curves is approximately the same with regression analysis yielding a velocity of 1.9 ± 0.2 km/s for the auroral feature and 2.3 ± 0.2 km/s for the *D'*-component crossover. Given the errors of up to 5% in the scaling factors used in transposing the all-sky camera photos to linear mappings, this agreement can be considered to be rather good.

In summary, we have shown an example where the surge moved from the eastern field of view over the Alberta array. The motion is smooth for ~ 5 minutes between 0626 and 0631, but was absent for two minutes between 0631 and 0633 after which time rapid westward motion was again initiated. Inspection of the longitude profiles indicate a regime of negative *D'* to

the west of the surge and a region of positive *D'* under the surge. There is some indication of negative *D'* behind the surge (see SMIT and URAN in Fig. 2b) but this is not well pronounced for this event. The *Z*-component is positive to the west of the surge and negative to the east of the surge. The *H'* component is negative under the surge and to the east of it, indicating the effect of the substorm westward electrojet. The *Z* and *D'* profiles are what one would expect from a three-dimensional current system involving anti-parallel Birkeland current sheets connected by ionospheric equatorward current flow.

(ii) Day 307 (2 November, 1976)

The substorm analysed in this section features the development and decay of several surge features which appear in different positions with respect to the Alberta array. Magnetograms from the east-west line are shown in Fig. 5 with records from the north-south line within the Alberta array being shown in Fig. 6 and riometer data from the Alberta array being shown in Fig. 7.

The substorm onset occurred at ~ 0625 as evidenced from the negative *H'* component perturbations visible across the Alberta sector. A latitude profile taken shortly after the substorm onset is shown as panel A of Fig. 8. The substorm westward electrojet is seen to lie in the latitude range $\sim 65^\circ$ – 69° . The lack of a positive *D'*-component perturbation indicates that the surge has not penetrated the Alberta sector at this time. This would suggest that the magnetic perturbations stem from enhanced current flow in the discrete auroral arc region immediately to the west of the substorm disturbed region (Kawasaki and Rostoker 1979b). The longitude profile for the same instant of time (panel A of Fig. 9) is consistent with this view in that it shows a negative *H'*-component perturbation which changes little across the east-west extent of the Alberta array. In addition, the east-west gradient in the *Z*-component and the steady negative *D'*-component perturbation indicate a westward jet which is tilted $15^\circ \pm 5^\circ$ to the north-west over the interval 0625–0640 (period A on the magnetograms).

At ~ 0640 the structure of the current flow in the Alberta sector is altered significantly by the growth of a surge structure. Initially the structure is highly localized near SMIT as one can

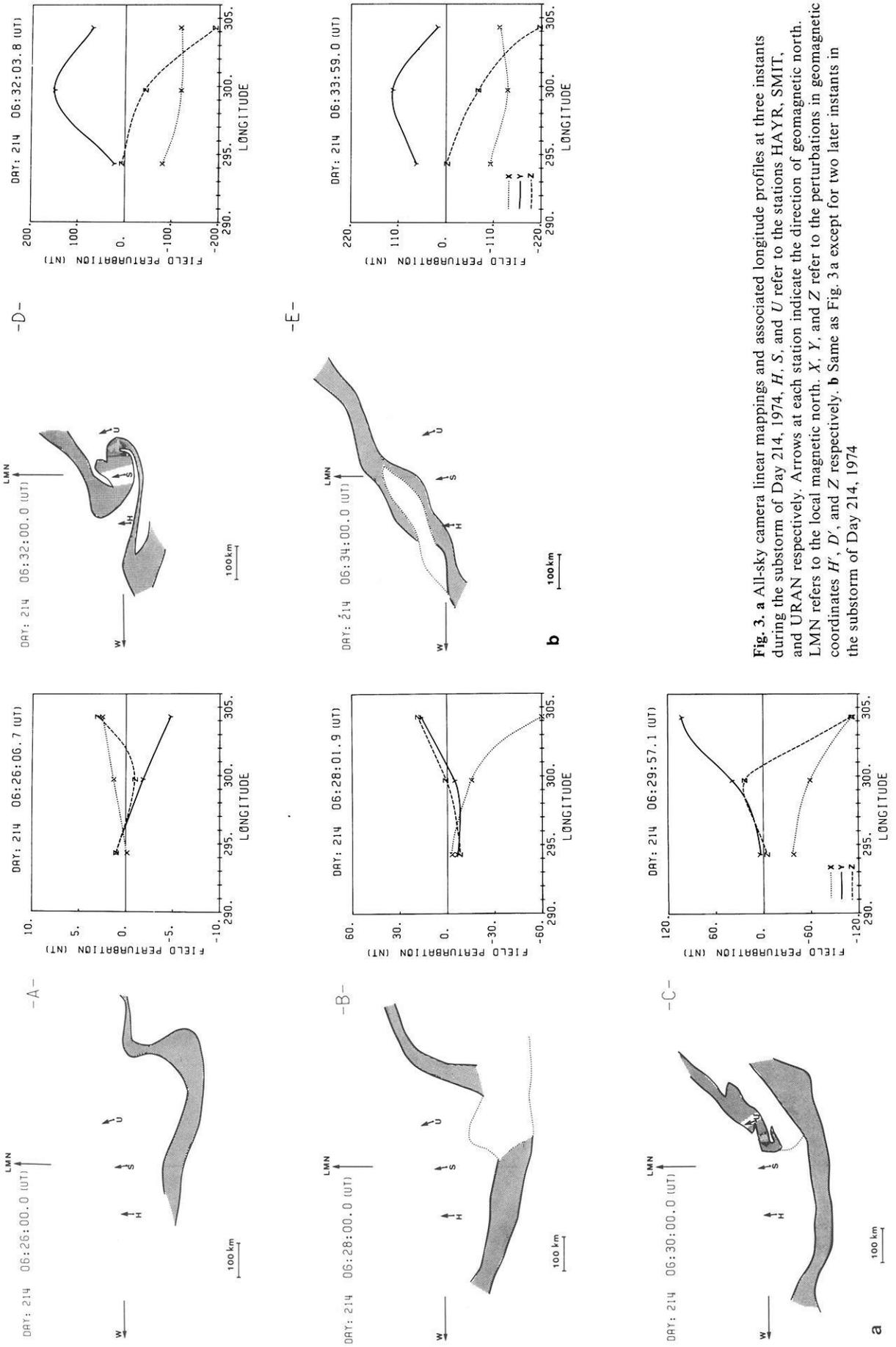


Fig. 3. a All-sky camera linear mappings and associated longitude profiles at three instants during the substorm of Day 214, 1974, *H*, *S*, and *U* refer to the stations HAYR, SMIT, and URAN respectively. Arrows at each station indicate the direction of geomagnetic north. LMN refers to the local magnetic north. *X*, *Y*, and *Z* refer to the perturbations in geomagnetic coordinates *H'*, *D'*, and *Z'* respectively. **b** Same as Fig. 3 a except for two later instants in the substorm of Day 214, 1974

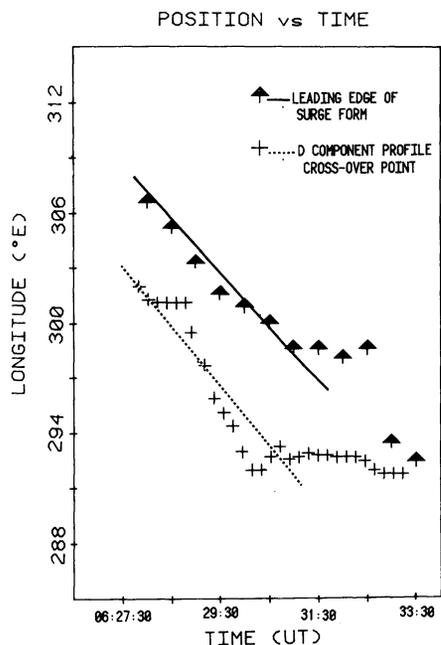


Fig. 4. Graph of the average position of the leading edge of the surge and the position of the polarity reversal in the D' -component (from negative in the west to positive in the east) as a function of universal time. Note the abrupt cessation of westward motion at ~ 0630

see comparing the D' -component perturbations to the east, west and south of SMIT to that at SMIT itself. In addition there is a significant precipitation event at SMIT which is not present either at PROV or HAYR to the west (see Fig. 7). The latitude and longitude profiles at two instants during the development of the surge are shown in panels B1 and B2 of Figs. 8 and 9. The tendency for the peak disturbance to remain localized at SMIT indicates that this surge form is not moving westward significantly. In fact, the perturbation pattern at URAN suggests that the surge did not move into the Alberta sector from the east. We conclude that this surge form has developed in a localized azimuthal sector centered near SMIT. While the disturbance

has the magnetic signature of the westward travelling surge, this surge is not travelling but decays away while remaining relatively stationary.

From 0645–0650 (period C on the magnetograms) another surge form develops. Panel C of Figs. 8 and 9 show the latitude and longitude profiles at one instant during the development of the surge. It can be seen that the peak D' -component disturbance is now near HAYR, about 5° to the west of SMIT. Minute by minute profiles (not shown here) indicate, as in the case of period C, that the surge grows and decays while showing no discernable westward motion. Based on the data from the east-west line and the enhanced D' -component perturbations at YKNF (relative to the earlier periods A and B) we would say that this surge formed to the north and west of the previous one (event B).

The major intensification in this substorm takes place near 0656, with the D' -component data (Panel D1 of Fig. 10) showing the surge developing in the vicinity of SMIT. The latitude profiles (panels D1 and D2 of Fig. 11) indicate that the surge expands rapidly poleward in conjunction with the development of a strong westward electrojet at the poleward edge of the array. The longitude profiles in Figs. 10 and 12 document the development of the surge and suggest clear westward propagation. While the peak of the surge form is near SMIT at ~ 0656 (D1), it lies between HAYR and PROV by ~ 0659 (D2) and well to the west of PROV by ~ 0703 (D3). Based on the motion of the point at which the polarity of the D' -component reverses (e.g., the crossover point is $\sim 300^\circ$ – 301° E in panel D3 of Fig. 10), the velocity of the surge is estimated to be ~ 1.2 km/s over the interval 0656–0703.

Following the passage of this major surge, the substorm continues to develop to the west of the Alberta array. Latitude and longitude profiles taken at 0718:35 (panel E in Figs. 11 and 12) show an intense electrojet peaking to the north of YKNF with a maximum H' -component perturbation of at least ~ 430 nT. The longitude profile shows an electrojet which features little longitudinal gradient in current intensity. The tilt in the Z -component profile taken together with the steady D -component perturbation can be used to estimate the angle which the electrojet makes with respect to lines of constant geomagnetic latitude. The estimated value of $20^\circ \pm 5^\circ$ is close to the estimate

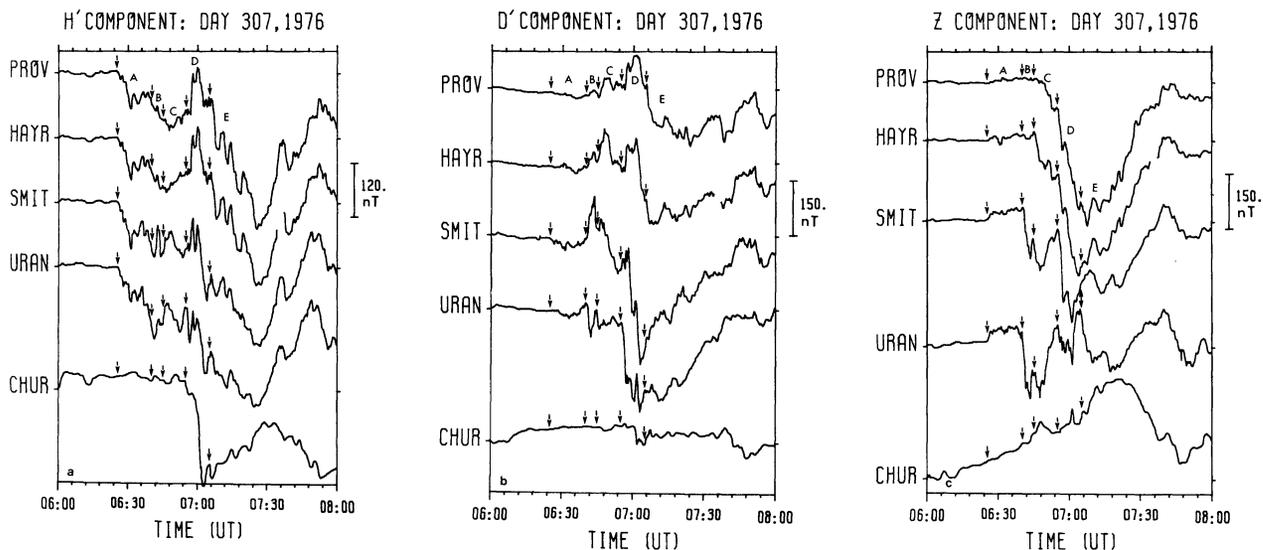


Fig. 5a–c. Stackplots of magnetograms along an east-west line stretching from PROV to CHUR for the substorm event of Day 307, 1976. The intervals A, B, ... refer to specific intervals referred to in the text. a for the H' -component, b for the D' -component, c for the Z -component

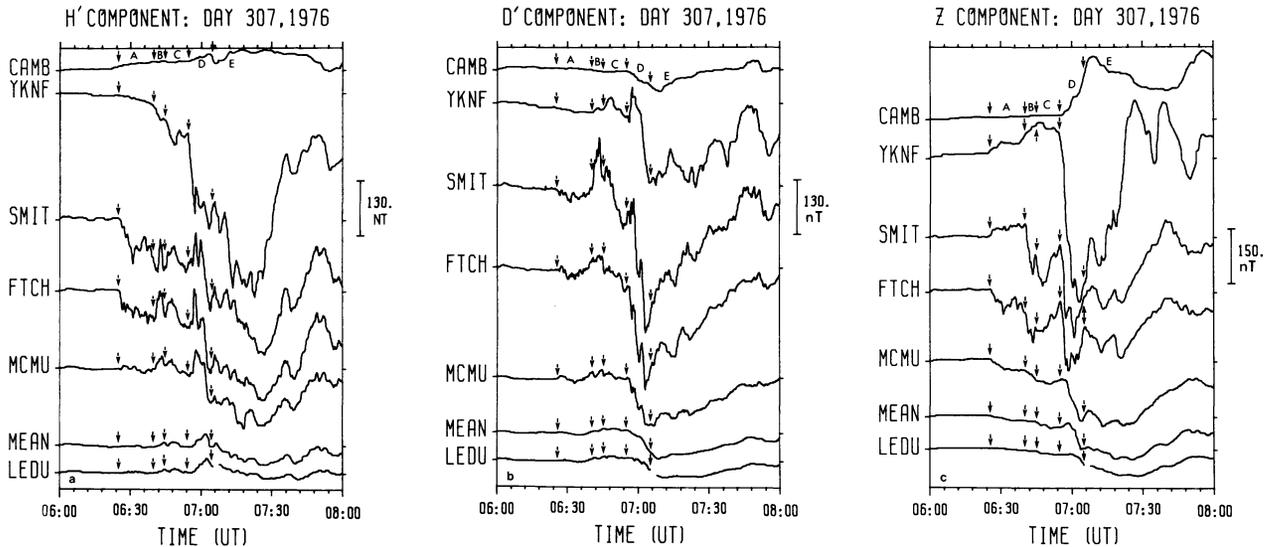


Fig. 6a-c. Stackplots of magnetograms along a north-south line through the Alberta sector for the substorm event of Day 307, 1976. a for the H' -component, b for the D' -component, c for the Z -component

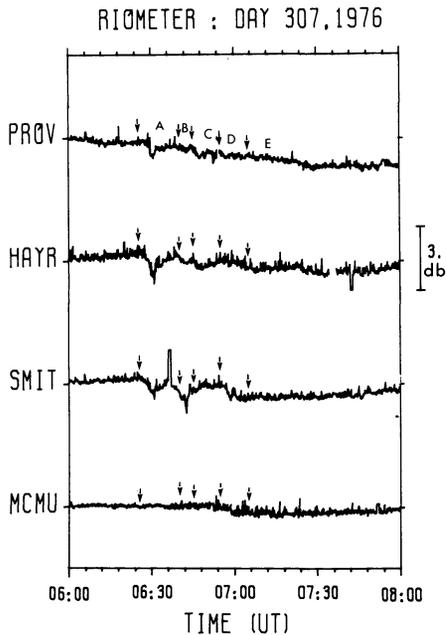


Fig. 7. Riometer data from four observatories in the Alberta sector recorded during the substorm event of Day 307, 1976. The intervals A, B, ... are referred to in Figs. 5 and 6 and in the text

of the angle which the electrojet made with lines of constant geomagnetic latitude ($\sim 15 \pm 5^\circ$) at the very beginning of the substorm (period A). The passage of the substorm has not apparently influenced the gross configuration of the auroral oval over this time scale of ~ 1 h.

In summary, we have seen in this substorm an example of a surge which propagates westward and examples of surges which grow and decay while remaining relatively stationary. In addition we have seen a surge form develop to the east of a previous surge in the body of the same substorm, in agreement with earlier observations by Pytte et al. (1976).

Modelling of Substorm Surges

The data presented in this paper include the first longitude profiles of westward travelling surges. As such, it is worthwhile itemizing some of the characteristic features of the surge perturbation pattern with emphasis on changes which occur as a function of longitude. The reader is referred to Figs. 3, 9, 10 and 12 for examples of the longitude profiles on which our synthesis is based:

(a) The dominant feature is a regime of positive D' which earlier studies (Kisabeth and Rostoker 1973; Rostoker and Hughes 1979) have found to be in the near vicinity of the western edge (or head) of the surge. At the positive peak of D' , the Z -component is normally negative, while the H' -component is either positive or negative depending on the position of the observing station with respect to the center of the surge.

(b) As one moves westward from the position of peak positive D' , the D' -component falls off in magnitude eventually becoming slightly negative. As this happens Z becomes positive reaching a positive peak before falling off to zero. The H' -component tends to fall off toward zero as one moves westward from the position of peak positive D' .

(c) As one moves eastward from the position of peak positive D' , the D' -component falls off toward zero eventually becoming negative. Normally the negative D' behind the surge is very large, often exceeding in value the magnitude of the peak positive D' , although occasionally the magnitude of the negative D' perturbation can be small. The peak value of negative Z is normally obtained at the longitude where D' switches from being positive to being negative. The H' -component reaches its maximum negative values in the region east of the polarity transition in D' .

A schematic diagram of a typical surge longitude profile is shown in Fig. 13.

A cursory comparison of the schematic diagram shown in Fig. 13 with model current system magnetic perturbation patterns suggests that most of the disturbance on the western portion of the surge is due to a three-dimensional Birkeland current system with downward current at the poleward edge linked to upward current at the equatorward edge by equatorward flowing ionospheric current, as suggested by Rostoker and Hughes

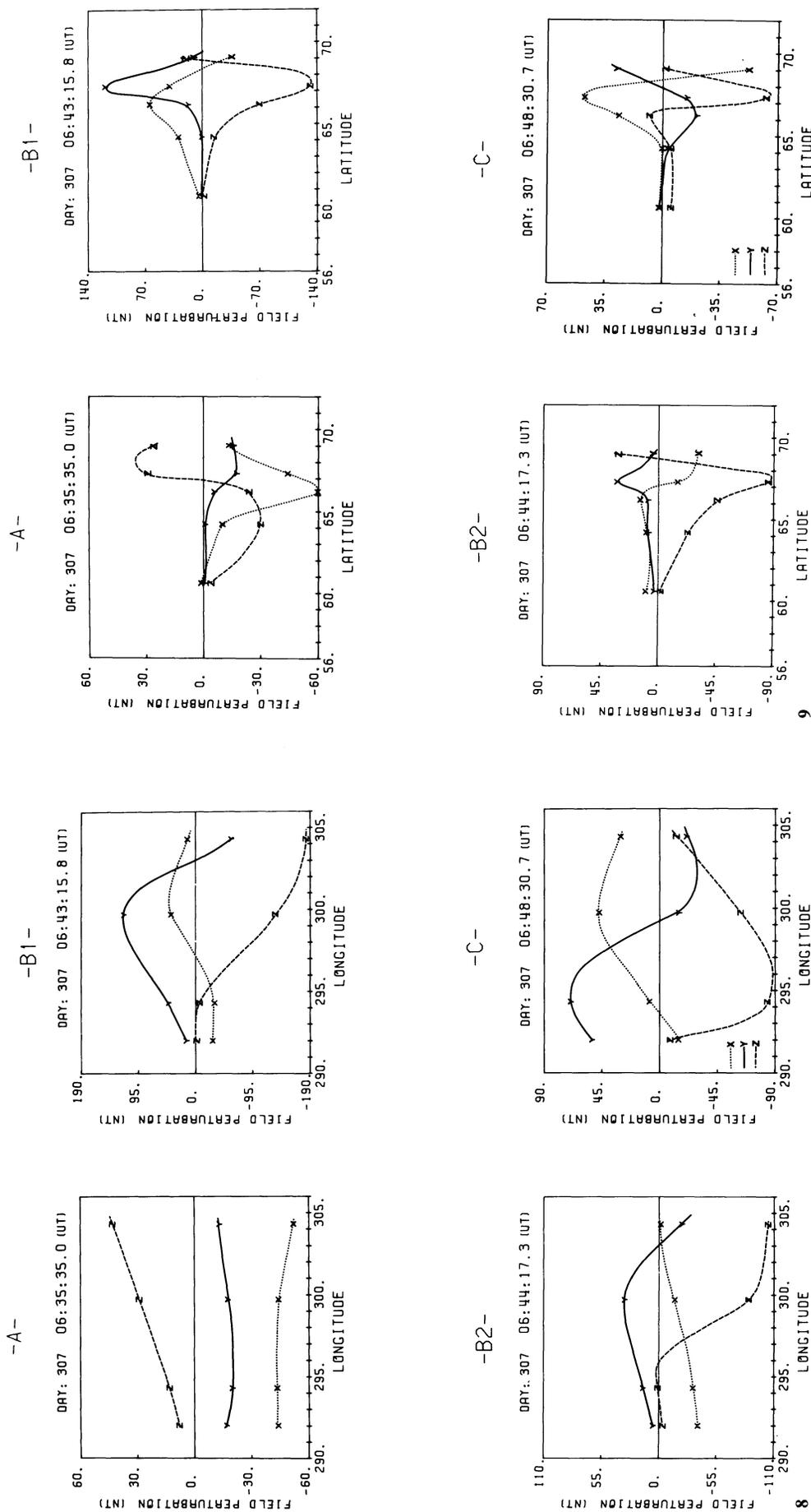


Fig. 8. Latitude profiles taken at instants during the intervals A, B, and C during the substorm of Day 307, 1976. The symbols X, Y, and Z refer to the perturbations in geomagnetic coordinates H', D' and Z respectively

Fig. 9. Same as Fig. 8 except the corresponding longitude profiles are shown

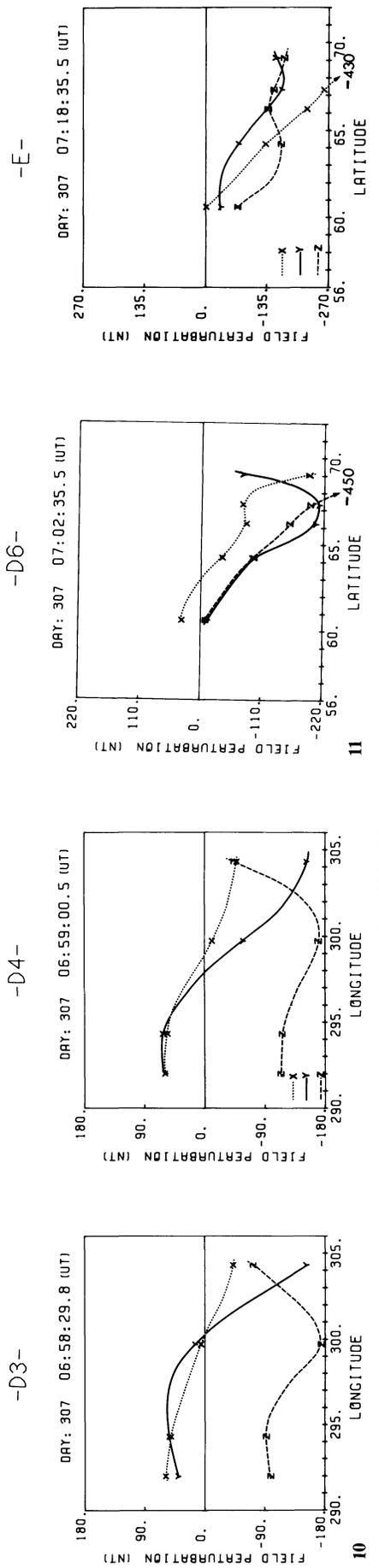
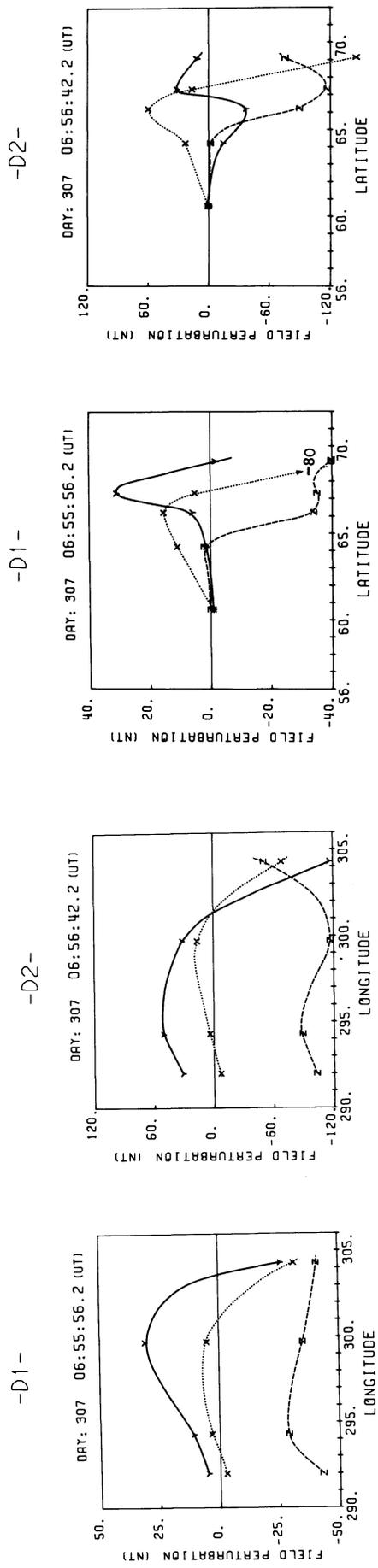


Fig. 10. Same as Fig. 9 except for instants during interval D of the substorm on Day 307, 1976

Fig. 11. Same as Fig. 8 except for instants during interval D of the substorm on Day 307, 1976

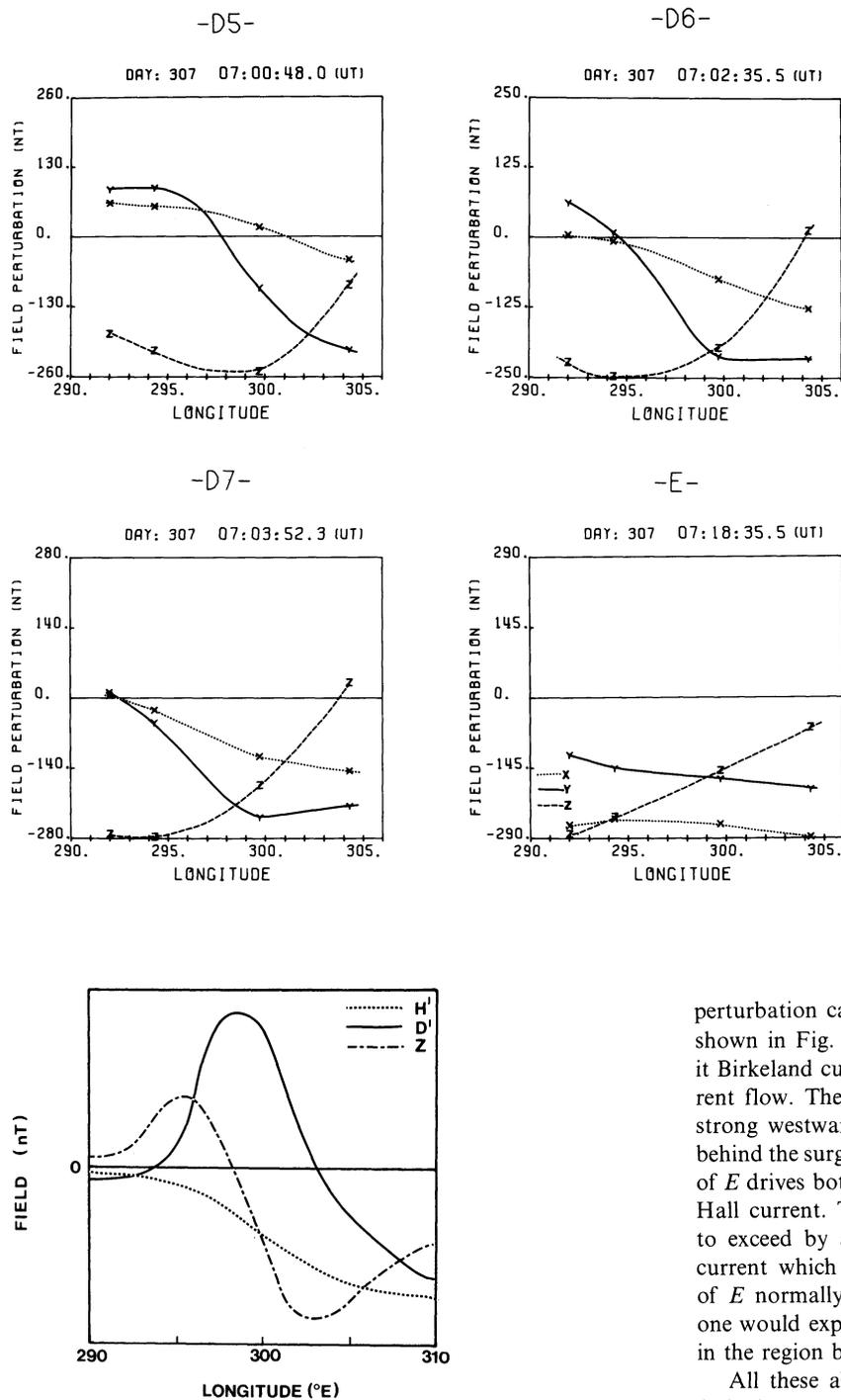


Fig. 12. Same as Fig. 9 except for instants during intervals *D* and *E* of the substorm on Day 307, 1976

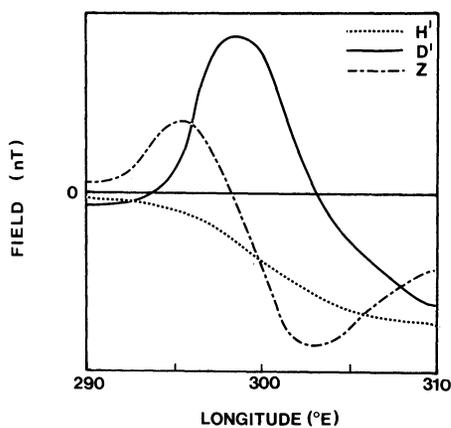


Fig. 13. Schematic diagram at a typical surge longitude profile in accordance with the observations reported in this paper

(1979). The perturbation pattern as a function of longitude for such a system is shown in Fig. 14. It is immediately apparent that the major differences between Figs. 13 and 14 are in the region to the east of the head of the surge. Those differences can, in the main, be explained by the presence of the westward electrojet for which the surge is the leading head. Two schematic profiles are shown in Fig. 15 which are typical of what might be expected for a location south of the center of a westward electrojet. Adding this type of perturbation to the surge-related

perturbation can easily produce the type of composite pattern shown in Fig. 13. Of course any electrojet has associated with it Birkeland current sheets and the connecting north-south current flow. There is now considerable evidence that there is a strong westward component of the electric field in the region behind the surge (Horwitz et al. 1978). This westward component of E drives both a westward Pedersen current and a northward Hall current. This northward Hall current would be expected to exceed by a considerable amount the southward Pedersen current which would be driven by the southward component of E normally found in the westward electrojet region. Thus one would expect to find the current flow to be northwestward in the region behind the surge.

All these above considerations were taken into account in designing a model current system to reproduce the magnetic field perturbation pattern associated with a surge. The model current system chosen is shown in Fig. 16. It consists of a westward electrojet with a longitudinal extent of L and a latitudinal width of W . This electrojet is considered to be a Hall current in the easternmost part of the system although the southwest orientation of the electric field sometimes makes it nearly a Cowling current. The westernmost part of the system is the site of intense downward fluxes of energetic electrons which represent a north-south aligned strip of negative charge. The electric field associated with this space charge distribution points eastward at the western edge of the surge and westward at the trailing edge of the surge. The upward current at the western edge of the electrojet is bled up over the surge region. Accordingly our model has a region of equatorward current flow of longitu-

dinal extent L_W at the western edge of the surge and a region of poleward current flow of longitudinal extent L_E at the trailing edge of the surge form (see Fig. 16). The north-south current is linked to the magnetosphere by Birkeland currents in the model. However it is quite possible that the northward and southward currents near the head of the surge are part of a Hall current vortex, as this would produce a magnetic perturbation pattern similar to the one produced by our three-dimensional system. Figure 17 shows sample latitude and longitude profiles computed for our model system along meridians and lines of constant latitude indicated in Fig. 16. The parameters for this model are given in Table 2. In this table, the west edge is defined as the longitude to the west of which no ionospheric current flows, that is the western edge of the region defined by L_W in Fig. 16. The total east-west length of the poleward/equatorward current system is given by $L_W + L_E$. Profile P covers the western edge of the surge and the region further west, Profile

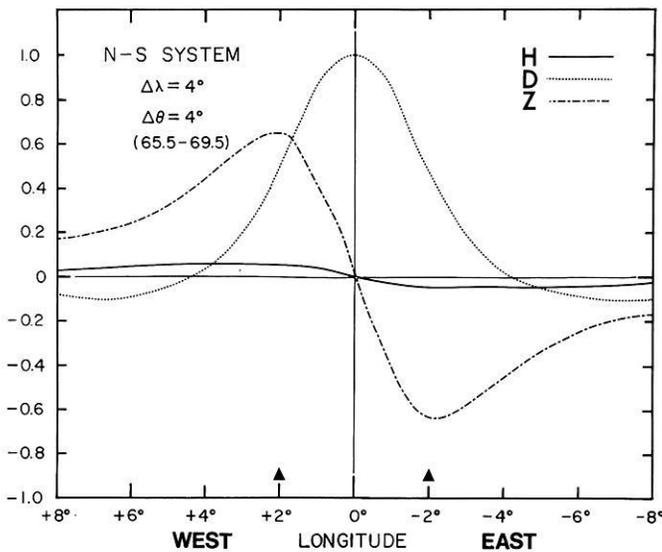


Fig. 14. Longitude profile showing the magnetic perturbation pattern for a three-dimensional current system involving a downward current sheet at 70.5° N, an upward current sheet at 66.5° N and equatorward flowing ionospheric current connecting the two field-aligned current sheets. The east-west extent of the current systems is 4° , the edges being indicated by triangles at $\pm 2^\circ$. The profile is taken along 67.5° N

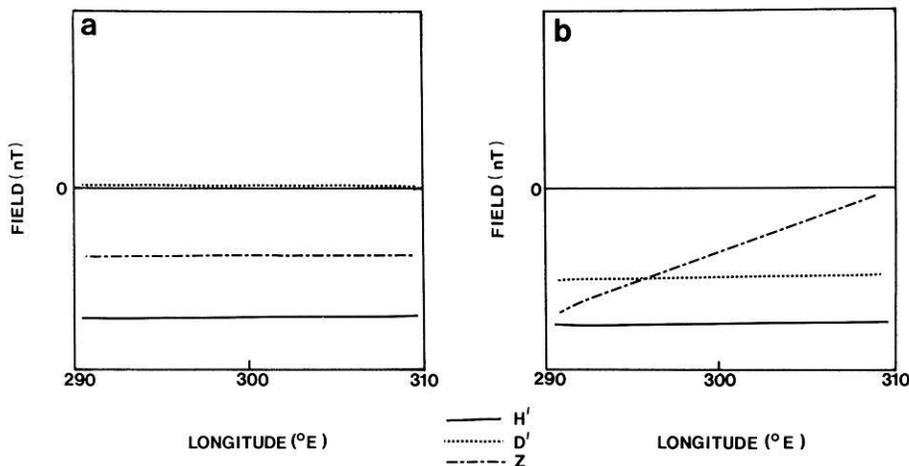


Fig. 15a, b. Schematic longitude profiles for an east-west array of field points to the south of a westward electrojet. For panel 'a' the westward electrojet lies along lines of constant geomagnetic latitude while for panel 'b' the electrojet runs from south-east to northwest at an angle to lines of constant geomagnetic latitude. Note that in the case depicted in panel 'b' the negative D' -component can account, at least in part, for the negative D' behind the surge

Q encompasses the entire surge form and Profile R emphasizes the eastern edge of the surge and regions further east of that. Each of these longitude profiles is taken at different latitudes to demonstrate how the position of the stations with respect to the electrojet can influence the observed perturbation pattern. The similarity of these profiles to the observational data we have presented in this paper is quite striking. The latitude profiles indicate how localized the positive D' regime can be, with that characteristic signature having disappeared completely in profile 3 at the trailing edge of the surge.

Finally we note that the calculations for the current systems were performed using the three-dimensional model systems de-

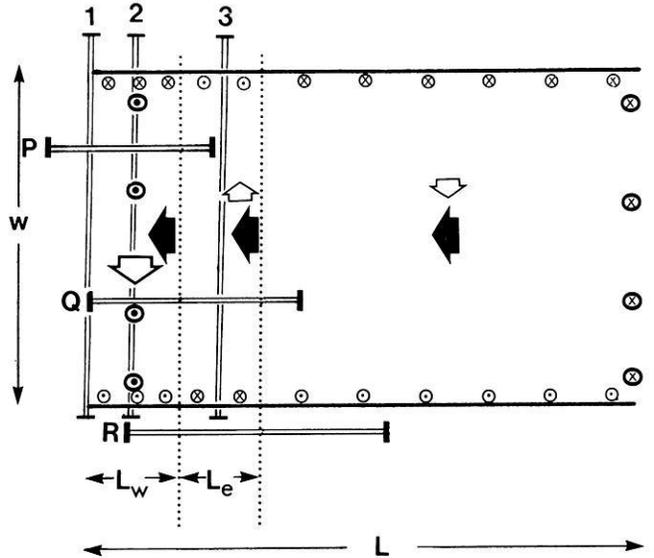


Fig. 16. Model current system proposed for the westward travelling surge and regions to the east of the surge. P , Q , and R refer to lines of constant latitude, along which theoretical longitude profiles are computed while 1, 2 and 3 refer to meridians along which theoretical latitude profiles are computed. The solid arrows indicate the direction of the east-west component of the ionospheric current, while the open arrows indicate the direction of the north-south current flow which connects the antiparallel Birkeland current sheets. Equatorward ionospheric current flows at the western edge of the surge region. The parameters shown in the figure are defined in the text. Parameters describing the actual system for cases computed in this study are shown in Tables 2 and 3

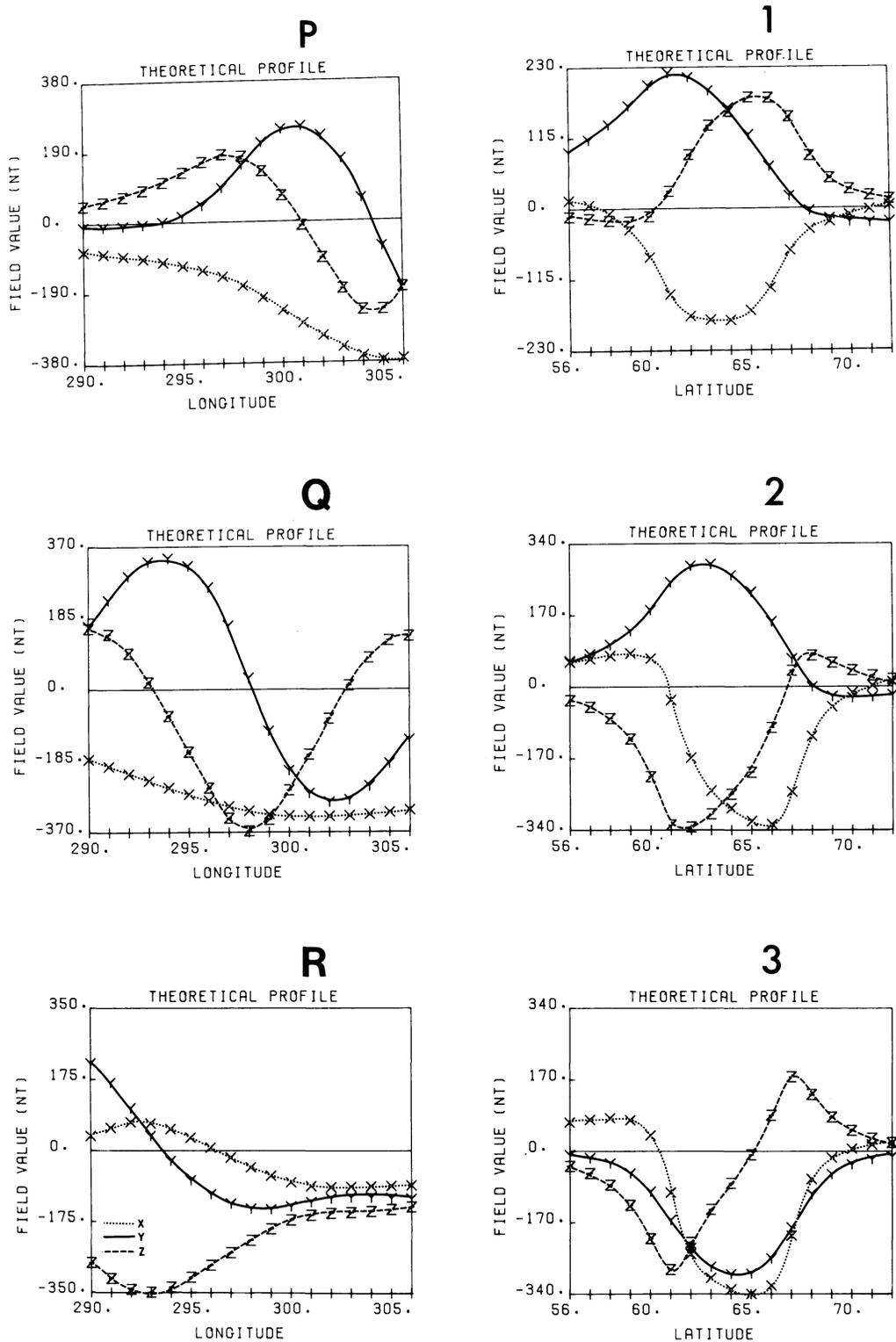


Fig. 17. Theoretical latitude and longitude sample profiles computed from the model current system shown in Fig. 16. Locations of the meridians 1, 2 and 3 and lines of constant geomagnetic latitude P, Q, and R are shown in Fig. 16. Profiles P, Q and R are designed to show the effects of surge motion. At the same time the latitudinal variation of the perturbations is illustrated by allowing profiles P, Q, and R to be taken at different latitudes across the electrojet. The current system parameters are shown in Table 2

Table 2. Model current system parameters for production of profiles shown in Fig. 17

Poleward/equatorward current system			Westward current system		
West Edge Center (°E) (°N)	E-W Length (deg.)	N-S Width (deg.)	West Edge Center (°E) (°N)	E-W Length (deg.)	N-S Width (deg.)
290.0	10.0	6.0	290.0	20.0	6.0
65.0			65.0		

veloped by Kisabeth and Rostoker (1977). For the model shown in Fig. 16 the ionospheric currents are at an altitude of 100 km and induction effects are simulated by placing a superconductor at depths below 200 km from the earth's surface. The ionospheric current density for each element of the system is 1 Am^{-1} .

(i) Model for Day 214, 1974

Figure 18 shows the observed longitude profiles chosen for model studies and the profiles obtained from the model whose parameters are shown in Table 3. One can see the increase in current density and the westward motion of the western edge of the

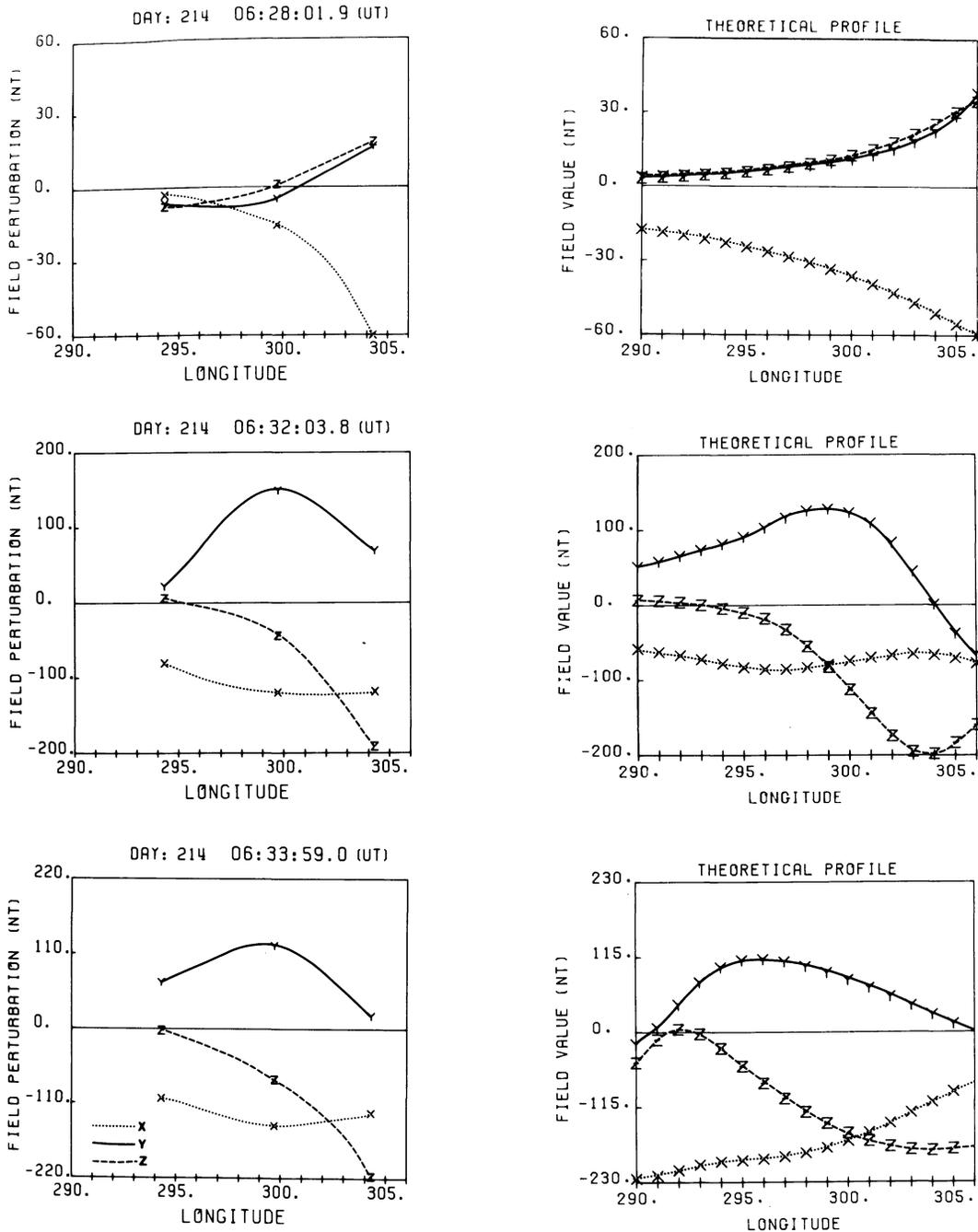
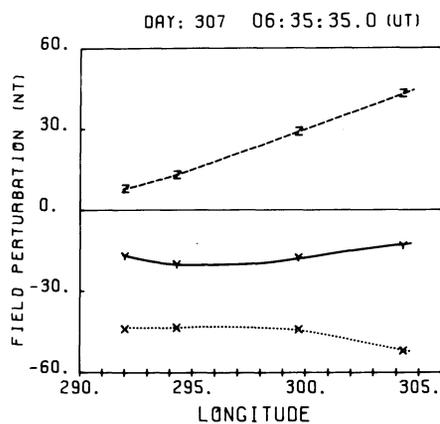


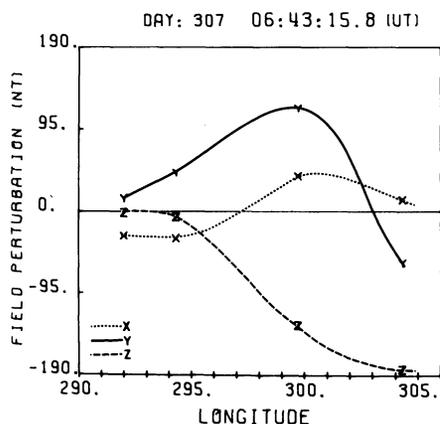
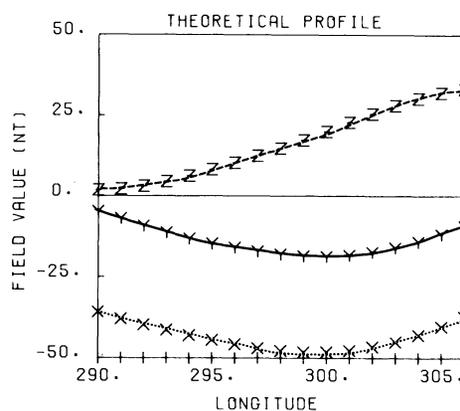
Fig. 18. Longitude profiles for observations made on Day 214, 1974 and the theoretical profiles which simulate them. X, Y, and Z indicate the geomagnetic north, east and vertical components. The current system parameters are shown in Table 3

Table 3. Model current system parameters for Day 214, 1974 and Day 307, 1976

Profile	Poleward/Equatorward current system					Westward current system				
	West Edge Center (° E) (° N)	E-W Length (deg.)	N-S Width (deg.)	Tilt (° east of south)	Current Density (Am ⁻¹)	West Edge Center (° E) (° N)	E-W Length (deg.)	N-S Width (deg.)	Tilt (° east of south)	Current Density (Am ⁻¹)
<i>Day 214/74</i>										
B	308.0 68.0	14.0	6.0	0	0.25	308.0 68.0	20.0	6.0	0	0.35
D	296.5 70.0	14.0	6.0	0	0.40	294.0 70.0	25.0	6.0	0	0.45
E	278.0 69.5	12.0	7.0	-10	0.90	278.0 69.5	30.0	7.0	-10	0.70
<i>Day 307/76</i>										
A						290.0 67.5	25.0	4.0	20	0.20
B1	292.0 71.5	19.0	6.0	20	0.55	290.0 71.5	25.0	6.0	20	0.12
D6	284.0 72.5	20.0	7.0	20	0.45	284.0 72.5	26.0	7.0	20	0.50
E						278.0 72.5	30.0	7.0	20	0.75



-A-



-B1-

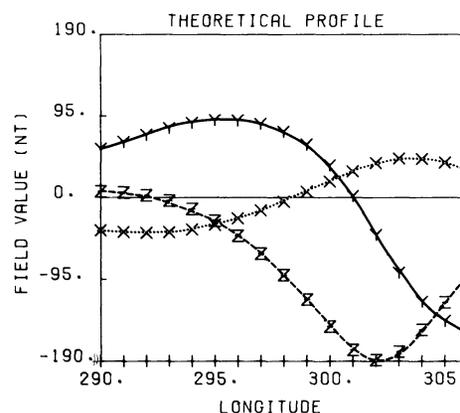
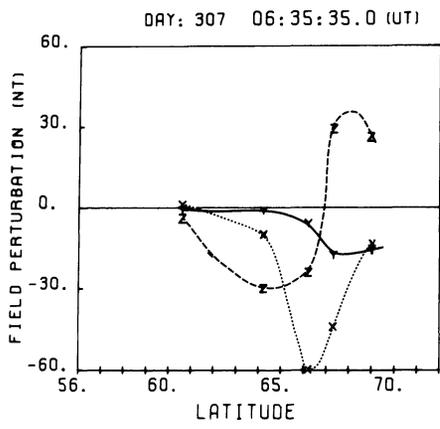
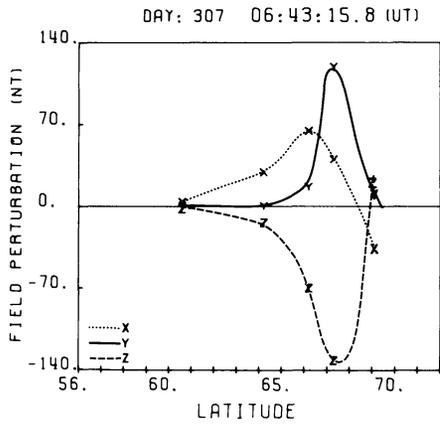
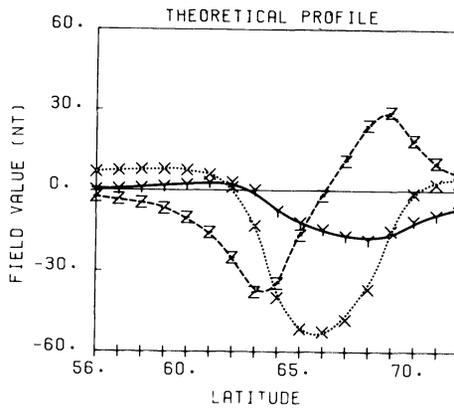


Fig. 19. Longitude profiles for observations made early in the event recorded on Day 307, 1976 and the theoretical profiles which simulate them. Note that profile *A* is taken before the arrival of the surge while profile *E* is taken after the surge form has disappeared from the Alberta sector. In both cases, only the westward current system is needed to model the magnetic effects. The current system parameters are shown in Table 3



-A-



-B1-

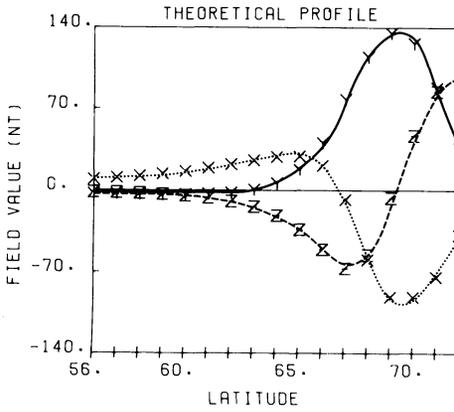
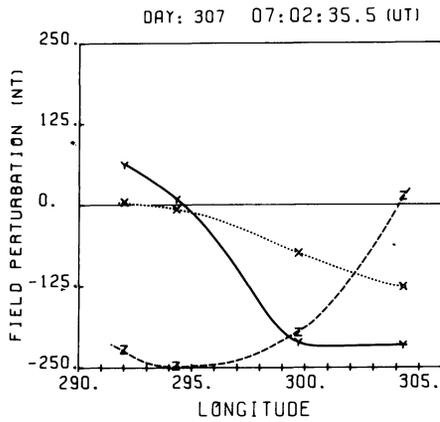
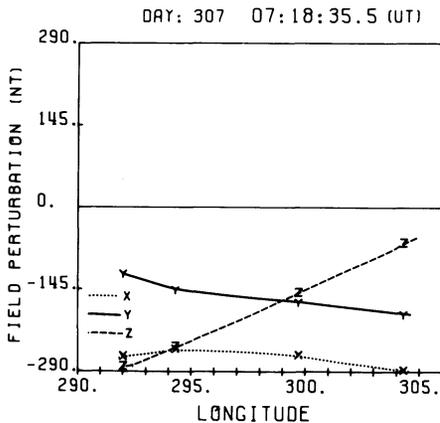
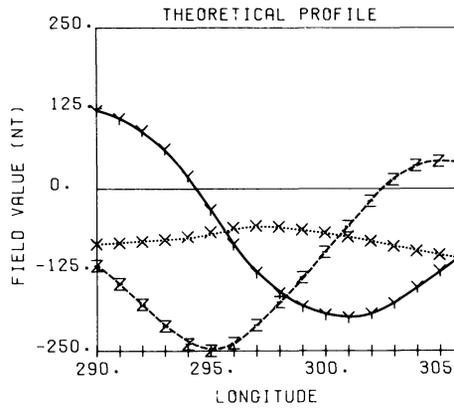


Fig. 20. Latitude profiles for the same instants during the event on Day 307, 1976 presented in Fig. 19



-D6-



-E-

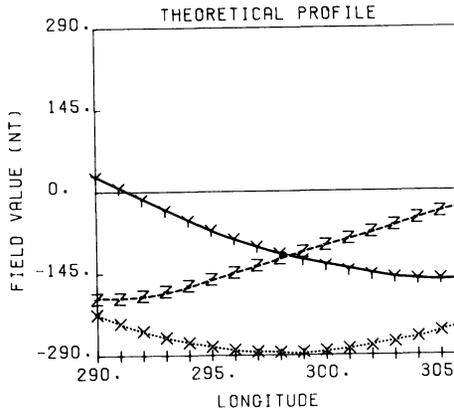


Fig. 21. Longitude profiles for observations later in the event recorded on Day 307, 1976 and the theoretical profiles which simulate them. The current system parameters are shown in Table 3

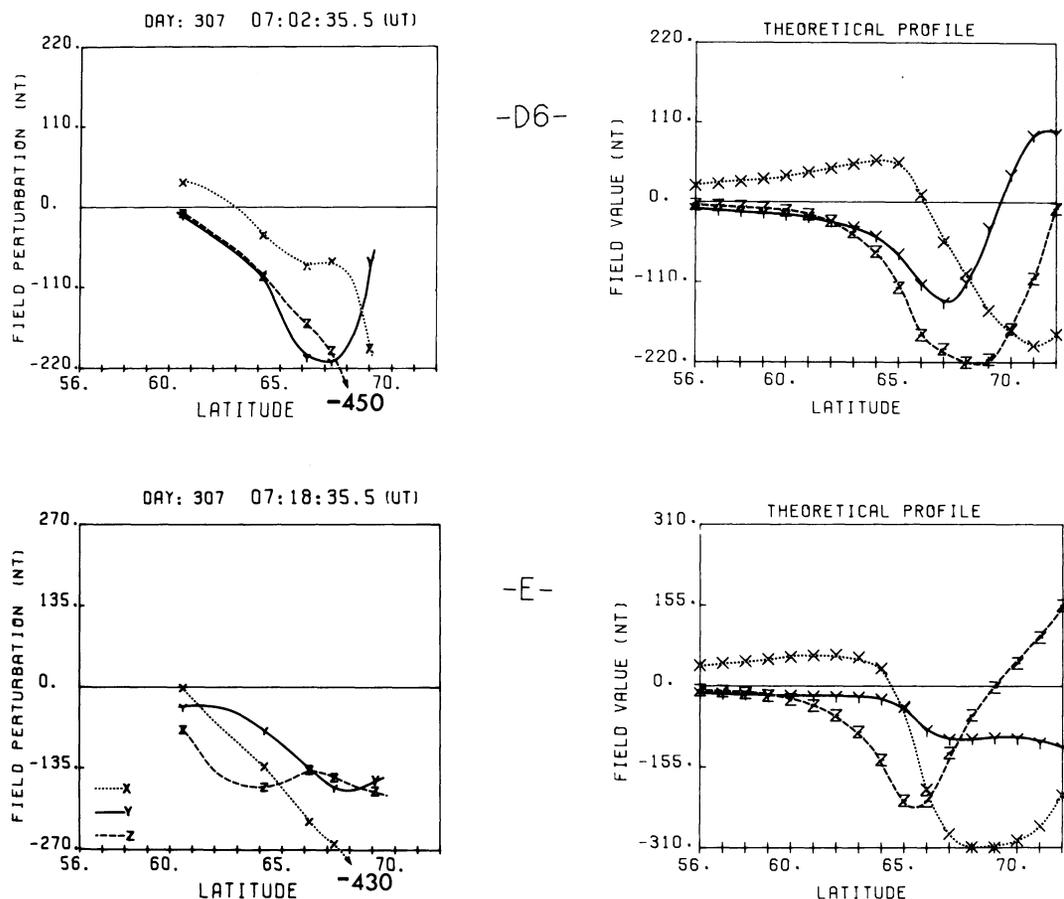


Fig. 22. Latitude profiles for the same instants during the event on Day 307, 1976 presented in Fig. 21. The maximum negative H -component at YKNF is off-scale. Good agreement of the model at YKNF is not expected since YKNF is several degrees to the west of the Alberta meridian line and marked azimuthal asymmetries occur in the vicinity of the surge

surge as the event develops. Clearly the head of the surge is to the east of Uranium City at the onset of the event but has moved west of Fort Providence by 0633:39. There are obviously some discrepancies between the model results and the observations, but this is to be expected in the light of the fact that we cannot faithfully reproduce the true current distribution in so complex a situation as a substorm disturbance.

(ii) Model for Day 307, 1976

For this case we shall use our model to produce latitude and longitude profiles, as both types of observational data were available for this event which was recorded while the Alberta IMS array was in operation. Figures 19 and 20 show sample longitude and latitude profiles. The model parameters are shown in Table 3. It will be noted that the profile at 0635:35 was taken prior to the arrival of the surge while that at 0643:15 was taken after the surge had developed in the Alberta sector. We consider the overall agreement between the model and observations to be particularly good in these cases, given the complexity of substorm disturbances.

The profiles in Fig. 21 and 22 represent the situation after the passage of the surge. The model parameters are again given in Table 3. While agreement for the longitude profiles (Fig. 21) seems reasonable, there are clear discrepancies in the case of the latitude profiles which might be expected considering the fact that all the stations are not truly on the same meridian (e.g., the longitude of Fort McMurray is 303.3° E while the

longitude of Yellowknife is 292.6° E). In addition, given the sharp spatial changes in the substorm perturbation pattern particularly in the vicinity of the surge, outstanding agreement should not be expected. In fact, this is a good indicator of how careful one must be in ensuring that stations used for meridian line profiles do not deviate significantly from a common meridian.

Discussion and Conclusions

The purpose of this research was to investigate the current configuration associated with the westward travelling surge and to trace the development of the surge form through following the spatial and temporal development of its current system. This paper is the first to utilize constraints provided by knowledge of the variation of the magnetic perturbation pattern as a function of longitude, and the longitude profiles presented are the first such profiles published in the literature to date. Using this information we can study the longitudinal motion of surge forms and gain some measure of knowledge of the longitudinal scale size of the current carrying regions. Based on the analysis presented in this paper we have reached the following conclusions.

(i) The motion of Surges

There seem to be many different ways in which a surge can develop and behave. Traditionally one considers the surge as

representing the head of a westward electrojet as it expands westward during the substorm expansion phase. Certainly, as in the case of Day 214, 1974, we have encountered cases where the surge had developed to the east of the Alberta sector and had moved westward across our sector at the head of an expanding westward electrojet. On a time scale of minutes this motion may be very irregular, and at times the surge may stop propagating for a few minutes before resuming its motion.

On the other hand, we have observed surge forms develop in which the current magnitude grows and decays without any significant westward expansion being observed (as observed early in the event on Day 307, 1976). In the cases we have studied, such surge forms tend to appear before the arrival of the substorm intensified-westward electrojet in our sector.

One may attempt to understand this surge behaviour in the following fashion. The westward electrojet connects regions of net downward field-aligned current flow in the noon sector to regions of net upward field-aligned current flow in the pre-midnight sector. The concept of the downward field-aligned current of the portion of the substorm three-dimensional current loop is implicit in the model studies of Rostoker and Hughes (1979). It is adopted because there appears to be no indication of strong net downward field-aligned current in the region just to the east of the average substorm disturbed region which would be identified by a positive level shift in the east-west component of the perturbation magnetic field going from south to north across the oval. It is, of course, possible that there is a net downward current to the east of the surge, however it would be distributed over a large longitude range resulting in a weak effect in the region of the surge. Since we are trying to model only the perturbation field in the near vicinity of the surge, we do not include any net downward current east of the surge in the model. The surge, itself, represents a region of intense upward field-aligned currents which are carried by keV electrons in the region below the acceleration region (Mozer et al. 1980). While the upward flow of cold ionospheric electrons in the noon sector can easily account for the net downward field-aligned current observed in that region, it is more difficult to explain the flow of upward current in the pre-midnight sector. It is presently felt that upward current is carried by the hot electrons which have been accelerated above the ionosphere by parallel electric fields, and that the acceleration is needed to give the electrons a sufficiently high velocity so as to provide the maximum possible current densities. The equation of continuity provides a limitation as to how large this current density can be and thus, in order to match the nightside upward field-aligned current to the dayside downward field-aligned current, it is necessary for the region of upward field-aligned current to expand in ionospheric cross sectional area. The growth of a surge region represents the development of a new section of the ionosphere from which upward field-aligned current may flow. For reasons which are presently not well understood, it would appear that regions of upward current flow (marked by discrete auroral arcs) attain limited latitudinal scale size. Development of discrete auroras involves the sudden creation of new auroral arcs rather than the continual growth in size of a single arc. The suddenness of discrete arc development is symptomatic of the substorm process and differs from the smooth appearance of increases in downward field-aligned current across the noon sector. Insofar as the actual development of a surge is concerned, initially the growth of the current density may satisfy the need for the magnetosphere to balance the upward and downward flow, so that the strength of the current in the surge may grow with no further growth of the area occupied by the surge. In that case, the

surge will simply intensify while showing no westward expansion. In other cases, the increase in downward current on the dayside may be so large that the development of a new surge region may be inadequate in terms of the ionospheric cross sectional area available and it will be necessary for the surge to expand westward. In essence, therefore, the motion of the surge region reflects the changes in downward current flow on the dayside in order to maintain current continuity in the magnetosphere/ionosphere current circuit. The most rapid spatial development of the surge would be expected to be associated with the strongest enhancements in current flow in the auroral oval.

(ii) *The Scale Size of Surges*

The analysis of the longitude profiles in the vicinity of westward travelling surges has allowed us, using modelling techniques, to estimate the east-west scale size of the surge region for several events. The data for these estimates are found in Table 3 where the length of the poleward/equatorward current system at the head of the westward electrojet represents the east-west extent of the surge region. Table 3 shows the estimates for five separate surge events of varying current strengths. It is seen that the east-west extent ranges between 12° – 20° at a latitude of approximately 70° N which amounts to a range of 455–760 km. According to our model, equatorward ionospheric current will flow in the western half of this region with poleward current flowing in the eastern half. Thus the east-west scale size of the region of equatorward current flow at the head of the surge would appear to be of the order of 230–380 km. Since the ground magnetometers are a minimum of 100 km away from the current-carrying region at all times, a region of width 230 km can be considered to be rather narrow in terms of the resolving power of ground based magnetometer arrays. For one to detect the surge region, it is necessary to have a magnetometer at the correct latitude and within approximately 500 km of the surge in an east-west direction before one can be sure of detecting the traditional positive D' -component perturbation associated with the surge.

Knowledge of the east-west scale size of surge regions allows us to make some estimate of the scale size of the substorm disturbed region in the magnetotail, which in turn reflects the probability of a satellite in the tail region finding itself in the volume of space to which the surge region maps. Rostoker and Boström (1976) developed mapping factors from the ionosphere to the magnetotail based on the contention that the Birkeland current region as observed just above the ionosphere mapped the half-width of the plasma sheet in the tail. Based on their contention, the mapping factor for the azimuthal direction was 56. Using this factor and the model calculations shown in Table 3, the azimuthal extent of the surge region in the magnetotail should be of the order of 34,000 km or $5.3 R_E$. Since the tail has an average width of about $40 R_E$, it can be seen that the surge region involves only about 10%–15% of the tail insofar as its azimuthal extent is concerned.

There is no doubt that the model for the surge presented in this paper is relatively crude. No effort was made to greatly refine the model so as to produce an optimum model based on some quantitative criterion involving the deviation of each model prediction from the observation. The relatively small number of stations used in data acquisition make such an effort unprofitable. Only the exploitation of two-dimensional array data such as that which has been recently acquired in the European sector over the period of the International Magnetospheric Study will make such improvements in analysis worthwhile to

undertake. In fact, a recent study by Inhester et al. (1981) has utilized two-dimensional magnetometer array data and auroral zone electric field data acquired by STARE to develop a model for the surge which is in good agreement with the results presented in our present study.

Acknowledgements. We would like to thank Drs. K. Kawasaki, J.V. Olson, and T.J. Hughes for useful discussions during the course of this research. We are greatly indebted to Transport Canada (Telecommunications Branch), to the Atmospheric Environment Service (Western Regional Headquarters) and to Mr. and Mrs. Ted Malewski of Fort Providence, Northwest Territories, for their help in arranging for the operation of the University of Alberta array. Standard magnetogram data were provided by World Data Center-A for the Solar Terrestrial Physics and by Earth Physics Branch of Dept. of Energy Mines and Resources of Canada. This research was supported by the Natural Sciences and Engineering Research Council of Canada.

References

- Akasofu, S.-I.: The development of the auroral substorm. *Planet. Space Sci.* **12**, 273–282, 1964
- Akasofu, S.-I., Meng, S.-I., Kimball, D.S.: Dynamics of the aurora – IV polar magnetic substorms and westward traveling surges. *J. Atmos. Terr. Phys.* **28**, 489–496, 1966
- Baumjohann, W.: Spatially inhomogeneous current configurations as seen by the Scandinavian magnetometer array. In: *Proceedings of the International Workshop on Selected Topics of Magnetospheric Physics*, Tokyo, March, 1979
- Chen, A.J., Rostoker, G.: Auroral polar currents during periods of moderate magnetospheric activity. *Planet. Space Sci.* **22**, 1101–1115, 1974
- Horwitz, J.L., Doupnik, J.R., Banks, P.H.: Chatanika radar observations of the latitudinal distributions of auroral zone electric fields, conductivities and currents. *J. Geophys. Res.* **83**, 1463–1481, 1978
- Inhester, B., Baumjohann, W., Greenwald, R.A., Nielsen, E.: Joint two-dimensional observations of ground magnetic and ionospheric electric fields associated with auroral zone currents. 3. Auroral zone currents during the passage of a westward travelling surge. *J. Geophys. Res.* **49**, 155–162, 1981
- Kawasaki, K., Rostoker, G.: Perturbation magnetic fields and current systems associated with eastward drifting auroral structures. *J. Geophys. Res.* **84**, 1464–1480, 1979a
- Kawasaki, K., Rostoker, G.: Auroral motions and magnetic variations associated with the onset of auroral substorms. *J. Geophys. Res.* **84**, 7113–7122, 1979b
- Kisabeth, J.L.: The dynamical development of the polar electrojets. Ph. D. Thesis, University of Alberta, Edmonton, Alberta, Canada, 1972
- Kisabeth, J.L., Rostoker, G.: Current flow in the auroral loops and surges inferred from ground-based magnetic observations. *J. Geophys. Res.* **79**, 5573–5584, 1973
- Kisabeth, J.L., Rostoker, G.: Modelling of three-dimensional current systems associated with magnetospheric substorms. *Geophys. J. R. Astron. Soc.* **49**, 655–683, 1977
- Meng, C.-I.: Polar magnetic and auroral substorms. M. Sc. Thesis, University of Alaska, 1965
- Mozer, F.S., Cattell, C.A., Hudson, M.K., Lysak, R.L., Temerin, M., Torbert, R.B.: Satellite measurements and theories of low altitude auroral particle acceleration. *Space Sci. Rev.* **27**, 155–213, 1980
- Pytte, T., McPherron, R.L., Kokubun, S.: The ground signatures of the expansion phase during multiple onset substorms. *Planet. Space Sci.* **24**, 1115–1132, 1976
- Rostoker, G., Boström, R.: A mechanism for driving the gross Birke-land current configuration in the auroral oval. *J. Geophys. Res.* **81**, 235–244, 1976
- Rostoker, G., Hughes, T.J.: A comprehensive model current system for high latitude magnetic activity. 2. The substorm component. *Geophys. J.* **58**, 571–581, 1979
- Rostoker, G., Akasofu, S.-I., Foster, J., Greenwald, R.A., Kamide, Y., Kawasaki, K., Lui, A.T.Y., McPherron, R.L., Russell, C.T.: Magnetospheric substorms – definition and signatures. *J. Geophys. Res.* **85**, 1663–1668, 1980
- Tighe, W.G.: The development of the westward travelling surge during magnetospheric substorms. M.Sc. Thesis, University of Alberta, Fall, 1979
- Wiens, R.G., Rostoker, G.: Characteristics of the development of the westward electrojet during the expansive phase of magnetospheric substorms. *J. Geophys. Res.* **80**, 2109–2128, 1975

Received November 3, 1980; Revised version June 22, 1981

Accepted June 23, 1981

Short Communications

Impulse in Global Geomagnetic “Secular Variation”, 1977–1979

H. Nevanlinna¹ and C. Sucksdorff²¹ University of Helsinki, Department of Geophysics, Fabianinkatu 24, SF-00100 Helsinki 10, Finland² Finnish Meteorological Institute, Division of Geomagnetism, Box-503, SF-00101 Helsinki 10, Finland**Key words:** Geomagnetic – Secular variation – Impulse

The global geomagnetic secular variation in the 1960's and 1970's was characterized by a rather rapid intensification which was shown to be of internal origin (Nevanlinna 1980). An external 11-year wave could also be identified (Nevanlinna, 1980; Yukutake and Cain 1979). In studying the global secular variation for the last few years, an even more rapid global change can be identified, namely a sudden enhancement followed by a deficit, lasting roughly two years from the end of 1977 to 1979. Here we shall demonstrate that the pulses of external origin and that it resembles an intensification of the magnetospheric ring-current system lasting for about two years.

The year to year change in the Z component in some European observatories is shown in Fig. 1. The enhancement in 1978 is clearly seen. The amplitude of ΔZ is largest at high latitudes, as can be seen from Fig. 2, which shows the corresponding amplitudes at all the observatories from which data were available at the appropriate time (Pushkov and Ivchenko 1979). The change in the H component is also shown in Fig. 2, its amplitude being largest at the equator. The first order harmonics, both external and internal, were calculated using the data shown in Fig. 2. The ΔZ and ΔH curves shown depict the calculated curves corresponding to the first order harmonics. The internal harmonic coefficient g_{1i}^0 was found to be -0.5 nT/yr and the corresponding external one $g_{1e}^0 = 15$ nT/yr, which shows, in spite of the large r.m.s. errors ($\sigma_Z = 12$ nT/yr, $\sigma_H = 3$ nT/yr, the dominant external character of the phenomenon. This, of course, can also be concluded from the negative sign of the H -variation corresponding to the positive change of Z at high northern latitudes.

By comparing monthly mean values of successive years from some Scandinavian observatories (Lovö, Nurmijärvi, and Sodankylä) we get the impression that the enhancement started as early as August 1977; it was clearly visible in November 1977, the maximum being in June 1978. From Fig. 1, where the point for 1981 is based on data for January to April only, we conclude that the phenomenon lasted throughout 1978 and 1979, but it does not seem to persist at the beginning of 1980, because the 1981–80 difference, calculated from the January to April data of these years, corresponds to the expected internal secular variation curve.

The curves in Fig. 2 depict ΔZ and ΔH in the geomagnetic dipole coordinate system supposing a homogeneous external field corresponding to a distant equatorial ring-current. The fit to ΔZ at high latitudes is not good. If the equatorial ring-current is brought to the normal ring-current distance of 4.5 Earth radii

($L=3.5$) and the ring-current is given a latitudinal extent up to 50° N and 50° S along the field lines, the fit is slightly improved.

In summary, a sudden change seems to have affected the magnetosphere in late 1977. The duration of the effect was some two years, which is longer than sudden changes are generally

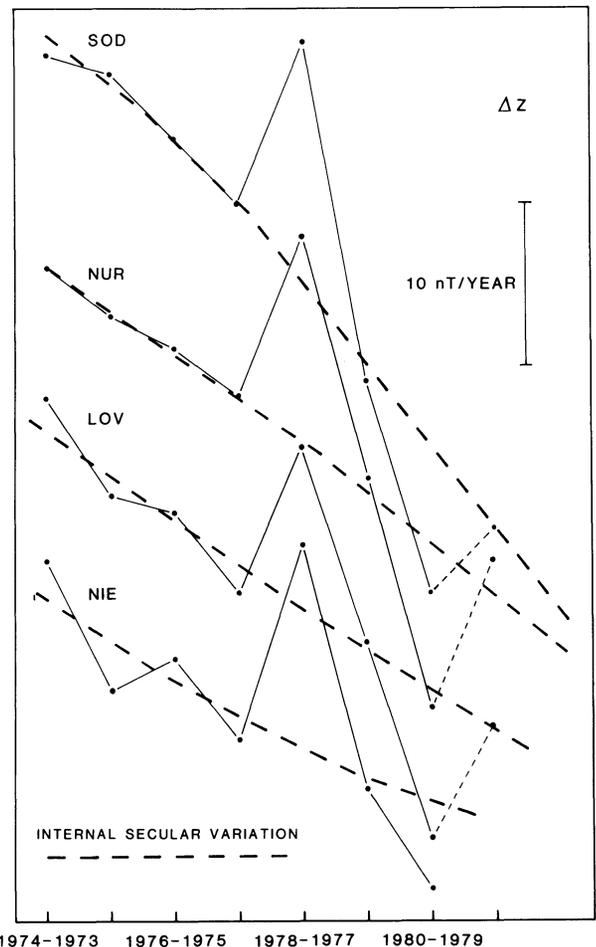


Fig. 1. Year to year change ΔZ at four Northern European observatories Sodankylä (SOD; $67^\circ 22.2'N$, $26^\circ 37.8'E$), Nurmijärvi (NUR; $60^\circ 30.5'N$, $24^\circ 39.3'E$), Lovö (LOV; $59^\circ 20.7'N$, $17^\circ 49.6'E$) and Niemegek (NIE; $52^\circ 04.3'N$, $12^\circ 40.5'E$). Broken lines: smoothed annual secular variation at the observatories. The points for 1981–1980 have been calculated from January to April data only

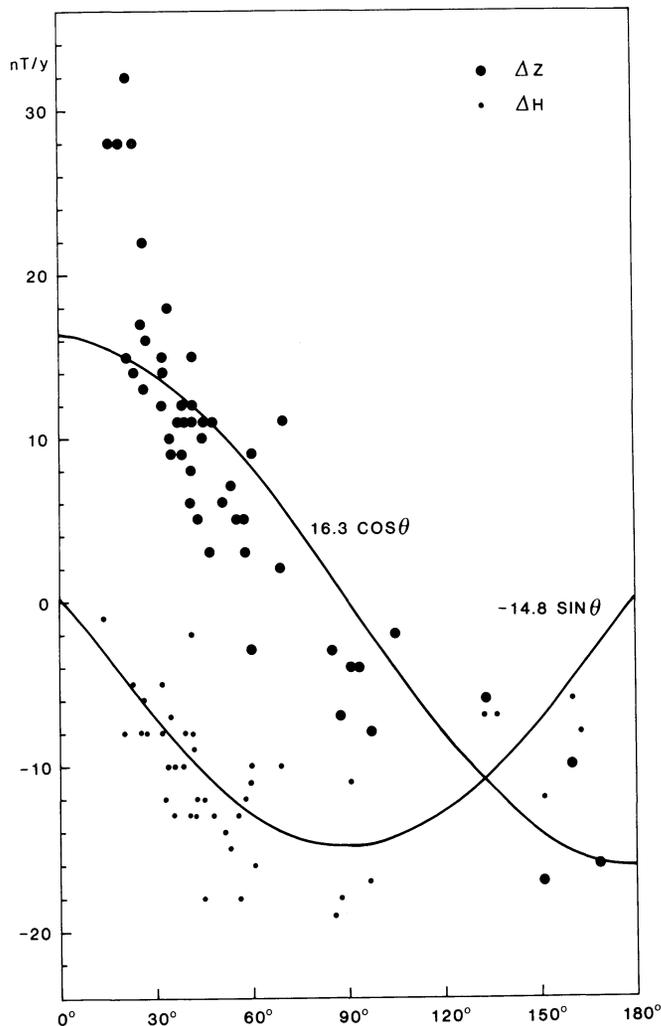


Fig. 2. *Large circles*: amplitude of enhancement in the secular variation of Z ($\Delta Z = (g_{1e}^0 - 2g_{1i}^0) \cos \theta$) at different observatories. *Small circles*: corresponding H values ($\Delta H = -(g_{1e}^0 + g_{1i}^0) \sin \theta$). θ is the dipole colatitude. The amplitude of the secular variation pulse was calculated by subtracting the smoothed extrapolated annual change 1979–1978 from the observed annual change 1978–1977. The smoothing was done by extrapolating the observed annual changes from 1974–1973, 1975–1974, 1976–1975, 1977–1976 to 1979–1978 linearly (see Fig. 1). The curves represent a least-squares fit of all points, the source being assumed to be a homogeneous axial field.

expected to last. Looking back in time, we have not been able to find a similar effect before, perhaps because the absolute measurements, especially in Z , have not previously been accurate enough. The effect seems not to be correlated with magnetic activity indices. The effect behaves like an enhancement of the ring-current with a rather large N-S dimension. The simple ring-current model does not, however, explain the very large amplitudes in ΔZ at the 50° – 70° northern latitudes, and this phenomenon needs additional explanation.

References

- Nevanlinna H (1980) Geomagnetic secular variation described by dipole models. Report Series in Geophysics No 14, Univ of Helsinki
 Pushkov AN, Ivchenko MP (1979) The summary of the annual mean values of magnetic elements at the world magnetic observatories. Vol 14, Izmiran, Acad of Sci of the USSR
 Yukutake T, Cain JC (1979) Solar cycle variations of the first degree spherical harmonic components of the geomagnetic field. *J Geomagn Geoelectr* 31: 509–544

Received April 24, 1981; Revised form June 16, 1981

Accepted June 16, 1981

Turbulent Acceleration of Electrons in Extended Extragalactic Radio Sources*

A. Ferrari¹ and E. Trussoni²

¹ Max-Planck-Institut für Extraterrestrische Physik, Garching bei München, Federal Republic of Germany

² Istituto di Cosmogeofisica del CNR, Corso Massimo d'Azeglio 46, I-10125 Torino, Italy

Key words: Stochastic acceleration of particles – MHD turbulence – Extragalactic radio sources

The physical properties and evolution of extended extragalactic radio sources (double, head-tail, jet) are supported energetically by the nuclear activity in the parent galaxies: accordingly, it is believed that the energy developed in the nuclei is transmitted to the extended radio components by supersonic (relativistic) beams or electromagnetic low frequency waves (Rees 1971; Scheuer 1974; Blandford and Rees 1974). The observed radio radiation is due to synchrotron emission, as shown by recent detailed observations, and for this process in situ acceleration of relativistic electrons is required.

Several theoretical papers have therefore been devoted to the problem of particle acceleration in physical conditions typical of radio sources (Pacholczyk and Scott 1976; Lacombe 1977; Eilek 1979; Ferrari et al. 1979). These are based on stochastic acceleration by magnetic inhomogeneities and shocks generated by magnetohydrodynamic (MHD) instabilities in supersonic flows. In particular a model may be envisaged in two steps:

(i) Resonant acceleration of electrons occurs by interaction with Alfvén waves of short wavelength λ : in this case we have a unique relation between the energy of the particle and wavelength: $\frac{2\pi}{\lambda_{\text{res}}} \cong \frac{\Omega}{\gamma c} \frac{m_p}{m_e}$, where γ is the Lorentz factor, Ω the proton gyrofrequency, c the speed of light and m_p, m_e the proton and electron mass. Alfvén waves are assumed to reach a stationary spectral distribution with index ν and total energy $(\delta B)^2$, respect to the background magnetic field B_0 , such that:

$$\left(\frac{\delta B}{B_0}\right)_A^2 \propto \int_{r_g}^{\lambda_0} \lambda^{\nu-2} d\lambda \quad (1)$$

where r_g is the proton gyroradius. As we need re-acceleration for frequencies ≥ 10 GHz, it is assumed that $\lambda_0 \geq 10^{15}$ cm and $r_g \sim \frac{\Omega}{v_A} \sim 10^{10}$ cm (v_A is the Alfvén velocity). For the parameters of a typical radio source we assume (n_e is the number density):

$$\left. \begin{aligned} B_0 &\sim 10^{-(4-5)} G \\ n_e &\sim 10^{-(4-5)} \text{ cm}^{-3} \end{aligned} \right\} \Rightarrow v_A \sim 10^{8-9} \text{ cm s}^{-1}. \quad (2)$$

(ii) The energy lost by Alfvén waves during acceleration is supplied by an MHD Kelvin-Helmholtz instability which develops at the interface between the moving radio component and intra-cluster gas.

Although appealing, the previous model is based on two crucial assumptions:

(a) If we have a typical inhomogeneity scale length L in the problem (e.g. the diameter of the jet, or, for a shear layer with a gradient of velocity, the scale length $L^{-1} \sim \frac{u'_0}{u_0}$, u_0 is the beam velocity), the fastest growing mode has wavelength $\sim L$. In the extended sources it is reasonable to assume L not much lower than the component radius i.e. \sim several Kpc. Conversely, for resonant acceleration, $\lambda_{\text{res}} \lesssim \lambda_0 \sim 10^{15-16}$ cm are required.

(b) The unstable modes with frequencies $\omega_0 \cong \frac{2\pi}{L} u_0$ for supersonic motion, are assumed to be compressive magnetic perturbations and not Alfvén waves.

In this paper we want to discuss how interaction of particles with long wavelength perturbations can lead to efficient electron acceleration. It can in fact be shown that the counteracting effect of large wavelength modes (which leads to anisotropy in the electron pitch-angle distribution) and Alfvén waves (which tend to remove such anisotropy) can lead to an equilibrium state of production of Alfvén waves and acceleration of particles. We shall discuss the phenomenon in terms of the time scales of the relevant processes, i.e. acceleration (t_{acc}) and scattering (t_{sc}) of particles, growth (t_{inst}) and absorption (t_{abs}) of MHD modes and Alfvén waves.

We start by considering the influence of a slowly varying magnetic field (associated with long wavelength MHD modes) on the particles, following a procedure similar to that applied by Melrose (1969; 1974).

An isotropic distribution function, under the effect of a time varying magnetic field $\left(\left(\frac{\delta B}{B_0}\right)_L^2\right)$ is the energy density of the long wavelength MHD perturbations)

$$B = B_0 \left[1 + \left(\frac{\delta B}{B_0}\right)_L \cos(\omega_0 t) \right] \quad (3)$$

* Paper presented at the Workshop on Acceleration of Particles by Shock Waves, 7-9 October, 1980, Max-Planck-Institut für Aeronomie, Lindau, Federal Republic of Germany.

develops anisotropy at a rate (a pitch angle of electrons)

$$\frac{d\varepsilon}{dt} \sim \omega_0 \left(\frac{\delta B}{B_0} \right)_L, \quad \varepsilon \approx \frac{f(p, a = \frac{\pi}{2}) - f(p, a = 0)}{f(p, a = \frac{\pi}{2}) + f(p, a = 0)} \quad (4)$$

with $f(p, a)$ the electron distribution function.

On the other hand for $\varepsilon \gtrsim \frac{v_A}{c}$, Alfvén waves grow and lead to acceleration and scattering of particles at a rate (Lacombe 1977):

$$t_{sc} \sim \frac{1.2 \cdot 10^{-10}}{\pi^2} \lambda_0 \left(\frac{B_0}{\delta B} \right)_L^2 F(\lambda_0, \nu) \\ t_{acc}^A \sim \left(\frac{c}{v_A} \right)^2 t_{sc}, \quad F(\lambda_0, \nu) = \frac{1}{\nu - 1} \left(\frac{\lambda_0}{\lambda_{ris}} \right)^{\nu - 2} \quad (5)$$

A quasi equilibrium situation (for $\omega_0 t_{sc} \ll 1$) can be reached when

$$\frac{\varepsilon}{t_{sc}} \sim \omega_0 \left(\frac{\delta B}{B_0} \right)_L, \quad \varepsilon \cong \frac{v_A}{c} \quad (6)$$

Taking into account that we are dealing with a resonant process and that we have a power law distribution of electrons, we can have Alfvén waves with wavelengths from $\sim r_g$ up to $\lambda_0 (\sim 10^{15} \text{ cm})$. From (5) and (6) it is possible to derive a relationship between $\left(\frac{\delta B}{B_0} \right)_L$ and $\left(\frac{\delta B}{B_0} \right)_A$

$$\left(\frac{\delta B}{B_0} \right)_L \sim 1.4 \cdot 10^6 \frac{\pi}{F(\lambda_0, \nu)} \frac{v_A}{u_0} \frac{L_{21}}{\lambda_{015}} \left(\frac{\delta B}{B_0} \right)_A^2, \\ \lambda_{015} = \frac{\lambda_0}{10^{15} \text{ cm}}, \quad L_{21} = \frac{L}{10^{21} \text{ cm}} \quad (7)$$

As we are dealing with perturbations, $\left(\frac{\delta B}{B_0} \right)_L < 1$, from (7) and the condition $\omega_0 t_{sc} < 1$ we have the following limits for $\left(\frac{\delta B}{B_0} \right)_A$

$$\left(\frac{\delta B}{B_0} \right)_A^2 > \frac{2.4 \cdot 10^{-8}}{\pi} u_{08} \frac{\lambda_{015}}{L_{21}} F(\lambda_0, \nu), \quad u_{08} = \frac{u_0}{10^8 \text{ cm/s}} \\ \left(\frac{\delta B}{B_0} \right)_A^2 < \frac{7.2 \cdot 10^{-7}}{\pi} \frac{u_{08}}{v_A} \frac{\lambda_{015}}{L_{21}} F(\lambda_0, \nu), \quad v_{A8} = \frac{v_A}{10^8 \text{ cm/s}} \quad (8)$$

These two inequalities are consistent independently of the wave spectrum if

$$v_{A8} < 30. \quad (9)$$

Scattering of particles controls the process of acceleration by long wavelength modes: following the quasi-linear theory (Kulsrud and Ferrari 1971), the rate of acceleration is fixed by the conditions $\left(k_0 = \frac{2\pi}{\lambda_0} \right)$

$$\left[\frac{k_0 c t_{sc}}{\omega_0 t_{sc}} \right] \gg \text{or} \ll 1. \quad (10)$$

Taking into account (5) and (8) together with the fact that previously we assumed $\omega_0 t_{sc} < 1$ in (10) we have $k_0 c t_{sc} > 1$ with the condition

$$u_{08} \lesssim 10 v_{A8}. \quad (11)$$

In this case the time scale of acceleration is given by

$$t_{acc}^L \sim \frac{10^8}{\pi^2} \frac{L_{21}}{(u_{08})^2} \left(\frac{B_0}{\delta B} \right)_L^2 \text{ yr}. \quad (12)$$

In the mean time Alfvén waves accelerate particles at a rate (taking into account Eq. (11))

$$t_{acc}^A \sim \frac{5 \cdot 10^7}{\pi} \frac{L_{21}}{v_{A8} u_{08}} \left(\frac{B_0}{\delta B} \right)_L \text{ yr}. \quad (13)$$

It is interesting that in our picture the acceleration times do not depend on the details of the Alfvén spectrum $F(\lambda_0, \nu)$ but only on the mechanism of the Kelvin-Helmholtz instabilities and therefore on the parameters $\left(\frac{B_0}{\delta B} \right)_L$, L_{21} , u_{08} , v_{A8} , which are strictly related to the physical conditions in the radio sources. Observational data suggest that both v_{A8} , $u_{08} \sim 1-10$ (we remember that relativistic motions lead to stabilization of modes). The values of L_{21} and $\left(\frac{B_0}{\delta B} \right)_L$ are more uncertain and can be estimated only roughly. The time scale of unstable modes is (Ferrari et al. 1978)

$$t_{inst} \sim \frac{1.7 \cdot 10^5 M L_{21}}{\pi \Phi u_{08}} \text{ yr}, \quad \Phi < 1 \quad (14)$$

where M is the Mach number and ϕ is determined from the dispersion relation. For consistency $t_{inst} \ll t_{rs}$ (t_{rs} is the age of the radio source $\sim 10^7 \text{ yr}$) so we have an upper limit for L :

$$L_{21} < 60\pi \frac{\Phi u_{08}}{M}. \quad (15)$$

Referring to $\left(\frac{\delta B}{B_0} \right)_L$ we can again assume an equilibrium situation between the growth rate of modes and their dissipation by interaction with particles.

For a power law spectrum of relativistic electrons, we know that the energy E is lost by the MHD modes during the process of acceleration at a rate

$$\frac{dE}{dt} = \int_{\gamma_{min}}^{\infty} \frac{d\gamma}{dt} dn(\gamma) \sim \\ 3.3 \cdot 10^{-16} \pi^2 \zeta \frac{u_{08}^2}{L_{21}} \left(\frac{\delta B}{B_0} \right)_L^2 \varepsilon_{rel} \text{ erg cm}^{-3} \text{ s}^{-1} \quad (16)$$

$$\zeta = \frac{\Gamma - 1}{\Gamma - 2} \gtrsim 1$$

where ε_{rel} and $n(\gamma)(\alpha \gamma^{-\Gamma})$ are the energy density and number density of relativistic particles. In a steady equilibrium state

$$\frac{dE}{dt} \sim \frac{d \left(\frac{\delta B^2}{8\pi} \right)_L}{dt} \sim 2 \frac{\left(\frac{\delta B^2}{8\pi} \right)_L}{t_{inst}} \quad (17)$$

from which we get (see (14))

$$\left(\frac{\delta B}{B_0} \right)_L \sim \frac{4 \cdot 10^2 \phi}{M u_{08} \eta \zeta} \quad \eta = \frac{\varepsilon_{rel}}{\left(\frac{\delta B^2}{8\pi} \right)_L} \quad (18)$$

From (18) we conclude that the process of acceleration in the weak turbulence limit is possible only for supersonic flows and for $\eta \sim 1$.

From (12), (13), and (18) we draw the main conclusion of the model. First of all we notice that long wavelength modes and Alfvén waves accelerate at almost the same rate:

$$\frac{t_{\text{acc}}^L}{t_{\text{acc}}^A} \sim \frac{2 v_{A8}}{\pi u_{08}} \left(\frac{B_0}{\delta B} \right)_L \quad (19)$$

Regarding the efficiency of the process (i.e., $t_{\text{acc}} < t_{\text{rs}}$) we can envisage two limiting cases (assuming δB less but of the order of B_0):

(i) For fast motion (u_{08} , possibly $v_{A8} \sim 10$) acceleration is effective even with smooth boundaries ($L_{21} \lesssim 1$). In this case the main difference from previous papers (Lacombe 1977; Eilek 1979; Ferrari et al. 1979) is that two different MHD perturbations can contribute to the acceleration.

(ii) For slow motion ($u_{08}, v_{A8} \sim 1$) the process is critically dependent on L_{21} . Again for very sharp boundaries ($L_{21} \lesssim 10^{-2}$) we have an efficient acceleration, but in the mean time the MHD modes evolve non linearly, unless they are well supersonic (see (18)). Conversely, for smooth boundaries we expect only nonlinear evolution of instabilities and no effective acceleration.

References

Blandford, R.D., Rees, M.J.: A 'Twin-Exhaust' model for double radio sources. *Mon. Not. R. Astron. Soc.* **169**, 395–415, 1974

- Eilek, J.A.: Particle reacceleration in radio galaxies. *Astrophys. J.* **230**, 373–385, 1979
- Ferrari, A., Trussoni, E., Zaninetti, L.: Relativistic Kelvin-Helmholtz instabilities in extragalactic radio sources. *Astron. Astrophys.* **64**, 43–52, 1978
- Ferrari, A., Trussoni, E., Zaninetti, L.: Magnetohydrodynamic instabilities and electron acceleration in extended extragalactic radio sources. *Astron. Astrophys.* **79**, 190–196, 1979
- Kulsrud, R., Ferrari, A.: The relativistic quasilinear theory of particle acceleration by hydromagnetic turbulence. *Astrophys. Space Sci.* **12**, 302–318, 1971
- Lacombe, C.: Acceleration of particles and plasma heating by turbulent Alfvén waves. *Astron. Astrophys.* **54**, 1–16, 1977
- Melrose, D.B.: Acceleration of ultrarelativistic electrons in the Crab Nebula. *Astrophys. Space Sci.* **4**, 165–181, 1969
- Melrose, D.B.: Resonant scattering of particles and second phase acceleration in the solar corona. *Sol. Phys.* **37**, 353–365, 1974
- Pacholczyk, A.G., Scott, J.S.: In situ particle acceleration and physical conditions in radio tail galaxies. *Astrophys. J.* **203**, 313–322, 1976
- Rees, M.J.: New interpretation of extragalactic radio sources. *Nature* **229**, 312–317, 1971
- Scheuer, P.A.G.: Models of extragalactic radio sources with a continuous energy supply from a central object. *Mon. Not. R. Astron. Soc.* **166**, 513–528, 1974

Received January 22, 1981, Revised Version June 1, 1981

Accepted June 4, 1981

Calibration of the Middle Triassic Time Scale by Conventional K–Ar and $^{40}\text{Ar}/^{39}\text{Ar}$ Dating of Alkali Feldspars

K.N. Hellmann and H.J. Lippolt

Laboratorium für Geochronologie der Ruprecht-Karls-Universität,
Im Neuenheimer Feld 234, D-6900 Heidelberg, Federal Republic of Germany

Abstract. Potassium-argon age determinations were made on 30 alkali feldspar samples from Middle Triassic bentonites of the “Grenzbitumenzone” of Monte San Giorgio (Ticino, Switzerland) which is stratigraphically well-defined by mollusc faunas. The feldspars are of primary volcanic origin (high sanidine) and of secondary authigenic origin (pure K-feldspar with low obliquity). The reported data were determined on the basis of samples from different vertical and horizontal positions within the rock sequence. The high sanidine has been dated at 233 ± 9 m.y., the authigenic feldspar at 226 ± 9 m.y. (IUGS constants, 1977). Stepwise heating experiments using the $^{40}\text{Ar}/^{39}\text{Ar}$ technique yield well-defined plateaus for the sanidine samples and support the former results. For various reasons the sanidines are thought to be originally volcanic and their K–Ar ages to be the age of eruption. Since the sediments of the Grenzbitumenzone were deposited close to the Anisian-Ladinian boundary, the high sanidine age can be used as calibration point for the Triassic time scale. An extrapolation based on biostratigraphic arguments favors a value close to 250 m.y. for the Permian-Triassic boundary.

Key words: Potassium-argon dating – $^{40}\text{Ar}/^{39}\text{Ar}$ dating – Triassic time scale – Grenzbitumenzone – Bentonite – Sanidine – Authigenic feldspar – X-Ray diffraction – Neutron activation-Ticino, Switzerland

Introduction

Dating on the Phanerozoic time scale is conducted on the basis of cross-calibration between paleontological correlations and isotopic data. Useful time-marking fossils for the Mesozoic are ammonoids which are cosmopolitan and have a high phylogenetic variability. Obviously, the definition of reliable chronometric calibration points requires thorough paleontological research above as well as below the object to be dated. Therefore, marine ash beds interbedded in fossiliferous sediments offer the best prerequisites for calibrating the fossil-based geological time scale.

Nevertheless, isotopic dating of volcanogenic sediments is often complicated by secondary rebedding, detrital contamination and/or secondary mineralization causing remobilization. Moreover, most submarine ash deposits are affected by halmyrolysis causing devitrification and argon loss. Therefore, potassium bearing minerals with high argon retentivity are the most useful minerals for potassium-argon dating.

Thus, the conditions which the material used for isotopic age determinations and for the definition of calibration points on

the biostratigraphic time scale must fulfil, can be summarized as follows:

1. the geological object to be dated should be extracted from primary and stratiform beds between fossiliferous strata of definite stratigraphic positions.
2. different components of the material used for dating must be clearly separable from detrital constituents.
3. there must be various cogenetic minerals of primary origin datable by different methods.
4. the material must be a closed system of radioactive and radiogenic isotopes.
5. the measured age must correspond to the real age of primary crystallization and thus correlate with the biostratigraphic age.

The aim of this study was to define a new calibration point on the Phanerozoic time scale which is significant in relation to biostratigraphic dating as well as to the experimental results. The lack of reliable age data suitable as calibration points for the Mesozoic made efforts in this direction quite challenging.

The Middle Triassic volcanisms of the Southern Calcareous Alps best fulfil the above-mentioned conditions. This is particularly the case for the Anisian-Ladinian boundary in Ticino, Switzerland. There are two reasons for giving preference to this region:

1. the stratigraphic position of the so-called “Grenzbitumenzone” (Frauenfelder 1916) is exactly defined by mollusc faunas.
2. the rock sequence contains volcanogenic intercalations (bentonites) already investigated from sedimentologic and petrographic aspects (Müller et al. 1964).

Among the phenocrysts in the bentonites only high temperature alkali feldspars are suitable for isotopic dating.

Geological Setting

The Mediterranean part of the Tethian geosyncline is characterized by intensive volcanic phases during the Permo-triassic and the Tertiary periods. Figure 1 compiles occurrences of Triassic volcanic rocks in the Mediterranean region which at first were considered candidates for our study. By the evaluation of rock qualities and trustworthiness of their stratigraphic correlation we came to the opinion that the area of this study (D1 in Fig. 1) was most favourable. All the other occurrences of Mesozoic volcanic or intrusive rocks disappointed us as far as their correlation with the geological time scale and/or the mineralogical environment were concerned.

Triassic volcanic rocks of the Mediterranean region

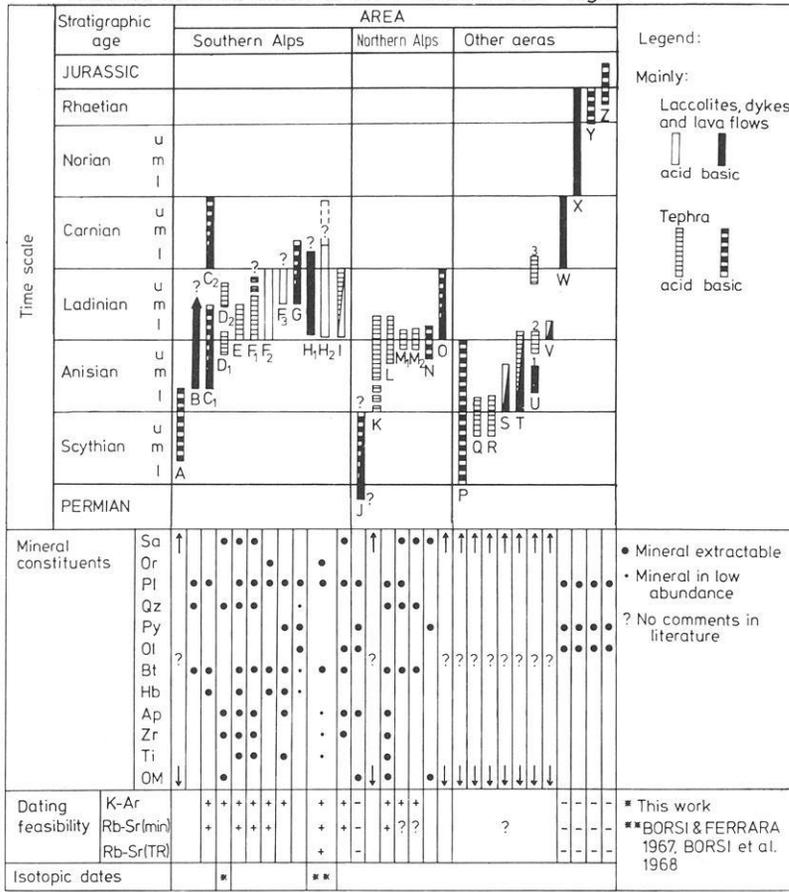


Fig. 1. Stratigraphic comparison of Triassic volcanic rocks in the Mediterranean region and evaluation of their dating feasibility. Rocks of column D₁ were chosen for this study. The notations A to Z stand for: A Loibl-Paß, Gartnerkofel (Kahler and Kahler 1953), B Valle del Dezzo, Lombardy (Assereto and Casati 1965), C₁ Malpensata/M. Guglielmo, C₂ Val Sabbia (Bianchi et al. 1971), D₁, D₂ Varesotto area (Müller et al. 1964; Wirz 1945), E Eastern Dolomites (Caprile, Marmolada, Val Gardena, Latemar (Baccele and Sacerdoti 1965; Callegari and De Pieri 1967), F₁₋₃ Recoaro - Schio - Posina area, Vicentinian Alps (Hummel 1931; Zanettin and De Vecchi 1968; Ott 1972; Assereto 1969; De Vecchi et al. 1974), G Western Dolomites (Hummel 1932; Rossi 1962; Sacerdoti and Somnavilla 1962), H₁, H₂ Predazzo (H₂)-Monzoni (H₁) area (+intrusive rocks) (Paganelli and Tiburtini 1964; Vardabasso 1945; Simboli 1966; Delmonte 1967; Borsi and Ferrara 1967; Borsi et al. 1968; Gallitelli and Simboli 1970), I Carnian Alps (Tarvisiano) (Spadea 1970; Castellarin and Pisa 1974; Assereto et al. 1968), J Salzkammergut/Hallstatt (Cornelius 1936, 1941; Zirkel 1957), K Schneeberg area, Mieminger mountains, Rhätikon, Wetterstein massive, Lechtal, Karwendel (Spengler and Stiny 1931; Cornelius 1951; Schmidegg 1951; H. Miller 1965; Kobel 1969), L Bavarian Alps (Vidal 1953), M₁ Großreifling/Styria, Innsbruck/Austria (Höllner 1963; Plöching and Wieseneder 1965; Gessner 1966; Summesberger and Wagner 1972; Bechstädt and Mostler 1974), M₂ Göstling/Ybbs, Austria (Faupl and Hamedani 1972), N Dobratsch, Gailtal/Austria (Pilger and Schönenberg 1958), O Lechtal/Austria (Mutschlechner 1954), P NW Bulgaria (Ganev in Zapfe 1974), Q Chios/Eastern Aegaic (Bender 1970; Assereto 1974; Jacobshagen and Tietze in Zapfe 1974), R Kithaeron mountains/N-Attika, Greece (Bender 1967), S Vilajets Skutari/Albania (Nopcsa and Reinhard 1912), T Carpathians, Montenegro (Ramovs in Zapfe 1974; Kuthan 1959), U Bükk mountains, Balaton, Bagolyhegy, Hungary (Balogh in Zapfe 1974), V Slovenia (Ramovs in Zapfe 1974), W Pyrenees (Misch 1934), X Keltiberian mountains (Richter and Teichmüller 1933), Y Calcena/Moncayo, Arcos de las Salines, Spain (Dubar 1925), Z Eastern Pyrenees (Dubar 1925)

In Ticino (Swiss Southern Alps) stratigraphically well-defined volcanic rocks occur. In the region of Lugano there are thin layers of bentonites (altered volcanogenic ejecta) interbedded in a series of fossiliferous bituminous dolostones and argillaceous shales. The fossil record relates this sequence to the Anisian-Ladinian boundary and has therefore been called the "Grenzbitumenzone" (GBZ, "bituminous boundary layer") by Frauenfelder (1916). Outcrops of the GBZ are mainly confined to the area of Monte San Giorgio, south of Lake Lugano.

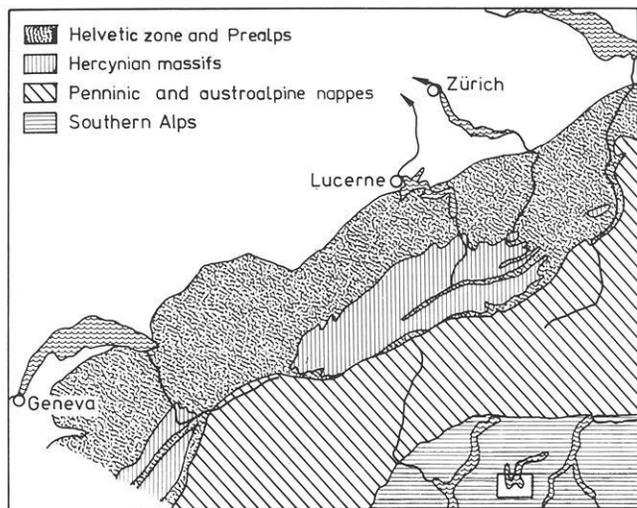
Seen geologically the Monte San Giorgio area is a fault block, flat-dipping to the south (20-30°), consisting of unfolded conformable beds with a total thickness of more than 800 m from Lower Triassic up to Jurassic, resting concordantly upon various Permian rocks (Fig. 2b).

The GBZ close to the Anisian-Ladinian boundary is a

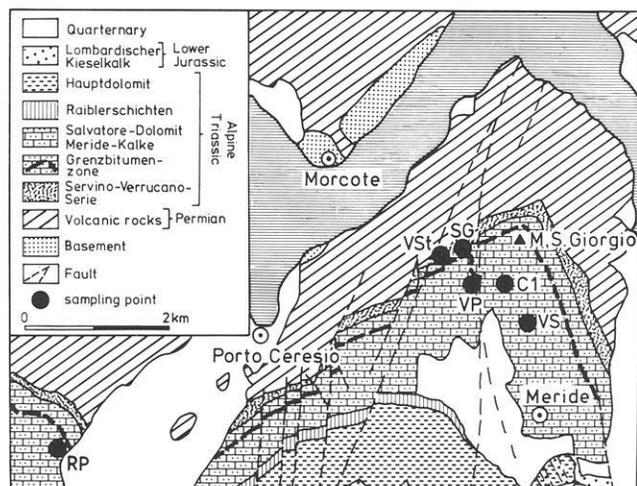
rhythmic alternation of early diagenetic dolomitized limestones and bitumenous shales (±12 m thickness) conformably interstratified between the laying massive Anisian dolomite and the hanging Ladinian Meride limestone. During the excavation of Middle Triassic vertebrates from 1950-1968 (Kuhn-Schnyder 1964; 1974) the GBZ profile was mapped by paleontologists from Zurich. They divided the whole sequence into layers (nos. 1-180), six of which were recognized as volcanic sediments.

Relation of the "Grenzbitumenzone" to the Triassic Zonal Successions

The Monte San Giorgio area is famous for its abundant deposits of Middle Triassic marine fishes and reptiles (Kuhn-Schnyder 1964). The exact chronostratigraphic position of the GBZ, how-



a: Location of sampling area



b: Sampling area and sample locations

Fig. 2. a Location (rectangle on the lower right side) and b geological sketch of the sampling area in the Monte San Giorgio region south of Lake Lugano (Ticino, Switzerland)

ever, is defined by mollusc faunas (several ammonoid genera; Rieber, 1973b). In the twenties it was already known that the GBZ fauna belongs to a Middle Triassic interval close to the Anisian-Ladinian boundary (Frauenfelder 1916; Senn 1924).

In this context we have to recall what is known about the stratigraphic significance of the GBZ, especially taking into account the results of the research done at the Mirigioli outcrop of Monte San Giorgio (SG in Fig. 2b; point 902 on folio 1973 Mendrisio 1:25000).

Within the excavated section 44 species and subspecies of 15 ammonoid genera and 12 *Daonella* species (lamellibranchiata of the family Halobiidae) have been found (Rieber 1969, 1973a). The small time range of these fossils permits a highly accurate resolution of the GBZ strata.

The ammonoids are abundant mainly in the dolomite beds of the lower and middle part of the GBZ. Among them Ceratitidae are clearly preponderant. Trachyceratidae are frequent in some beds of the upper part of the sequence. The genus *Protrachyceras Mojsisovics* makes its first appearance in bed no. 98 of the section (compare Fig. 3) thus establishing the lower limit of the Ladinian between bed no. 97 and bed no. 98. Based on *Daonellae*, the Anisian-Ladinian boundary would coincide with the interval between bed no. 83 and bed no. 104 of the Mirigioli outcrop (Rieber 1969). Comparable *Daonella* faunas are known from profiles of the Tridentinian Alps and the Dolomites in Southern Tyrol belonging to the latest Anisian and earliest Ladinian.

The Triassic chronostratigraphy, however, is defined only by the zonal succession of ammonoids. Ammonoids are remarkably cosmopolitan and therefore they are of great use for correlating Triassic rocks worldwide (Tozer 1971). Genera of six families of sub-order Ceratitina which are found at Monte San Giorgio are also represented in the Middle Triassic of North America (Silberling and Tozer 1968).

Figure 4 shows the relation of the GBZ to the standard zonal successions of the Alpine-Mediterranean and the North American Triassic. Taking into account the greatest possible error, the GBZ definitely lies above of the Mediterranean *Paraceratites trinodosus* zone and below the *Protrachyceras curionii* zone, which are the middle of *Gymnotoceras meeki* zone and the base of *Protrachyceras subasperum* zone of North America respectively. This corresponds to an interval of two faunal zones. The latest Anisian of the GBZ is characterized by a new faunal

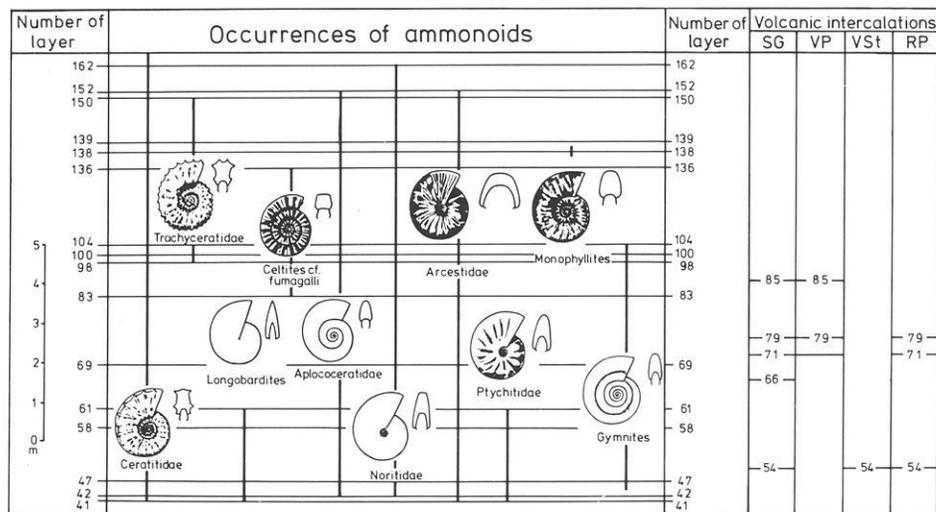


Fig. 3. Occurrence, range and subdivision of stratigraphically significant ammonoids in the Monte San Giorgio limestones (Rieber 1973b) combined with the sequence of bentonites at various outcrops of the area. The numbers characterize the stratigraphic position of fossils and volcanic layers. (SG Mirigioli outcrop, VP Val Porina mine gallery, VSt Valle Stelle gorge, RP quarry near Rossago/Pogliana)

STAGES	SUBSTAGES	ZONES	Grenzbitumen-Zone (RIEBER 1969, 71, 73)	ZONES
		Alpine-Mediterranean Triassic (ASSERETO 1966, 69, 71)		North America (Silberling & Tozer 1968)
LADINIAN	UPPER	Longobard	Protrachyceras archelaus	Frankites sutherlandi Maclearnoceras maclearni Meginoceras meginiae
	LOWER	Fassan		Protrachyceras curionii (Protrachyceras reitzii)
ANISIAN	UPPER	Illyr	Aplococeras avisianus Paraceratites trinodosus	Ticinites polymorph. Gymnotoceras occidentalis Daonella dubia Gymnotoceras meeki Daonella moussoni D.cf.elongata Daonella sturi Gymnotoceras rotelliformis Daonella americana
		Pelson		(Balatonites-Horizon)
	MIDDLE	Bithyn	Anagymnotoceras ismidicus	Anagymnotoceras varium
		LOWER	Aege	(Paracrocordiceras- beds)

The chronostratigraphic position of the "Grenzbitumenzone", dated by ammonoids (Ceratitidae, Hungaritidae, Aplococeratidae, Trachyceratidae) compared with the revised zonal sequence of the Alpine Mediterranean Triassic (left side) and the standard zones of North America (right side)

Fig. 4. Relation of the "Grenzbitumenzone" of Monte San Giorgio to the standard zone succession of the Alpine-Mediterranean (left) and the North American Triassic (right)

zone, the zone of *Ticinites polymorphus* RIEBER. This zone may be partly equivalent to the *Aplococeras avisianus* zone defined by Assereto (1969) in the Southern Alpine Triassic taking into account the faunal content. The definition of the upper limit of the *polymorphus* zone in Fig. 4, column 4 (dashed line) as coinciding with the Anisian-Ladinian boundary of the Alpine-Mediterranean succession is justified by the fact that bed no. 98 contains the first Trachyceratidae represented by the genus *Protrachyceras*. Rieber (1973a) considers the layers above bed no. 97 to belong to the *reitzii* zone, even though *Protrachyceras reitzii* has not been found. This means that the GBZ also includes the lower part of the Fassan substage, which is indicated by the top line in Fig. 4, column 4. Admittedly, Kozur (1974) doubts that genus *Protrachyceras* appears isochronously on a global scale. He pointed out that this fossil already occurs within the *avisianus* zone. This could mean that the GBZ is really confined to the upper part of the *Aplococeras avisianus* zone and therefore to the latest Anisian. According to Kozur (1974), Silberling & Tozer (1968) have not taken into account representatives of *P. reitzii* when they defined the Anisian-Ladinian boundary because the significant development of genus *Protrachyceras* does not begin until the *P. subasperum* zone.

To sum up briefly we can say that the GBZ is definitely younger than the *trinodosus* zone, older than, or possibly isochronous with the *reitzii* zone of the Alpine standard succession and probably of the same age as parts of the *meeki* and *occidentalis* zone of the North American zonal succession. This implies that the GBZ was sedimented during an interval which lasted as long as one or may be two zones. The exact stratigraphic position therefore may have an error of \pm one zone. Assuming that the 34 Triassic zones of more or less the same zonal range have lasted 40–50 m.y., the stratigraphic error is in the order of 1.2–1.6 m.y. This, however, lies within the experimental errors of chronometric dating.

Composition of the Tuffs from the Grenzbitumenzone and Its Implications

The volcanogenic beds are intercalated within the lower to middle part of the GBZ (from bed nos. 29–85) and consist of pure tuffs altered to bentonites without any detrital admixture. The thickness of the layers varies considerably (from a few mm up to several cm). The base of the bentonite beds are often crystal tuff layers, the thickness of which depends on the thickness of the bentonites. This may be due to silica migration resulting from the alteration of the tuffs during diagenetic processes. The volcanic material consists of a micro-crystalline groundmass of illite-montmorillonite composition which contains a small portion of intratelluric phenocrysts of alkali feldspar, quartz and, in some layers, leached euhedral biotite. Euhedral crystals of zircon and apatite are accessory components. Plagioclase, however, is lacking in all horizons. This absence of plagioclase may be an original feature but later alteration into montmorillonite by halmyrolysis cannot be excluded (Helgeson 1972; Wirsching 1976).

The phenocrysts are normally fragments of euhedral crystals with grain sizes less than 1 mm. They amount to $\pm 5\%$ of the samples. The groundmass fraction (less than 2 μm) consists of a mixed-layer mineral of montmorillonite and illite (Müller et al. 1964). The volcanic layers show graded bedding which explains their tuffogenic origin. The mainly splintery, sharp-edged to euhedral phenocrysts, their random distribution in the beds and the lack of detrital compounds are features of aeolic transport followed by sedimentation in the Tethys Sea. Secondary transport and reworking can be excluded.

Sample Val Serrata (VS in Fig. 2b) of our study is stratigraphically younger than the GBZ (from the base up to Lower Ladinian). Contrary to the other samples, it is a tuffite in the sense that it is solidified by calcareous cement. Its origin may

be very similar to the genesis of the GBZ bentonites since it contains quartz and feldspar phenocrysts in graded bedding but no detrital constituents.

Since feldspar turned out to be the only mineral in these tuffs which could be used for dating, we were confronted with the difficulty that only K–Ar measurements could be carried out. Since the Rb/Sr ratio of these feldspars is around 2 (Rb ~ 100 ppm, Sr ~ 50 ppm), Rb/Sr dating would have been possible if a second mineral of volcanic origin had been present in sufficient amounts to permit its separation and the analysis of the initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratio. Unfortunately the apatite in the tuff samples did not meet this requirement.

Because no second supporting dating method could be applied in this case, we were forced to look for other criteria to check the internal consistency of our dating study. We therefore carefully sampled the material for the K–Ar measurements. This implied a detailed examination of the feldspar types in the tuffs on the one hand, and the preparation of different grain size fractions from those samples from which enough feldspar could be extracted on the other hand. As a further approach we decided not only to apply the conventional K–Ar dating method but also $^{40}\text{Ar}/^{39}\text{Ar}$ dating with stepwise heating.

Remarks on K–Ar Dating of Feldspar

Müller et al. (1964) have shown that feldspars contained in the bentonites of Monte San Giorgio are only alkali feldspars. As we shall show later the alkali feldspars are sanidine and adularia-like potassium feldspars which may be of authigenic origin. The use of alkali feldspar for K–Ar dating is not unequivocal. Orthoclase and microcline are normally not considered suitable because they obviously do not retain argon as well as sanidine and anorthoclase (compare Dalrymple and Lanphere, 1969, pp 168–170). In general, K–Ar ages will only be geologically meaningful if the object to be dated has initially been free of radiogenic argon and has remained a closed system for argon and potassium since its crystallization.

In the case of the binary system of alkali feldspar the closed system condition can be affected by subsolidus transformations. They are caused by a rearrangement of Si and Al atoms within the lattice in the course of slow cooling of the rocks (diffusive transformation; Laves 1960) and by exsolution or separation of the binary system into sodium and potassium phases.

Random distribution of Al and Si within the lattice of high alkali feldspars is only preserved by rapid quenching near 1,000°C (Stewart and Wright 1974). Long term annealing at temperatures below 800°C causes the aluminium to migrate to electrostatically preferable $T_1(0)$ sites. Because of this the crystal structure becomes more and more triclinic. Only rapid cooling of a melt produces monoclinic potassium rich alkali feldspar of high sanidine structure with maximum Al, Si disorder.

In the course of slow cooling the high sanidine structure will be affected by formation of triclinic domains which produces an apparent integrated monoclinic structure characteristic of "orthoclase" (low sanidine), or it will break down to microcline with maximum Al, Si order, often accompanied by separation of potassium and sodium phases.

The usefulness of high sanidine as a K–Ar clock has long since been established by many authors (Byström-Asklund et al. 1961; Baadsgaard and Dodson 1964; and others). Argon diffusion experiments in a vacuum could be interpreted by simple diffusion mechanisms (Baadsgaard et al. 1961; Frechen and Lipolt 1965; compare Kalbitzer and Fechtig 1966 for general

discussion). The argon activation energies are comparable to those of biotites (57 Kcal/mol for the Kinnekulle bentonite sanidine, for instance). The K–Ar results on orthoclase and microcline are confusing. Some data agree quite well with those of cogenetic biotite and hornblende, others not. It is not possible to predict a priori that orthoclase and microcline will yield inconsistent K–Ar results. Perthitization (Sardarov 1957) and obliquity play an important role. The patterns of argon diffusion rates in microcline are quite complicated (Evernden et al. 1960). Low Ar activation energies may be simulated by Ar diffusion from internal surfaces caused by perthitization (Gentner and Kley 1957).

On the contrary, homogeneous orthoclase shows an Ar diffusion pattern comparable to that of sanidine and an activation energy of 44 Kcal/mole (Foland 1974a, b). Na and K diffusion experiments suggest in the case of homogeneous orthoclase and of high sanidine that Ar diffusion mainly takes place on lattice sites of alkali ions, whereas in the case of inhomogeneous orthoclase, triclinic domains are also affected by surface diffusion effects.

The influence of triclinic domains on Ar retentivity has been studied on the basis of $^{40}\text{Ar}/^{39}\text{Ar}$ stepwise heating experiments of adularias with different obliquities (Halliday and Mitchell 1976). Saddle-shaped $^{40}\text{Ar}/^{39}\text{Ar}$ spectra were observed, while feldspar with obliquity equal zero yield nearly model release patterns for undisturbed K–Ar systems (rhyolite sanidine, measured by Dalrymple and Lanphere 1974).

In summary we may conclude that the Monte San Giorgio bentonites provide suitable dating material, whereas other feldspar types cannot be considered suitable as long as they have not been tested systematically.

Little is known about the Ar retentivity of authigenic alkali feldspar formed at surface p, T conditions. Most of them have triclinic symmetry. Monoclinic crystals are only found growing on primarily existing nuclei of detritic origin (Füchtbauer 1950, Kastner 1971). Since subsolidus and solvus reactions in authigenic feldspar can be excluded, argon loss is not expected a priori. For an Upper Jurassic tuff from the Southern Alps, Hunziker (1979) found K–Ar ages of sanidines with authigenic K-feldspar rims, which scatter between 136 and 100 m.y., depending on the relative amounts of rim and core analysed. Hunziker assumes on the basis of investigations by Bernoulli and Peters (1974) that the authigenic feldspar rims are products of a later devitrification of the tuff. A value of about 100 m.y. can be derived as devitrification date. This result has to be kept in mind when discussing our results on a presumably authigenic feldspar phase. In case the authigenic feldspar was formed at nearly the same time as the tuff deposition of the bentonite no substantial effect on the ages can be expected. However, if we are dealing with devitrification taking place much later than the formation of the volcanic layers, $^{40}\text{Ar}/^{39}\text{Ar}$ stepwise heating experiments could be of help in distinguishing between the formation of core and rim.

Sampling and Dressing Procedures

Samples were taken from five sediment outcrops belonging to the GBZ (SG, RP, Vst, VP in Fig. 2b). In addition, material was used from the excavation at Cassina (C1, Lower Ladinian) and from the tuffite within the Lower Meride Limestone of Val Serrata (VS, Lower Ladinian, 100 m above the GBZ). Sample C1, however, did not contain any datable minerals.

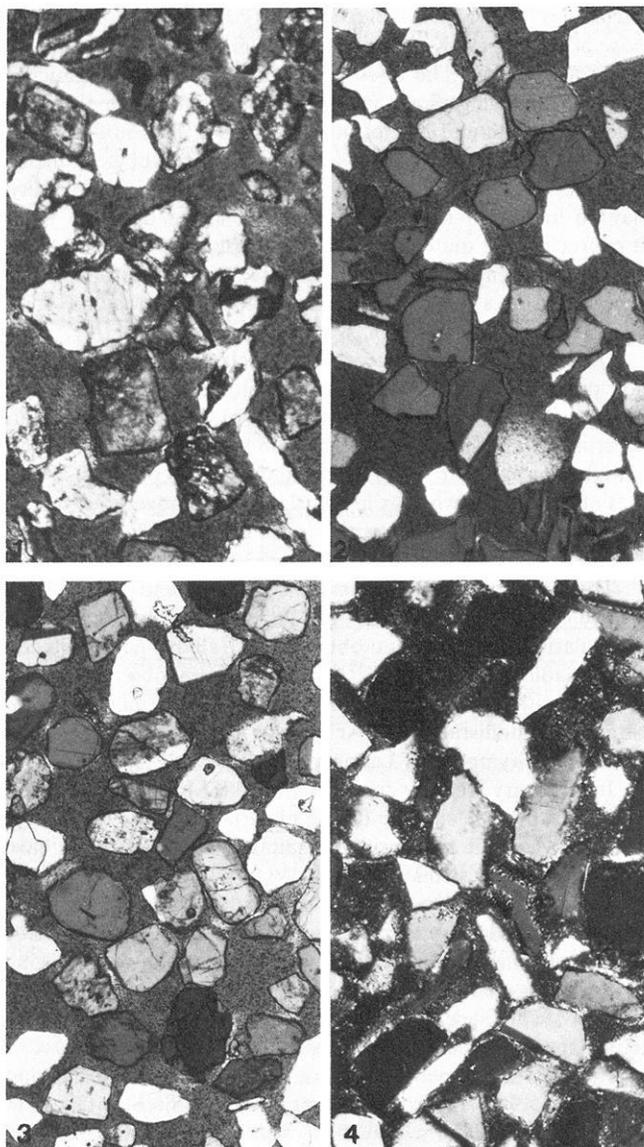


Fig. 5. Photomicrographs of four feldspar concentrates from the bentonite of bed no. 79 at the Mirigioli outcrop (Monte San Giorgio). Grain sizes 200–315 μm , 315–500 μm in Fig. 5-3. 1 Authigenic K feldspar showing inhomogeneous extinction and beginning caolinitization (type W in Figs. 6 and 9), 2 clear high sanidine, poor in inclusions (type G), 3 round-edged, cloudy sanidine with inclusions (type R), 4 Sharp-edged, cloudy sanidine of high sanidine quality with overgrowth of authigenic feldspar (type T)

We selected five horizons of the Mirigioli outcrop (SG) and one horizon of the other GBZ outcrops in such a way that three horizons (54, 79, 85) are represented by two independently taken specimens. In total nine different samples were processed. The original samples weighed 20 to 50 kg.

The preparation of the mineral concentrates was begun by decomposing the rocks with suitable solutions, (benzine, hydrogen peroxide) which do not attack the minerals selected for dating. After this treatment the decomposed samples consisted mainly of phenocrysts in their original habit. They were washed several times in twice-distilled water and sieved into five grain size fractions. The fractions 0.1–0.2 mm, 0.2–0.315 mm and 0.315–0.5 mm were chosen for further dressing work. After drying the phenocryst concentrates the feldspars were separated

Table 1. Samples of volcanic feldspar prepared from Grenzbitumenzone bentonites and a Lower Ladinian tuffite (VS) of the Monte San Giorgio area, Ticino, Switzerland. (G, W, T: feldspar types described in the text, M mixtures of the three types)

Grain size (μm)	Feldspar types			
	M	G	T	W
100–200	–	–	–	SG 66
	SG 71	–	–	–
	SG 79	SG 79	SG 79	SG 79
	–	RP 79	–	–
	–	VP 85	VP 85	–
200–315	–	–	–	SG 66
	SG 79	SG 79	SG 79	SG 79
	–	–	–	RP 79
	–	SG 85	–	–
	–	VS	–	–
315–500	–	–	–	SG 66
	SG 79	SG 79	SG 79	SG 79
	–	VP 85	VP 85	VP 85
500–1,000	–	VP 85	–	–

by heavy liquids and magnetic separation. The remaining minerals were examined under the binocular microscope for further datable constituents, however without success. Detrital compounds have not been found in any of the mineral concentrates.

The feldspars are alkali feldspars. Optical and morphological differences were observed among the feldspar crystals. On the basis of these differences three feldspar types could be distinguished:

1. A splintery, hypidiomorphic, homogeneous and water-clear type, called "G" (compare Fig. 5, no. 2)
2. A cloudy, partly round-edged type, called "T" (Fig. 5 no. 4)
3. An optically inhomogeneous, hypidiomorphic up to euhedral feldspar type, rich in inclusions, called "W" (Fig. 5, no. 1).

Since a detrital provenience of the round-edged feldspar type "T" could not be excluded a priori, we decided to aim for pure concentrates of the different types, apart from the feldspar mixtures (symbol "M", Fig. 5, no. 3). The selection of the different feldspar types could only be done by handpicking grain by grain. The degree of sample purity reached in this manner was better than 95%. The residual 5% consisted of the two other feldspar types. Each feldspar concentrate was treated with ultrasonics till no further crushing of mineral grains was detectable. The samples were dried and divided into aliquots for different investigations. Each aliquot was examined again under the microscope for its composition. This procedure resulted in a sample reservoir of about 2 g, representing 26 different feldspar samples from seven sampling sites. Bed 54 of the Mirigioli outcrop (SG) and of the Valle Stelle hollow did not contain feldspar; only quartz and montmorillonite were found. Table 1 summarizes the final state of our sample dressing. The majority of the samples consists of the various types and grain sizes of the Mirigioli bentonites 79 and 85.

Experimentation and Results

Structural Examination by X-Ray Diffraction

The structure of the different feldspar types extracted from the tuffs were identified with powdered aliquots using the Phillips X-ray diffractometer of the Mineralogical Institute of the Univer-

Table 2. Certain lines of X-ray diffraction scans (Cu K α 1) from different feldspar types from Monte San Giorgio bentonites

Reflections (<i>hkl</i>)	Feldspar type	
	G and T 2 θ ($^\circ$)	W 2 θ ($^\circ$)
($\bar{2}01$)	21.26 \pm 0.01	20.96 \pm 0.02
(111)	22.67 \pm 0.01	22.51 \pm 0.02
(130)	23.60 \pm 0.01	23.50 \pm 0.02
($\bar{1}12$)	25.74 \pm 0.02	25.71 \pm 0.02
(220)	27.14 \pm 0.02	26.81 \pm 0.02
($\bar{2}02$)	27.29 \pm 0.02	27.11 \pm 0.03
(040)	27.38 \pm 0.02	—
(002)	27.66 \pm 0.01	27.60 \pm 0.02
(131)	29.94 \pm 0.01	29.81 \pm 0.02
(041)	30.76 \pm 0.02	30.78 \pm 0.02
($\bar{1}32$)	32.36 \pm 0.02	32.35 \pm 0.02
($\bar{2}41$)	34.94 \pm 0.02	34.72 \pm 0.02
(060)	41.60 \pm 0.01	41.63 \pm 0.02
($\bar{2}04$)	50.88 \pm 0.01	50.80 \pm 0.03
(280)	62.00 \pm 0.01	62.01

sity of Heidelberg. Scans were run repeatedly between 65 $^\circ$ and 5 $^\circ$ at a speed of 1/4 $^\circ$ (2 θ) per min (time constant position 4). Si metal added to each smear mount was used as an internal standard. The reflections selected for spacing were evaluated in accordance with Orville (1967) and identified on the basis of the powder patterns developed by Wright and Stewart (1968).

The most important 2 θ angles are given in Table 2. In this connection the following features must be noted:

1. the pertinent reflections appear at the same angles for the feldspar types G and T, but at significantly different angles for type W.
2. the d-values based on the ($\bar{2}01$) reflections are comparable for all types, within the experimental error, to those reported earlier by Müller et al. (1964) for the tuffogenic feldspar of the Monte San Giorgio.
3. the (130) and (131) reflections for feldspar types G and T are sharp single lines without peak broadening. Therefore the presence of a triclinic component can be excluded. Type W however, shows broadened, diffuse peaks caused by lattice inhomogeneities.
4. the $\Delta 2\theta$ ((040)-(002)) value for the feldspars G and T amounts to 0.27 $^\circ$ and is close to that for high sanidine (Borg and Smith, 1969). For type W this value could not be determined because of peak broadening.
5. the d-values obtained from the feldspar types G and T are comparable to those from the Drachenfels sanidine, Eifel (ASTM 19-11227, 1967) whereas the d-values from the feldspar type W are similar to adularia from Selkingen, Switzerland (ASTM 19-931, 1966).

Apart from this it should be mentioned that the 20.95 $^\circ$, 2 θ line found in the type W spectrum can also be seen as a small peak near the ($\bar{2}01$) reflection of type T.

Least squares unit-cell refinements were calculated using a computer program which was developed by Nuber started from calculations made by Berdesinski and Nuber (1966). Only single, well-indexed lines of which the 2 θ values did not scatter more than $\pm 0.03^\circ$ in repeated runs were used. The results are plotted in Fig. 6.

The lattice constants of the water-clear feldspar type G and the cloudy type T do not differ significantly. They correspond

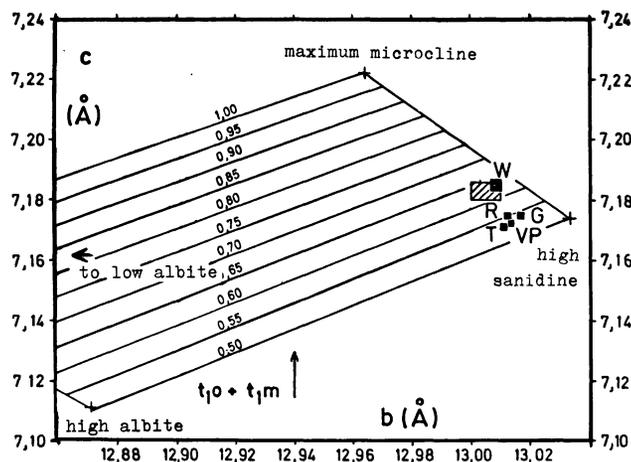


Fig. 6. Lattice parameters of four alkali feldspar types from various bentonite beds of the Monte San Giorgio region (G, VP clear high sanidine; T core of cloudy, sharp-edged high sanidine; R round-edged high sanidine; W authigenic K-feldspar crystals and rims of T and R). Four samples with nearly complete Al,Si disorder ($T_1=0.55$) lie in the high sanidine corner, whereas sample W shows more triclinity ($T_1=0.64$). The sizes of the symbols correspond to the analytical errors. The size of the error rectangle for sample W (cross-hatched) is due to some diffuse reflections caused by triclinic domains

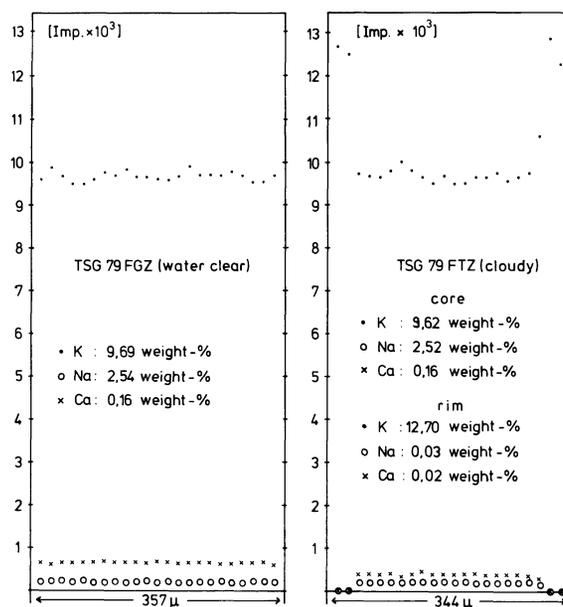


Fig. 7. Concentration profiles of K, Na and Ca across crystals of the feldspar types G (left) and T (right) showing the overgrowth of the sharp-edged cloudy feldspar crystals. (electron microprobe measurements 15 kV/10 μ A, beam diameter 3 μ m, integration time 20 s., relative errors $\pm 1\%$)

to high sanidine. The type W feldspar, however, has lattice constants equal to those of monoclinic authigenic K-feldspar as reported by Baskin (1956, Table 8) within the experimental errors. Because of the diffuse nature of the X-ray reflections for type W, it was not possible to determine whether the lines are actually split. Fine twinning of triclinic domains can result in an apparent monoclinic symmetry (Goldsmith and Laves, 1954). The broad and diffuse peaks may therefore be due to a domain structure resulting from poor crystallographic continuity.

Table 3. K–Ar data and ages of volcanogenic and authigenic alkali feldspars of certain grain sizes from one bentonite bed (no. 79) of Mirigioli-outcrop (point 902) at Monte San Giorgio, Ticino (Switzerland). The determined ages of high sanidine and high sanidine with authigenic overgrowths are comparable within the errors

Material analysed	Grain size (mm)	Weight (mg)	K (%)	Rad. ^{40}Ar ($\times 10^{-6}$ cm 3 STP/g)	Atmos. ^{40}Ar (%)	Calculated age (m.y.)
SG 79 type G (primary high sanidine)	0.315–0.500	34	9.95	90.8	3.7	221 \pm 7
				91.5	3.5	222 \pm 7
	0.315–0.500	44	9.95	96.2	1.5	233 \pm 7
				95.6	1.7	232 \pm 7
	0.200–0.315	45	9.47	91.8	1.2	234 \pm 7
				92.5	1.4	235 \pm 8
	0.100–0.200	40	9.27	83.6	1.9	218 \pm 7+
				83.9	2.0	219 \pm 7+
	0.100–0.200	52	9.27	89.8	0.6	233 \pm 7
				89.6	0.8	233 \pm 7
Mean age:						228 \pm 7
Mean age without +:						230 \pm 6 (\pm 11)
SG 79 type T (primary high sanidine with authigenic feldspar rims)	0.315–0.500	39	9.83	93.7	2.2	230 \pm 7
				94.9	2.2	233 \pm 8
	0.315–0.500	41	9.83	96.8	1.3	237 \pm 7
				97.1	1.4	238 \pm 8
	0.200–0.315	45	10.07	98.1	1.8	235 \pm 7
				98.1	1.8	235 \pm 8
	0.100–0.200	39	10.17	99.6	1.7	236 \pm 7
				98.9	1.7	234 \pm 8
Mean age:						235 \pm 3 (\pm 6)
SG 79 type W (authigenic K-feldspar)	0.315–0.500	38	12.12	108.7	2.4	217 \pm 6
				110.0	2.5	220 \pm 7
	0.315–0.500	37	12.12	113.5	1.0	226 \pm 7
				113.8	1.1	227 \pm 7
	0.200–0.315	44	12.22	115.1	1.5	227 \pm 7
				114.7	1.5	227 \pm 7
	0.100–0.200	45	11.62	110.0	1.4	229 \pm 7
				110.3	1.4	229 \pm 7
Mean age:						225 \pm 4 (\pm 9)
SG 79 type M (three component feldspar mixture)	0.315–0.500	38	10.15	95.2	1.7	226 \pm 7
				96.5	1.8	229 \pm 7
	0.315–0.500	39	10.15	97.7	3.3	232 \pm 7
				97.6	3.3	231 \pm 7
	0.315–0.500	39	12.18	116.3	1.8	230 \pm 7
				116.8	2.0	231 \pm 7
	0.315–0.500	39	10.57	101.1	2.0	231 \pm 7
				101.3	2.2	231 \pm 7
	0.200–0.315	44	10.16	96.8	1.7	230 \pm 7
				96.7	1.9	230 \pm 7
0.100–0.200	44	10.19	98.1	2.1	232 \pm 7	
			98.2	2.2	232 \pm 7	
Mean age:						230 \pm 2 (\pm 4)

Using decay constants $\lambda_e = 0.581 \times 10^{-10} \text{a}^{-1}$, $\lambda_\beta = 4.962 \times 10^{-10} \text{a}^{-1}$, and atomic abundance $^{40}\text{K}/\text{K} = 1.167 \times 10^{-4}$

Errors in K = 1.0% (standard deviation of the average on 14 determinations), errors in ^{40}Ar (rad) $\pm 2.5\%$ per each individual determination. The \pm figure is the estimate of precision of the determination at 1 s.d. criterion, 2 s.d. in parentheses

Potassium and Microprobe Analyses

The potassium analyses were carried out by emission flame photometry using a Zeiss spectrometer and applying a modification of Cooper's (1963) method. Standard and sample solutions are

buffered with Na and Ca. The measurements are checked by repeated analyses of the NBS 70a standard feldspar, for which Flanagan (1974) recommends 9.80% K as reference. Long time determinations in our laboratory gave $9.78 \pm 0.04\%$ potassium (1 s.d. error, 14 values).

Table 4. Conventional K–Ar and $^{40}\text{Ar}/^{39}\text{Ar}$ total fusion ages on volcanic and authigenic alkali feldspars and feldspar mixtures of certain grain sizes from different bentonite beds of the Mirigioli outcrop (except bentonite bed no. 79) and other localities (sample number corresponds to the number of bentonite bed)

Locality and material analyzed	Grain size (mm)	Weight (mg)	K (%)	Radiogenic ^{40}Ar ($\times 10^{-6}$ cm ³ STP/g)	$^{40}\text{Ar}/^{39}\text{Ar}$	atm ^{40}Ar (%)	Calculated age (m.y.)
loc. Val Serrata (Lower Ladinian) (717.740/084.300, fol. 1373 Mendrisio)							
VS-G (high sanidine)	0.200–0.315	20	–	–	70.477	2.4	227 ± 8
					70.160		226 ± 8
	0.200–0.315	21	–	–	70.477	2.8	223 ± 7
					70.160		223 ± 6
						Mean age:	225 ± 2 (± 4)
loc. Val Porina (716.580/085.300)							
VP 85-G (high sanidine)	0.500–1.000	19	–	–	71.264	2.0	229 ± 8
		12	–	–	70.881	2.3	228 ± 8
	0.315–0.500	40	9.31	89.5		2.0	232 ± 8
				91.1		2.0	236 ± 8
	0.100–0.200	46	9.46	91.6		2.0	233 ± 7
				91.3		2.2	232 ± 8
						Mean age:	232 ± 3 (± 6)
VP 85-T (high sanidine with authig.) feldspar rim	0.315–0.500	43	9.48	88.8		1.6	226 ± 7
				88.5		1.7	225 ± 7
loc. Quarry Rossago/Pogliana (711.100/082.800)							
RP 79-W (authig. K-feldspar)	0.200–0.315	27	13.24	123.4		2.1	225 ± 7
				123.5		2.2	225 ± 7
RP 79-T (high sanidine with authig. feldsp. rims)	0.100–0.200	47	9.85	96.5		1.8	236 ± 7
				96.9		1.9	237 ± 8
loc. Mirigioli (716.512/085.537)							
SG 85-G (high sanidine)	0.200–0.315	10			72.951	4.4	235 ± 15
		11			78.966	6.4	253 ± 11
SG 85-W (authigenic K-feldspar)	0.200–0.315	27	13.24	121.9		2.1	222 ± 3
				122.6		2.3	224 ± 3
SG 71-M (feldspar mixture)	0.100–0.200	65	10.34	95.7		1.4	224 ± 7
				97.0		1.4	227 ± 7
	0.100–0.200	40	10.98	102.5		1.6	225 ± 7
				103.1		1.7	227 ± 7
						Mean age:	226 ± 2 (± 5)
SG 66-W (authigenic K-feldspar)	0.200–0.315	40	13.00	124.1		1.9	230 ± 7
				125.0		2.1	232 ± 7
	0.200–0.315	38	13.00	125.1		0.8	232 ± 7
				124.6		0.9	231 ± 7
	0.100–0.200	45	12.97	121.4		1.1	226 ± 7
				122.4		1.2	228 ± 7
						Mean age:	230 ± 3 (± 5)

The potassium concentrations of the alkali feldspar types of this study are listed in Tables 3 and 4. The data are mean values of duplicate determinations. The concentrations of the three feldspar types W, G and T differ significantly. They are in agreement with the concentrations which can be calculated from the structural properties (section above). Feldspar type W has the highest K content (about 13%), type G 9.5% and type T gave values in between.

Additional measurements have been made by electron micro-

probe scans. Representative grains of feldspar types G and T were analysed. Figure 7 shows the chemical profiles for the elements K, Na and Ca in the crystals. The water-clear and homogeneous feldspar type G can be distinguished by a homogeneous distribution of the three elements across the crystals, whereas the cloudy feldspar type T exhibits concentration changes towards the rims of the crystal. This suggests that the outer part of this feldspar has a different chemical constitution. The K values increase while the Na and Ca concentrations decrease.

SAMPLE	LOCALITY	Type, grain- ϕ	POTASSIUM (%)					AGE (Ma)							
			9,0	10,0	11,0	12,0	13,0	PERMIAN	240	230	220	210	200	JUR-ASSIC	190
VS	G	200/315													
SG 85	T	100/200													
	T	200/315													
SG 79	M	100/200													
	M	200/315													
	M	315/500													
	G	100/200													
	G	200/315													
	G	315/500													
	T	100/200													
	T	200/315													
	T	315/500													
	W	100/200													
	W	200/315													
	W	315/500													
	M	200/315													
	M	200/315													
SG 71	M	100/200													
SG 66	W	100/200													
	W	200/315													
RP	W	200/315													
	T	100/200													
VP	G	500/1000													
	G	500/1000													
	T	500/1000													

SG = San Giorgio (Mirigioli) ; VP = Val Porina ; VS = Val Serrata ; RP = Rossago/Pogliana
 SAMPLE WEIGHTS 40/50 mg; Ar(atm) - 2 %
 * $^{40}\text{Ar}/^{39}\text{Ar}$ ages (total fusion); neutron dose - $5 \cdot 10^{17} \text{ n/cm}^2$

Fig. 8. Potassium contents and K-Ar ages (conventional and $^{39}\text{Ar}/^{40}\text{Ar}$ total fusion data) of alkali feldspars from different bentonite beds of the Mirigioli outcrop (SG 66, 71, 79, 85) and some other localities (VP, Val Porina, corresponding to SG 85; RP, Rossago-Pogliana, to SG 79; VS, Val Serrata, is already stratigraphically Lower Ladinian)

The intermediate K values of the type T as determined by flame photometry are therefore interpreted as mean values between the concentrations of the core and the rim.

From the Table 3 it can be inferred that the feldspars G and T show systematic K concentration differences according to grain size. Type T has higher values for smaller grain sizes and type G lower values. Both effects are due to admixtures of feldspar with higher K content. In the case of type G the larger grain sizes are richer in this component than the smaller ones; in the case of type T the K-content increases with the decreasing core to rim ratio.

Neutron Activation

Feldspar samples of around 40–50 mg meant for stepwise degassing and samples of around 30 mg for total fusion were neutron irradiated in the reactor of the Nuclear Research Center, Karlsruhe-Linkenheim, West Germany. Sodium-sanidine of the arfvedsonite trachyte from Hohenburg, Siebengebirge (Vieten 1965; Frechen 1976, p14) was used as the irradiation standard. This also serves as the laboratory standard in our dating projects on Tertiary volcanics. The integrated fast neutron dose averaged 5×10^{17} neutrons/cm². For irradiation, samples and standards were wrapped in aluminium foil and enclosed in quartz-glass vials. ^{37}Ar was not measured because the low Ca/K ratios of about 1/1000 did not bring about measurable concentrations after radioactive cooling. As standard for the $^{40}\text{Ar}/^{39}\text{Ar}$ total fusion experiments, we chose sample G (200–315 μm) from bed no. 79 of the Mirigioli outcrop.

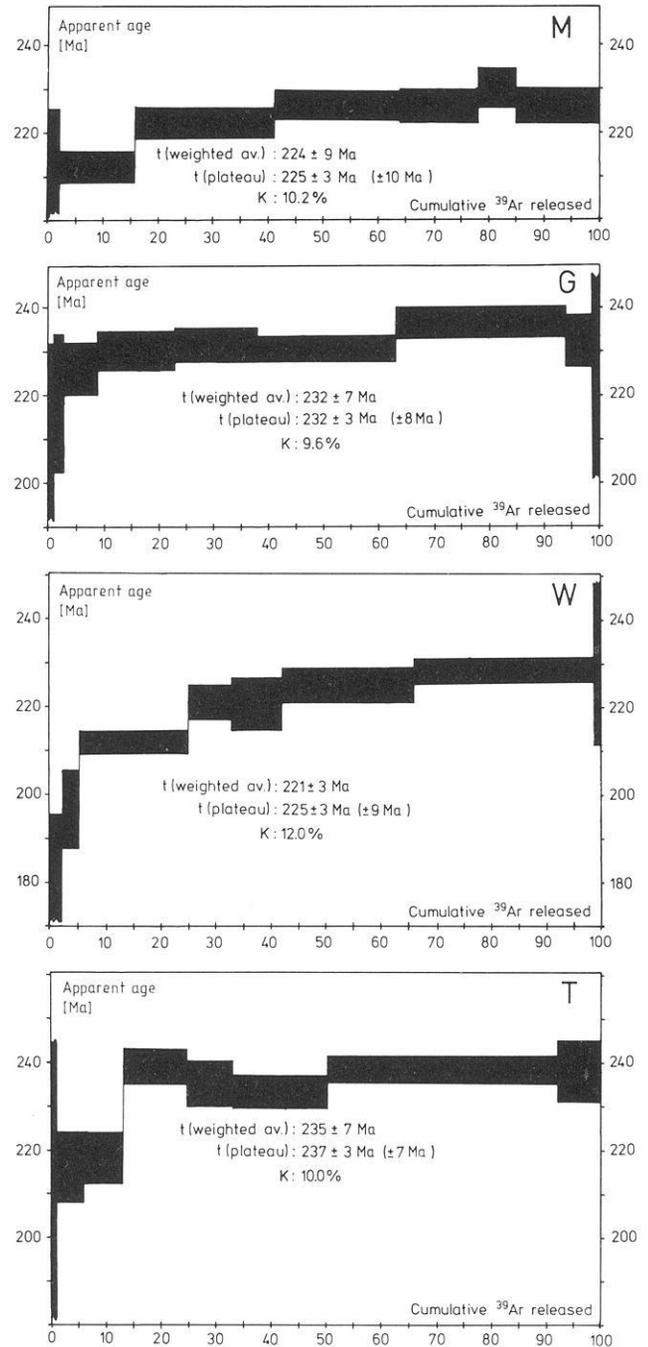


Fig. 9. $^{39}\text{Ar}/^{40}\text{Ar}$ stepwise degassing plots of three alkali feldspar types found in bentonite SG 79 of the "Grenzbitumenzone" at Monte San Giorgio and of the feldspar mixture (M). W feldspar with triclinic domains; T high sanidine with overgrowth of W, G clear high sanidine. Sanidine of the alkali trachyte Hohenburg (Siebengebirge) was used as standard

Argon Analysis

The $^{40}\text{Ar}/^{40}\text{K}$ measurements were performed between January 1976 and April 1977 (experiment numbers 2330–2885). The $^{40}\text{Ar}/^{39}\text{Ar}$ analyses were made in July 1976 (step experiments 2596–2601) and May 1977 (total fusion experiments 2906–2937).

The argon isotopes ^{36}Ar , ^{38}Ar , ^{39}Ar and ^{40}Ar were measured on a Varian MAT GD-150 mass spectrometer (180°, 5 cm radius) in static mode, with the acceleration voltage being changed for mass scanning. The total argon blank averaged about

Table 5. Summary of $^{39}\text{Ar}/^{40}\text{Ar}$ results (step experiments) for feldspar samples W, G, T and M from outcrop Mirigioli (SG 79). Standard: Laboratory standard sanidine Hohenburg, conventional age: 24.9 ± 0.4 m.y.

Sample	Weight (mg)	Run-No.	Steps	^{40}Ar (atm) (%)	Potassium (a) flame phot. (b) ^{39}Ar (%)	(40)/(39) Sample	(40)/(39) Standard	Averaged (A) and Plateau (P) ages (m.y.)
M	39.6	2601	9	5.96	10.17 (12.9)	57.20	6.01 ± 0.24	A 223.6 ± 9.0 P 225 ± 3 ± 10
G	40.7	2600	9	3.10	9.56 (10.6)	55.06	5.57 ± 0.15	A 231.6 ± 7.0 P 232 ± 3 ± 8
W	38.9	2596	10	4.81	12.99 (11.8)	50.52	5.37 ± 0.18	A 221.0 ± 8.0 P 225 ± 3 ± 9
T	44.6	2598	9	4.19	10.02 (8.8)	48.13	4.80 ± 0.11	A 235.0 ± 7.0 P 237 ± 3 ± 7

$1 \times 10^{-8} \text{ cm}^3$ for the low temperature runs and about $2 \times 10^{-8} \text{ cm}^3$ for the high temperature and fusion runs. The samples were heated and fused by induction heating in molybdenum crucibles. Stepwise heating was started at around 400°C , using nine to ten steps with progressively higher temperatures. The argon was released essentially at around $1,500^\circ\text{C}$. Prior to the measurements the samples for the conventional isotope dilution analysis (30–50 mg) were preheated at temperatures below 250°C for a duration of about 12 h. The samples for stepwise heating experiments ($^{40}\text{Ar}/^{39}\text{Ar}$) were preheated at temperatures below 180°C for 24 h. The extracted gases were cleaned by zirconium and titanium getters. The measurements were checked continuously by blank and second degassing runs. Mass spectrometer discrimination was monitored by analysing atmospheric argon quantities comparable to the sample argon.

Decay Constants and Calibrations

The ages were calculated on the basis of the IUGS convention decay constants (Steiger and Jäger 1977). The argon measurements for the K–Ar age determinations were performed in series of eight to ten fusion experiments. Within each series aliquots of the USGS standard biotite LP-6 were measured (Ingamells and Engels 1976; bottles 7-II-C8 and 7-III-D6). During the time devoted to this study forty analyses on this biotite were made. The atmospheric argon contributions were about 5% when using about 50 mg of sample. The standard deviation of the LP-6 measurements was 0.072, or 1.6%. All argon measurements were normalized to an average value of $43.16 \times 10^{-6} \text{ cm}^3$ NTP for LP-6. This argon content corresponds (together with our K value for LP-6 of 8.37%) to an age of 127.8 m.y. Ingamells and Engels (1976) compiled the results of a dozen K–Ar researchers and recommended $43.26 \times 10^{-6} \text{ cm}^3$ as radiogenic argon and 8.325% as potassium concentration (corresponding to an age of 128.9 m.y.). A recent $^{40}\text{Ar}/^{39}\text{Ar}$ comparison in our laboratory of the micas P-207 and LP-6 (Lippolt et al. 1981) gave 128.6 m.y. for the latter based on 82.6 m.y. for P-207. The degassing spectrum for LP-6 turned out flat with the exception of the first step which showed a higher $^{40}\text{Ar}/^{39}\text{Ar}$ value than the plateau, similar to the results of Ozima et al. (1979) for LP-6.

Based on the individual errors of the potassium and argon analyses the conventionally determined K–Ar ages in Tables 3

and 4 (measured argon volumes of about $5 \times 10^{-6} \text{ cm}^3$, some percent atmospheric Ar contribution) have standard deviations of about 2% for measurements performed twice and about 3% for single measurements.

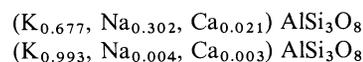
The $^{40}\text{Ar}/^{39}\text{Ar}$ total fusion measurements (Table 4) are based on the K–Ar age of one of the feldspar samples of this study (type G of bentonite 79). This means that they are linked to the calibration described above.

The $^{40}\text{Ar}/^{39}\text{Ar}$ stepwise heating experiments are calibrated with the K–Ar age of a sanidine from the Siebengebirge (locality Hohenburg). Vieten (1965) and Frechen (1976) give geological and petrographical details. For this sanidine, which for a long time served as our laboratory standard, we used a K–Ar age of 24.9 ± 0.9 m.y. This value is our best estimate for the age of this trachyte, though the long time average is 25.0 m.y. It can therefore not be excluded that the $^{39}\text{Ar}/^{40}\text{Ar}$ step ages should be higher by 0.4% than given in Table 5 and Fig. 9. The Hohenburg sanidine itself exhibits the degassing pattern of an undisturbed mineral as shown by H.J. Jensen in our laboratory (unpublished). Incomplete melting would therefore not result in severe dating errors. The errors given in Table 5 are standard deviations of the total-argon ages (A) and of the plateau steps (P). As a third value an error estimate for the plateau ages is presented which also comprises the errors of standard calibrations and standard measurements.

Discussion of the Experimental Results

Origin of the Alkali Feldspar Types

The chemical composition of the feldspar phenocrysts from the Monte San Giorgio bentonite based on electron probe measurements can be expressed by the following formulas:



The first one stands for the type G feldspar and for the core of the cloudy feldspar type T; the latter for the type W feldspar and the rim of type T. The cloudiness of the type T feldspar seems to be due to overgrowth by another feldspar not present in the core. At first we thought we had to distinguish between a sharp-edged and a round-edged cloudy feldspar type, but later on we discovered that the rounding of the feldspar grains might

have been caused by the ultrasonic treatment which partly cleans the T-type feldspar grains from the overgrowth.

The chemical and crystallographic results suggest the existence of two different feldspar phases which originated from two different sources:

a) a high temperature alkali feldspar with a high sanidine structure (G);

b) a low temperature potassium feldspar with poor crystallographic continuity (W).

We have good reason to believe that the high temperature feldspar is of volcanic and the low temperature feldspar is of authigenic origin.

The Authigenic Feldspar: The feldspar considered authigenic has the composition: $Or_{96-98}Ab_{4-2}$ and occurs in all the investigated bentonite beds. The adularia-like habit, the cell dimensions, the chemical composition and the weak Al, Si order are typical for authigenic feldspar. It appears as uniform crystals and as overgrowth on the type G feldspar. According to Baskin (1956) crystallization of potassium feldspar with nearly a stoichiometric composition takes place at temperatures below 100° C. As potassium source we can assume devitrification of rhyolitic glass in a H₂O solvent.

Althaus and Wirsching (1979) showed that the alteration of rhyolitic glass under hydrothermal conditions of 250° C leads to the formation of K feldspar and montmorillonite. The latter mineral can also be formed by alteration of plagioclase. Montmorillonite is a major component of the bentonites. The absence of plagioclase need not be a primary feature, but might be due to this alteration process. Halmyrolytic alteration (devitrification) of submarine ash beds is well known as a potassium source for the formation of new minerals (Füchtbauer 1950; Heim 1960). The authigenic feldspar must be younger than the formation of the bentonite bed. The isotopic age of the feldspar type W therefore could serve only as an age limit for the formation of the Grenzbitumenzone at Monte San Giorgio. Provided that this feldspar type is a retentive K–Ar mineral the isotopic age could give us information on the time of alteration of the sediments. It cannot be excluded that the alteration occurred almost contemporaneously with the deposition.

The Volcanic Feldspar: The feldspar regarded as volcanic has the composition: $Or_{68-69}Ab_{32-31}$ and its occurrence is restricted to the bentonite beds 71, 79 and 85 of our sample reservoir from the GBZ. The water-clear and homogeneous appearance, the high Al, Si-disorder, the cell parameters, and the chemical homogeneity are typical features of crystallization at high temperatures (above 800° C) and of rapid quenching (Ribbe 1975). Obviously there is no reason to doubt the volcanic origin of this feldspar type. It occurs as uniform crystals in type G and as core in the feldspar type T. As the rims of the cloudy feldspar crystals (type T) give reflections different from those of the core, the data points of the water-clear feldspar type G and VP (Fig. 6) from the same bentonite horizon at two different localities and the cloudy type T and R from the locality SG (Fig. 2) cluster close to the high sanidine corner. Whereas the rims of the type T feldspar overlap with the data points of sample W.

The Total Fusion Ages

The total fusion ages of the volcanogenic and authigenic feldspars are compiled in Tables 3 and 4. To facilitate a comparison, figure 8 presents graphically the relations between the samples. Table 3 contains the result obtained from feldspars of bed no. 79

of the Mirigioli outcrop, arranged according to type and grain size. The feldspar data of all other beds of this outcrop and from other localities are given in Table 4 (K–Ar and $^{40}Ar/^{39}Ar$ results).

Since the bentonite from bed no. 79 contains all different feldspar types mentioned above in sufficient concentrations, the argon retention behaviour of the various feldspar types can easily be compared. The ages of the water-clear high sanidine (G: 232 ± 9 m.y. (2 s.d.), 15 analyses) and those of the high-sanidine with overgrowth (T: 235 ± 6 m.y. (2 s.d.), 8 analyses) are the same within the errors, even within 1 s.d. errors. These averages do not include three analyses of the water-clear feldspar which yielded distinctly lower (218 m.y., 219 m.y.) or higher (253 m.y.) ages, though by means of thin section studies no evidence of inhomogeneity or impurities was found. The authigenic feldspar on the average (W: 226 ± 8 m.y. (2 s.d.); 18 analyses) yielded an age, which, even on the 1 s.d. basis, is not significantly lower than the mean age of the volcanic feldspar. For this reason also feldspar mixtures (M: 229 ± 5 m.y. (2 s.d.); 16 analyses) with may be 30% W-type feldspar give ages close to those of high sanidine. The results of the cloudy sanidine type (T) indicate that the ages of the high sanidine cores are not seriously affected by the authigenic overgrowths.

The consistency of the age data is satisfactory not only for aliquots of the same sample but also for sample splits with different grain sizes. One may infer from this that there are no crystallographic differences within the spectrum of grain sizes which could have produced a low-retentivity feldspar.

Additionally the data for the sanidines and alkali feldspars from the other bentonite beds of the Mirigioli outcrop (66, 71, 85) and from the corresponding beds of other localities (VP, RP) are consistent with the results of the SG 79 bentonite. The analytical results in Table 4 do not reveal any vertical effect. The ages of sub- and toplayers within the GBZ are concordant within the analytical errors. This is not surprising since the bentonites concerned are not more than 2 m apart. Obviously the sedimentation interval documented by this sequence cannot be resolved by isotopic dating at present. Based on the lifespan of some *Daonella* species, the accumulation of the 4 m thick rock column could have taken only 1–4 m.y. (Rieber 1969). The Lower Ladinian crystal tuff from Val Serrata, which lies 110 m above the GBZ stratigraphically, yielded a high-sanidine age of 225 ± 4 m.y. From this, for the sedimentation rate, an order-of-magnitude calculation suggests a value of 10 m/m.y. From the data of Tables 3 and 4 an estimate of the interval between formation of the volcanic feldspar and the growth of the authigenic feldspar can be inferred. It is probably 23 m.y. at the most and much narrower than in the example described by Hunziker (1979).

The $^{40}Ar/^{39}Ar$ Stepwise Heating Ages

The results of our $^{40}Ar/^{39}Ar$ stepwise heating experiments on the samples SG 79 G, T, W and M are listed in Table 5. In Fig. 9 the individual apparent ages are plotted against the cumulative percentage of released ^{39}Ar . The $^{40}Ar/^{39}Ar$ total argon ages (weighted by means of steps) agree fairly well with the total fusion $^{40}Ar/^{40}K$ ages discussed in the last paragraph. The two differently determined ages (K–Ar; $^{40}Ar/^{39}Ar$) are (2 s.d. errors):

Samples G	232 ± 9 m.y.,	232 ± 14 m.y.
Samples T	235 ± 6 m.y.,	235 ± 14 m.y.
Samples W	226 ± 8 m.y.,	221 ± 16 m.y.
Samples M	229 ± 5 m.y.,	224 ± 18 m.y.

The mean ages of the samples W and M deviate by 4–5 m.y. Again the authigenic feldspar (W) turns out to be younger than the volcanic feldspar represented by samples G and T. All the degassing spectra show a well-developed plateau, but they differ in the first degassing steps.

Sample G behaves as an undisturbed sample. The plateau value is reached after the first two steps which amount to only 3% of the argon released. This behaviour is in accordance with the high argon retentivity of high sanidine.

Sample W, representing the authigenic feldspar, yields a curve which indicates argon loss. After the first three degassing steps (about 23% of the argon) a value of 212 m.y. is reached. From there on it increases slowly to a very narrow plateau value of 225 m.y. (35% of the argon). This plateau value agrees well with the total fusion age. The degassing behaviour of sample W reflects the lower argon retentivity of feldspar with crystal inhomogeneity.

Sample T, sanidine with overgrowth, reaches the plateau value after a 15% ^{39}Ar release. Between the third and the fourth step the $^{40}\text{Ar}/^{39}\text{Ar}$ ratio increases rapidly, indicating that the core now contributes noticeably to the argon released.

Sample M, which consists of a mixture of the three feldspar types G, W and T (40%, 20%, 40%), reaches the plateau after 40% ^{39}Ar release, but this value is smaller than expected. Actually, the sample may have contained up to 35% weight of feldspar type W, taking into account that the rims of the feldspar grains T are made up of authigenic overgrowth. This does not explain, however, why the plateau value of feldspar types G and T were not reached. It may result from irradiation inhomogeneity or from an underestimation of the type W feldspar component in the mixture. This result may emphasize the importance of using very pure mineral separates, since obviously the individual apparent ages are mineral ages between those of feldspar type G and of feldspar type W.

The results of the $^{40}\text{Ar}/^{39}\text{Ar}$ step experiments in a $^{40}\text{Ar}/^{36}\text{Ar}$ versus $^{39}\text{Ar}/^{36}\text{Ar}$ diagram define straight lines within the analytical errors; the maximum $^{40}\text{Ar}/^{36}\text{Ar}$ coordinates being $3 \times 10^4 - 5 \times 10^4$. Ages inferred from the isochrons agree well with the averages of the steps, either on the basis of all steps or of the plateau steps alone. The intercepts are 290 ± 90 (G), 223 ± 140 (T), 190 ± 192 (W) and 157 ± 190 (M). The relatively high standard deviations of the intercepts are due to the fact that the important steps are very radiogenic and therefore have very high coordinates. The combined plateau – and isochron evaluation shows clearly that the samples are undisturbed in the sense of the four criteria formulated by Lanphere & Dalrymple (1974). The losses of radiogenic ^{40}Ar prior to the analysis amount to about 2% at the most (sample W).

Conclusions

As has been shown, the bentonites intercalated in the Grenzbitumen zone are of volcanic origin and contain high sanidine and authigenic K-feldspar, which are both suitable for K–Ar dating. The total fusion ages of 26 samples (pure sanidine and volcanic feldspar cores) average at 232 ± 9 m.y. (2 s.d.). The authigenic feldspar samples over-all yield an age of 226 ± 8 m.y. (2 s.d.). The combined results of the total fusion (conventionally and $^{39}\text{Ar}/^{40}\text{Ar}$) and the step degassing experiments are 233 ± 7 m.y. (2 s.d.) for the volcanic feldspar samples and 226 ± 7 m.y. (2 s.d.) for the authigenic feldspar samples. For these averages the plateau ages have been used. When taking into account possible calibration errors the standard deviations increase by 2 m.y. to ± 9 m.y.

These results and the well-defined stratigraphic position of

the GBZ justify the proposition that this age be used as the new calibration point for the Anisian-Ladinian boundary of the Triassic time scale:

The ages of the high sanidine, however, are about 20 m.y. higher than one would expect on the basis of the Phanerozoic time scale by Harland and Francis (1971). To a certain degree the difference may be due to the new decay constants (Steiger and Jäger 1977), which produce an increase of about 5 m.y. The remaining 15 m.y. are a real discordance and suggest that the Triassic time scale has to be recalibrated.

Furthermore it is evident that the formation of authigenic feldspar can be dated. Because of the lower argon retentivity of this feldspar, the $^{40}\text{Ar}/^{39}\text{Ar}$ technique is considered a better way to obtain significant results.

Even sanidine with authigenic overgrowth may be of use if the rim/core ratio is favorable. With certain caution the ages obtained for the authigenic feldspar can be considered the age of diagenesis of the Monte San Giorgio rock sequence, especially of the alteration of the tuffs into bentonites. This event took place about 10 m.y. after eruption and deposition of the volcanic material.

As a final remark it should be mentioned that Ferrara and Innocenti (1974) have furnished proof of a thermal event in the Southern Alps which has obviously affected biotite ages of Hercynian basement rocks. This event is thought to be synchronous with the Triassic volcanic phase in this area. As the rocks of our study are sediments and at that time were certainly in a shallow position it is not likely that their volcanic components have been influenced by this event.

Consequences for the Triassic Time Scale

From the well-defined stratigraphic position of the “Grenzbitumenzone” and their bentonites and from the chronometric results of this study we may infer that the Anisian-Ladinian boundary is best dated at 232 m.y. This age is in conflict with the estimates of the currently accepted Triassic time scale according to which the Permian-Triassic boundary would lie close to this age (230 m.y.) and the Anisian-Ladinian boundary at about 210 m.y. These ages are based mainly on K–Ar mineral ages of magmatic rocks from Australia which scatter between 220 and 240 m.y. (Webb and McDougall 1967; Evernden and Richards 1962). Unfortunately the stratigraphic record of these rocks is not really significant and therefore the age data should not be used for time scale calibration. A close look at the more recent publications on the geochronology of Triassic rocks reveals that in fact our result are not in conflict with other information on this subject. Figure 10 summarizes graphically what is known from fifteen papers about isotopic and stratigraphic ages of dated Triassic rocks. The straight line on the left side represents the scale proposed in 1964 (Francis and Harland 1964). It is quite evident that most of the data point to significantly higher values for the Triassic period. The straight line on the right in Fig. 10 represents a calibration line based on more recent data. A more or less similar conclusion has already been drawn by Armstrong and McDowall (1974) and Armstrong (1978). They proposed a value of 212 m.y. for the Triassic-Jurassic boundary and a value of ≥ 247 and 248 respectively for the Permian-Triassic boundary. Webb (1981), reexamining the available data on the Triassic system, set the boundaries at 200 ± 5 and 245 ± 5 m.y. The calibration point we present here lies right in the straight line on the right side of Fig. 10. Our result is a further proof that the Triassic is confined to an older time

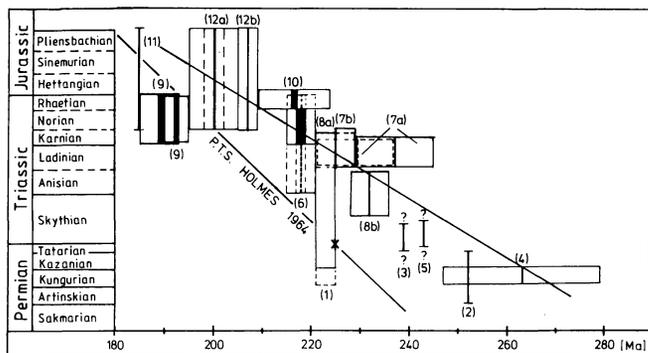


Fig. 10. Relation of isotopic age determinations of Triassic rocks to the pertinent stratigraphic levels. K-decay constants according to Wetherill (1966), Rb-decay constants as proposed by Steiger and Jäger (1977). The K–Ar based rectangles move 2.3% to the right when the IUGS constants are used.

The numbers refer to: 1, 2 Evernden and Richards, 1962, 3 Webb and McDougall, 1967, 4 Kaemmel et al. 1970, 5 Rosenkrantz et al., 1964, 6 Webb and McDougall, 1967, 7a Borsi and Ferrara, 1967, Borsi et al., 1968, 7b Borsi et al., 1968, 8 Valencio et al., 1975, 9 Erickson and Kulp, 1961, Fanale and Kulp, 1962, Armstrong and Besancon, 1970, 10 Armstrong and Besancon, 1970, 11 Baadsgaard et al., 1961, 12a White et al., 1967, 12b Christmas et al., 1969. Most ages are K–Ar ages; only 6 (black box), 7a and 12b are Rb–Sr results

interval than adopted in 1964, probably between 250 and 210 m.y.

In an earlier discussion Armstrong and Besancon (1970) had assumed that the Triassic period started between 255 and 250 m.y. This estimate was based on an extrapolation using maximum sediment thicknesses of the Triassic. Probably a more reliable extrapolation can be made using the range of ammonoid zones, as the rate of phylogenetic variability is much more sensitive and more time constant than sediment thicknesses. The Anisian-Ladinian boundary divides the 34 ammonoid zones of the Triassic into 16 earlier and 18 later zones. The data of White et al. (1967), Christmas et al. (1969), Armstrong and Besancon (1970) for uppermost Triassic rocks (205–210 m.y.) and our data allow the calculation of a mean length for a zone in this period (1.2–1.5 m.y.). If we apply the same rate to the sixteen earlier zones, the beginning of the Triassic lies close to 250 m.y.

Acknowledgements. This study was supported by several grants from the Deutsche Forschungsgemeinschaft (project name "Paläovulkanite") for which we are deeply grateful. We are further indebted to the Gesellschaft für Kernforschung, Isotopenstelle, in Karlsruhe-Linkenhein for performing the neutron irradiations. Mr. H. Funke and A. Lutz assisted us with the age determinations, U. Fuhrmann, J. Hess, K. Limbach and M. Hanel with calculations, drawings, etc. This we gratefully acknowledge.

Professor Berdesinski and his co-workers and Dr. B. Nuber helped us to perform the X-ray diffraction work. The electron microprobe analyses were made by Dr. V. Stähle. Professor Kuhn-Schnyder and Professor Rieber (Zürich, Switzerland) gave us considerable advice concerning the Monte San Giorgio paleontological situation. To all of them we would like to express our sincere appreciation.

References

Althaus, E., Wirsching, U.: Experimente zur Bildung von Feldspäten bei der hydrothermalen Umwandlung von Gesteinsgläsern. Fortschr. Mineral. **57**, Beih. 1, I–II, 3–4, 1979
 Armstrong, R.L.: Precenozoic phanerozoic time scale – Computer File of critical dates and consequences of new and in Progress Decayconstant revisions. In: 'Contributions to the Geological Time

scale', Studies in Geology, No. 6 G.V. Cohee, M.F. Glaessner, H.D. Hedberg, eds.: pp 73–91. Tulsa, Oklahoma, USA: AAPG 1978
 Armstrong, R.L., Besancon, J.: A Triassic time scale dilemma: K–Ar dating of Upper Triassic mafic igneous rocks, eastern USA and Canada and post-Upper Triassic plutons, western Idaho, USA. *Eclogae Geol. Helv.* **63/1**, 15–28, 1970
 Armstrong, R.L., McDowell, W.G.: Proposed refinement of the Phanerozoic time scale. Int. Meeting for Geochron., Cosmochron. and Isotope Geol., Paris, 1974
 Assereto, R.: Sul ritrovamento di Cefalopodi anisici nella Val Romana (Alpi Giulie occidentali). *Riv. Ital. Paleontol.* **72**, 3, 591–606, 1966
 Assereto, R.: Sul significato stratigrafico della "Zona ad Avisianus" del Trias Medio della Alpi. *Boll. Soc. Geol. Ital.* **88**, 123–145, 1969
 Assereto, R.: Aegean and Bithynian: Proposal for two new Anisian substages. *Schriftenr. Erdwiss. Kommiss. Österr. Akad. Wiss.* **2**, 23–39, 1974
 Assereto, R., Casati, P.: Revisione della stratigrafica permotriassica della Val Camonica meridionale (Lombardia) - *Riv. Ital. Paleontol.* **71**, 4, 999–1077, 1965
 Assereto, R., Desio, A., Di Colbertaldo, D., Passeri, L.D.: Note illustrative della Carta geologica d'Italia alla scala 1:100.000, F.^a 14^A Tarvisio Serv. Geol. d'Italia, Florenz 1968
 ASTM: Spec. Techn. Papers on Powder diffraction standards, compiled by JCPDS, American Soc. Testing materials, Swarthmore, Pennsylvania
 Baadsgaard, H., Dodson, M.H.: Potassium-Argon dates of sedimentary and pyroclastic rocks. *Quart. J. Geol. Soc. Lond.* **120 S**, 119–127, 1964
 Baadsgaard, H., Folinsbee, R.E., Lipson, I.: Potassium-Argon dates of biotites from Cordilleran granites. *Bull. Geol. Soc. Am.* **72**, 689–702, 1961
 Baccele, L., Sacerdoti, M.: Una serie di Strati di Livinallongo (Ladino inferiore) nei pressi di Caprile (Dolomiti Bellunesi). *Studi Trent., Sci. Nat. ser. A.* **42**, 2, 113–162, 1965
 Baskin, Y.: A study of authigenic feldspars *J. Geol.* **64**, 132–155, 1956
 Bechstädt, T., Mostler, H.: Mikrofazies und Mikrofauna mitteltriassischer Beckensedimente der Nördlichen Kalkalpen Tirols. *Geol. Paläontol. Mitt. Innsbruck* **4**, H. 5/6, 1–74, 1974
 Bender, H.: Zur Gliederung der mediterranen Trias II. Die Conodontenchronologie der mediterranen Trias. *Ann. Géol. Pays Helléniques* **19**, 465–540, 1967
 Bender, H.: Der Nachweis von Unter-Anis ("Hydasp") auf der Insel Chios. *Ann. Géol. Pays Helléniques* **19**, 412–467, 1970
 Berdesinski, W., Nuber, B.: Vorschlag zur Bestimmung der Bestwerte der Gitterparameter und ihrer Fehlerschranken aus den beobachteten Glanzwinkeln. *Neues Jahrb. Mineral. Abh.* **104**, 2, 113–146, 1966
 Bernoulli, D. & Peters, T.: Traces of rhyolitic-trachitic volcanism in the upper jurassic of the Southern Alps: Reply. *Eclogae Geol. Helv.* **67/1**, 209–219, 1974
 Bianchi, A., Boni, A., Callegari, E., Casati, P., Cassini, G., Comizzoli, G., Dal Piaz, G.B., Desio, A., Guiseppetti, G., Martina, E., Passeri, L.D., Sassi, F.P., Zanettin, B., Zirpoli, G.: Note illustrative della carta geologica d'Italia alla scala 1:100.000, F.^o 34 Breno, Serv. Geol. d'Italia, Florenz 1971
 Borg, I.Y., Smith, D.K.: Calculated powder patterns. II six potassium feldspars and barium feldspar. *Am. Mineral.* **54**, 163–181, 1969
 Borsi, S. & Ferrara, G.: Determinazione dell'età delle rocce intrusive di Predazzo con i metodi del Rb/Sr e del K/Ar. *Mineral. Petrogr. Acta* **13**, 45–65, Bologna, 1967
 Borsi, S., Paganelli, L. & Simboli, G.: Isotopic age measurements of the M. Monzoni intrusive complex. *Mineral. Petrogr. Acta* **14**, 171–183, Bologna, 1968
 Byström-Asklund, A.M., Baadsgaard, H. and Folinsbee, R.E.: K/Ar age of biotite, sanidine and illite from middle Ordovician bentonites at Kinnekulle, Sweden. *Repr. Geol. Föreningens Förhandlingar* **83**, H. 1, 92–96, Stockholm, 1961
 Callegari, E., De Pieri, R.: Unmixing in the sanidines of the "Pietra

- verde" of the Dolomites (Italy). *Schweiz. Mineral Petrogr. Mitt.* **47**, 1, 111–119, 1967
- Castellarin, A., Pisa, G.: Le vulcaniti ladiniche di Forni di Sopra (Carnia occidentale). *Mem. Mus. Trid. Sci. Nat.* **20**, 99–140, 1974
- Chrismas, L., Baadsgaard, H., Folinsbee, R.E., Fritz, P., Krouse, H.R., Sasaki, A.: Rb/Sr, S and O isotope analyses indicating source and date of contact metasomatic copper deposits, Craigmont, British Columbia, Canada. *Econ. Geol.* **64**, 479–488, 1969
- Cooper, J.A.: The flame photometric determination of potassium in geological materials used for potassium argon dating. *Geochim. Cosmochim. Acta* **27**, 525–546, 1963
- Cornelius, H.P.: Eruptivgesteine in den Werfener Schichten der steierisch-niederösterreichischen Kalkalpen. *Verh. Geol. B.-A.* **10**, 197–202, 1936
- Cornelius, H.P.: Zur magmatischen Tätigkeit in der alpidischen Geosynklinale. *Ber. Reichsst. Bodenforsch.* **1941**, 89–94, 1941
- Cornelius, H.P.: Die Geologie des Schneeberg-Gebietes. *Jahrb. Geol. Bundesanst., Sonderb.* **2**, 115 p., Wien 1951
- Dalrymple, G.B.: Critical tables for conversion of K–Ar ages from old to new constants. *Geology* **7**, 558–560, 1979
- Dalrymple, G.B., Lanphere, M.A.: Potassium-argon dating. Principles, techniques and applications to geochronology. San Francisco: Freeman 1969
- Dalrymple, G.B., Lanphere, M.A.: $^{40}\text{Ar}/^{39}\text{Ar}$ age spectra of some undisturbed terrestrial samples. *Geochim. Cosmochim. Acta* **38**, 5, 715–738, 1974
- De Vecchi, G.P., De Zanche, V., Sedea, R.: Osservazioni preliminari sulle manifestazioni magmatiche triassiche nelle Prealpi Venete (Area di Recoaro-Schio-Posina). *Boll. Soc. Geol. Ital.*, 1974
- Delmonte, M., Paganelli, L., Simboli, G.: The Monzoni intrusive rocks. A modal and chemical study. *Mineral. Petrogr. Acta* **13**, 75–118, 1967
- Dubar, G.: Etudes sur le Lias des Pyrénées françaises. – *Mém. Soc. Géol. Nord* **9**, s. 44ff, 214–216, Lille, 1925
- Erickson, G.P., Kulp, J.L.: Potassium-argon measurements on the Palisades sill, New Jersey. *Bull. Geol. Soc. Am.* **72**, 649–652, 1961
- Evernden, J.F., Curtis, G.H., Kistler, R.W., Obradovich: Argon diffusion in glauconite, microcline, sanidine, leucite and phlogopite. *Am. J. Sci.* **258**, 583–604, 1960
- Evernden, J.F., Richards, J.R.: Potassium-argon ages in Eastern Australia. *J. Geol. Soc. Australia* **9**, pt. 1, 1–50, 1962
- Fanale, F.P., Kulp, J.L.: The Helium method and the age of the Cornwall Pennsylvania magnetite ore. *Econ. Geol.* **57**, 735, 1962
- Faupl, P., Hamedani, A.: Ein Trachyt-Tuffit aus dem Reiflinger Kalk bei Göstling a.d. Ybbs, Niederösterreich. *Mitt. Geol. Ges. Wien* **65** (1972), 109–116, 1972
- Ferrara, G., Innocenti, F.: Radiometric age evidences of a Triassic thermal event in the Southern Alps. *Geol. Rundsch.* **63**, 2, 572–581, Stuttgart, 1974
- Flanagan, F.S.: Reference samples for the earth sciences. *Geochim. Cosmochim. Acta* **38**, 1731–1744, 1974
- Foland, K.A.: Ar^{40} diffusion in homogenous orthoclase and an interpretation of Ar diffusion in K-feldspars. *Geochim. Cosmochim. Acta* **38**, 151–166, 1974a
- Foland, K.A.: Alkali Diffusion in Orthoclase. In: *Geochemical Transport and kinetics* ed. Hofmann, A.W. et al., eds.: pp. 77–98, Carnegie Inst. of Washington, publ. 634, 1974b
- Francis, E.H., Harland, W.B.: The Phanerozoic time-scale. A symposium dedicated to Professor Arthur Holmes. *Quart. J. Geol. Soc. Lond.*, Vol. 120s (Geological Society Phanerozoic time-scale 1964, pp. 260–262), 1964
- Frauenfelder, A.: Beiträge zur Geologie der Tessiner Kalkalpen. *Ecol. Geol. Helv.* **14**, 2, 247–371, 1916
- Frechen, J.: Siebengebirge am Rhein, Laacher Vulkangebiet, Maargebiet der Westeifel. *Bornträger, Berlin & Stuttgart*, pp. 209, 1976
- Frechen, J., Lippolt, H.J.: Kalium-Argon-Daten zum Alter des Laacher Vulkanismus, der Rheinterrassen und der Eiszeiten. *Eiszeitalter und Gegenwart* **16**, 5–30, Öhringen, 1965
- Füchtbauer, H.: Die nicht karbonatischen Bestandteile des Göttinger Muschelkalkes mit besonderer Berücksichtigung der Mineralneubildungen. *Heidelberger Beitr. Mineral. Petrogr.* **2**, 235–254, 1950
- Gallitelli, P., Simboli, G.: Ricerche petrografiche e geochemiche sulle rocce di Predazzo e dei Monzoni (prov. Trento, Italia) svolte nell'Istituto Mineralogica e Petrografica dell'Università di Bologna negli anni 1962–1970. *Mineral. Petrogr. Acta* **16**, 221–238, 1970
- Gentner, W., Kley, W.: Argon-Bestimmungen an Kaliummineralien IV. Die Frage der Argonverluste in Kalifeldspäten und Glimmermineralien. *Geochim. Cosmochim. Acta* **12**, 323–329, 1957
- Gessner, D.: Gliederung der Reiflinger Kalke an der Typlokalität Großreifling an der Enns (Nördl. Kalkalpen). *Z. Geol. Ges.* **116** (1964), 696–708, 1966
- Goldsmith, J.R., Laves, F.: The microcline-sanidine stability relations. *Geochim. Cosmochim. Acta* **5**, 1–19, 1954
- Halliday, A.N., Mitchell, I.G.: Structural, K–Ar and ^{40}Ar – ^{39}Ar age studies of adularia K-feldspars from Lizard complex, England. *Earth Planet. Sci. Lett.* **29**, 227–237, 1976
- Harland, W.B., Francis, E.H.: The Phanerozoic Time-scale. A Supplement. *Geol. Soc. Lond. Spec. Publ.* **5**, 356 S., 1971
- Helgeson, H.C.: Chemical interaction of feldspars and aqueous solutions. In: *The feldspars Mackenzie and Zussmann eds.* pp. 184–217. New York: Manchester Univ. Press, Crane, Russak & Comp., Inc. 1972
- Heim, D.: Über die Petrographie und Genese der Tonsteine aus dem Rotliegenden des Saar-Nahe-Gebietes. *Beitr. Mineral. Petrogr.* **7**, 281–317, 1960
- Höller, H.: Ein vulkanischer Tuff aus den Reiflinger Kalken, E von Großreifling. *Anz. Österr. Akad. Wiss. Math. Natw. Kl.* **100**, 323–324, 1963
- Hummel, K.: Grünerden Südtirols und sonstige halmyrolytische Eisensilikate. *Chemie der Erde. Bd. 6*, 469–551, Jena, 1931
- Hummel, K.: Oberflächennahe Intrusionen und Trümmerlaven in der südalpinen Mitteltrias. *Fortschr. Geol. Paläont.* **11**, H. 33 (Deecke Festschr.), 44–74, Berlin, 1932
- Hunziker, J.C.: Potassium Argon Dating. In: *Lectures in Isotope Geology*, E. Jäger, T.C. Hunziker, eds.: 329 pp. Berlin, Heidelberg, New-York: Springer 1979
- Ingamells, C.O., Engels, J.C.: Preparation, analysis and sampling constants for a biotite. National Bureau of Standards special publications 422: Accuracy in Trace analysis: Sampling, Sample Handling and Analysis. *Proceedings of the 7th IMR Symposium 1976*
- Kaemmel, T., Pilot, J., Rösler, H.J., Schwab, M.: Radiogeochronologische Daten vom Perm der DDR zur Gewinnung von Eichpunkten für die internationale geochronologische Skala. *Z. Angew. Geol.* **16**, 2, 57–63, 1970
- Kahler, F., Kahler, G.: Das Muschelkalkkonglomerat der Südalpen. *Der Karinthin, F. 23*, Klagenfurt, 1953
- Kalbitzer, S., Fechtig, H. In: *Potassium-Argon Dating* (Genter volume), O.A. Schaeffer, J. Zähringer, eds.: Springer Heidelberg 1966
- Kastner, M.: Authigenic feldspars in carbonate rocks. *Am. Mineral.* **56**, 1403–1442, 1971
- Kobel, M.: Lithostratigraphische und sedimentologische Untersuchungen in der kalkalpinen Mitteltrias des Rätikon. *Mitt. Geol. Inst., T.H. Zürich, N.F.* **118**, 149 p., 1969
- Kozur, H.: Probleme der Triasgliederung und Parallelisierung der germanischen und thetyalen Trias, Teil I.: Abgrenzung und Gliederung der Trias. *Freib. Forsch. H.*, **S 298**, 139–197, Leipzig, 1974
- Kuhn-Schnyder, E.: Die Wirbeltierfauna der Trias der Tessiner Kalkalpen. *Geol. Rdsch.* **53/1**, 393–412, 1964
- Kuhn-Schnyder, E.: Die Triasfauna der Tessiner Kalkalpen. *Neujahrsblatt Nat.-Forsch. Ges. Zürich*, 119 S. Zürich: Leemann 1974
- Kuthan, M.: Spuren der vulkanischen Tätigkeit in der mittleren Trias des Slowakischen Karstes. *Geol. Práce, Zoš.* **56**, 55–74, Bratislava, 1959
- Lanphere, M.A., Dalrymple, G.B.: Final compilation of K–Ar and Rb–Sr measurements on P-207, the USGS interlaboratory standard muscovite. In: F.J. Flanagan, ed.: *Descriptions and analyses of eight new USGS rock standards.* *Geol. Surv. Prof. Pap.* **840**, 127–130, 1976
- Lanphere, M.A., Dalrymple, G.B.: The use of $^{40}\text{Ar}/^{39}\text{Ar}$ data in evaluation of disturbed K–Ar systems, in R.E. Zartman (ed.) *Geol. Surv. Open-File report 78–701* (Snowmass colloquium 1978 abstracts), 241–243, Denver, 1978

- Laves, F.: Al/Si – Verteilungen, Phasen – Transformationen und Namen der Akalifeldspäte. *Z. Kristallogr.* **113**, 265–296, 1960
- Lippolt, H.J., Raczek, I., Schleicher, H.: Das Isotopenalter ($^{40}\text{Ar}/^{39}\text{Ar}$; Rb–Sr) eines Autun-Biotits aus der Bohrung Wrzesnia/Polen. Der Aufschluß, im Druck, Jahrgang 1981
- Miller, H.: Die Mitteltrias der Mieminger Berge mit Vergleichen zum westlichen Wetterstein-Gebirge. *Verh. Geol. B.-A.*, 187–212, Wien 1965
- Misch, P.: Der Bau der mittleren Südpirenen. In: Beiträge zur Geologie der westlichen Mediterrangebiete, Nr. 13. *Abh. d. Ges. Wiss. Göttingen, math.-phys. Kl. III. Folge*, H. 12, 1934
- Müller, W., Schmid, R., Vogt, P.: Vulkanogene Lagen aus der Grenzbitumenzone (Mittlere Trias) des Monte San Giorgio in den Tessiner Kalkalpen. *Eclogae Geol. Helv.* **57**, 2, 431–450, 1964
- Mutschlechner, G.: Die Massengesteine der Nordtiroler und Voralberger Kalkalpen. *T. Cherm. Mineral. Petrogr. Mitt.* **4**, 386–395, 1954
- Nopsca, Baron F., Reinhard, M.: Zur Geologie des Vilajets Skutari in Nordalbanien. *An. Ist. Geol. al Românică*, 5, Fasc. 1, 1–24. *Jb. d. K.K. Geol. R.-A.* **61**, 2, 1–24, 1912
- Orville, P.M.: Unit-cell parameters of the microcline-low albite and the sanidine-high albite solid solution series. *Am. Mineral.* **52**, 55–85, 1967
- Ott, E.: Die Kalkalgen-Chronologie der alpinen Mitteltrias in Angleichung an die Ammoniten-Chronologie. *Neues Jahrb. Geol. Palaeontol. Abh.* **141**, 1, 81–115, 1972
- Ozima, M., Kaneoka, I., Yanagisawa, M.: Temperature and Pressure Effects On $^{40}\text{Ar}/^{39}\text{Ar}$ Systematics. *Earth Planet. Sci. Lett.* **42**, 463–472, 1979
- Paganelli, L., Tiburtini, R.: The Predazzo granite (North Italy). *Mineral. Petrogr. Acta* **10**, 57, 1964
- Pilger, A., Schönenberg, R.: Der erste Fund mitteltriadischer Tuffe in den Gailtaler Alpen (Kärnten) *Z. Geol. Ges.* **110**, 205–215, 1958/59
- Plöschinger, B., Wieseneder, H.: Ein Biotitandesit-Tuffit im Reiflinger Kalk des Schwarzkogels bei St. Gallen im Ennstal, O. Österreich. *Verh. Geol. Bundesanst.* 59–69, 1965
- Ribbe, P.H.: The chemistry, structure and nomenclature of feldspars (R-11). In: *Feldspar mineralogy*, P.H. Ribbe ed. *Mineral. Soc. Am., short course notes*, 2, Blacksburg, Virginia: Southern Printing Comp. 1975
- Richter, G., Teichmüller, R.: Die Entwicklung der Keltiberischen Ketten. *Beitr. Geol. westl. Mediterrangebiete* Nr. 9 *Abh. Ges. Wiss. Göttingen, meth.-phys. Kl., III F.*, 7, 1067–1186. Berlin: Weidmann 1933
- Rieber, H.: Daonellen aus der Grenzbitumenzone der mittleren Trias des Monte San Giorgio (Kt. Tessin, Schweiz). *Eclogae Geol. Helv.* **62**, 2, 657–683, 1969
- Rieber, H.: Cephalopoden aus der Grenzbitumenzone (Mittlere Trias) des Monte San Giorgio (Kt. Tessin, Schweiz). *Schweiz. Palaeontol. Abh.* **93**, 1–96, 1973a
- Rieber, H.: Ergebnisse paläontol.-stratigraph. Untersuchungen in der Grenzbitumenzone (Mittl. Trias) des Monte San Giorgio (Kt. Tessin, Schweiz). *Eclogae Geol. Helv.* **66/3**, 667–685, 1973b
- Rosenkranz, A.A., Semenova, T.P., Monich, V.K., Kovaleva, V.V.: The Geochronological data on the upper palaeozoic continental formations in Kazakhstan. *XXII. Int. Geol. Congr. Probl.* **3**, 172–176, 1964
- Rossi, D.: Geologia della parte meridionale del gruppo della Marmolada. *Mem. Mus. Storia nat. Venezia Trident.* **XXV–XXVI**, 1962–1963, vol. XIV, Fasc. 1/B., pp. 189, Trento, 1962
- Sacerdoti, M., Sommariva, E.: Pillowlave, jaloclastiti ed altre formazioni vulcanoclastiche nella Regione Dolomitica Occidentale. *Studi Trent. Sci. Nat.*, a. XXXIX, n. 3, 423–473, Trento, 1962
- Sardarov, S.S.: Retention of radiogenic argon in microcline. *Geochimistry* **3**, 233–237, 1957
- Schmidegg, O.: Die Stellung der Haller Salzlagertstätte im Bau des Karwendelgebirges. *Jahrb. Geol. Bundesanst.* **94** (Festband), 159–205, Wien, 1951
- Silberling, N.J., Tozer, E.T.: Biostratigraphic classification of the marine Triassic in North America. *Geol. Soc. Am. Spec. Pap.* **110**, 1–63, 1968
- Simboli, G.: Ricerche petrochimiche e considerazioni petrologiche sulle formazioni vulcaniche triassiche di Val Gardone (Predazzo). *Mineral. Petrogr. Acta* **12**, 37–60, 1966
- Spadea, P.: Le ignimbriti riolitiche del membro superiore delle vulcaniti di Rio Freddo, nel Trias medio della regione di Tarvisio (Alpi Giulie Occidentali). *Stud. Trent. Sci. Nat., Sez. A* **47**, 2, 287–358, 1970
- Spengler, E., Stiny, J.: Erläuterungen zu Blatt Schneeberg, St. Ägyd 1:75000. *Geol. Bundesanst. Wien*, 1931
- Steiger, R.H., Jäger, E.: Subcommission on Geochronology: Convention on the use of decay constants in geo- and cosmochronology. *Earth Planet. Sci. Lett.* **36**, 359–362, 1977
- Stewart, D.B., Wright, T.L.: Al/Si order and symmetry of natural alkali feldspars and the relationship of strained cell parameters to bulk composition. *Bull. Soc. Fr. Mineral. Cristallogr.* **97**, 356–377, 1974
- Summesberger, H., Wagner, L.: Der Stratotypus des Anis (Trias). *Annalen Naturhist. Mus. Wien* **76**, 515–538, 1972
- Tozer, E.T.: Triassic time and ammonoids: problems and proposals. *Can. J. Earth. Sci.* **8**, 989–1031, 1971
- Vardabasso, S.: Lo stato attuale delle nostre conoscenze sulla provincia petrografica di Predazzo. *Atti. Ist. Veneto di Sci. T. CIV, Parte 2^a, Venezia*, 1944/1945
- Valencio, D.A., Mendia, J.E., Vilas, J.F.: Paleomagnetism and K–Ar ages of Triassic igneous rocks from the Ischigualasto-Ischichuca basin and Puesto Viejo formation, Argentina. *Earth Planet. Sci. Lett.* **26**, 319–330, 1975
- Vidal, H.: Neue Ergebnisse zur Stratigraphie und Tektonik des nord-westlichen Wettersteingebirges und seines nördlichen Vorlandes. *Geologica Bavarica* **17**, 56–88, 1953
- Vieten, K.: Manganreicher Fluotaramit aus dem Alkalitrachyt der Hohenburg bei Berkum (Siebengebirge). *Neues Jahrb. Mineral. Monatsh.* **6**, 166–171, Stuttgart, 1965
- Webb, J.: A radiometric time scale of the Triassic. *J. Geol. Soc. Australia* **28**, 107–121, 1981
- Webb, A.W., McDougall, J.: Isotopic dating evidence on the age of the Upper Permian and Middle Triassic. *Earth Planet. Sci. Lett.* **2**, 483–488, 1967
- Wetherill, G.W.: Radioactive decay constants and energies. In: S. Clark (editor) *Handbook of Physical constants*. *Geol. Soc. Am. Mem.* **97**, 513–519, 1966
- White, W.H., Erickson, G.P., Northcote, K.E., Dirom, G.E., Harakal, J.E.: Isotopic dating of the Guichon batholith, B.C. *Can. J. Earth. Sci.* **4**, 677–690, 1967
- Wirsching, U.: Experimente zur Bildung von Tonmineralien bei der hydrothermalen Umwandlung von Feldspäten. *Fortschr. Mineral.* **54**, Beih. 1, I–II, 104–105, Stuttgart, 1976
- Wirz, A.: Beiträge zur Kenntnis des Ladinikums im Gebiet des Monte San Giorgio. *Schweiz. Pal. Abh.* **65**, 1–84, 1945
- Wright, T.L., Stewart, D.B.: X-ray and optical study of alkali feldspars: I. Determination of composition and structural state from refined unit-cell parameters and 2 V. *Am. Mineral.* **53/1**, 38–87, 1968
- Zanettin, B., De Vecchi, G.P.: Porfiriti triassiche. In: *Note illustrative della Carta geologica d'Italia*, Foglio Schio. *Serv. Geol. Italia*, 1968
- Zapfe, H. (Hersg.): Die Stratigraphie der alpin-mediterranen Trias. *Schriftenr. Erdwiss. Kommiss. Österr. Akad. Wiss.* **2**, 251 p. Wien: Springer 1974
- Zirkel, E.J.: Der Melaphyr von Hallstadt. *Jahrb. Geol. Bundesanst.* **100**, H. 2, 137–178, 1957.

Received January 12, 1981 ; Revised version July 30, 1981

Accepted September 26, 1981

On Modelling the Lithosphere in Mantle Convection with Non-Linear Rheology

H. Schmeling and W.R. Jacoby*

Institut für Meteorologie und Geophysik, Johann Wolfgang Goethe-Universität, Feldbergstr. 47, D-6000 Frankfurt 1, Federal Republic of Germany

Abstract. Numerical convection experiments were carried out with the aim of simulating the lithosphere as a strong mechanical boundary layer participating in the circulation, and to study its dynamical role and the governing parameters. The rheological model parameters were successively refined, effective viscosity depending on (1) depth, (2) temperature and pressure, and (3) temperature, pressure, and stress. In all cases a high-viscosity plate rested on a low-viscosity asthenosphere; in the two latter cases it could in principle subduct, but did so only if zones of weakness were built into it. It was possible to model active or inactive plates (moving faster or slower than the asthenosphere below). Because of a lack of numerical resolution it was however, not possible to simulate a narrow sinking slab; rather a broad zone of cooled and highly viscous material developed, often limiting the rate of descent and leading to non-steady convection. The circulation, including subduction, was stabilized by introduction of stress-dependence of viscosity (non-linearity), dissipation, and adiabatic heating. The parameter chiefly responsible for deciding the (active or passive) role of the plate is its decoupling from its neighbours, achieved in the models by assuming weakness zones. Another important result seems to be that the assumption of plausible mantle rheologies and heat input leads to equally plausible effective viscosities, plate velocities, and to upper-mantle temperatures which are relatively low by current ideas, but conforming to earlier estimates based on convection theory. Viscosity distribution and flow pattern are also in reasonable agreement with more detailed boundary layer computations. The main obstacles to our modelling are the numerical limitations, forcing upon us such artificialities as two-dimensionality, rectangular model boxes, coarse grids, and generalized weakness zones.

Key words: Mantle convection – Finite-difference model – Rheology – Mantle temperatures – Lithosphere

Introduction

The aim of our numerical experiments of mantle convection has been to model the lithosphere and its role in the whole circulation. Most previous studies have stressed either the aspect of thermal convection *or* the mechanical aspect of plate sliding and sinking. The gap between the two ap-

proaches is chiefly the consequence of the deficiencies of the modelling procedures available. Nevertheless, we try with limited means to gain some understanding of earth dynamics.

A model of mantle convection should include the plates to be realistic. They provide reliably known kinematic boundary conditions (rigid rotations typically of about 10^{-16} rad/s relative to each other: Minster et al. 1974; Minster and Jordan 1978). The outline of the plates varies with time and their lateral boundaries migrate in any frame of reference. Furthermore, simple force estimates (Jacoby 1970; Richter and McKenzie 1978) and global force equilibrium models (Forsyth and Uyeda 1975; Solomon et al. 1975; Harper 1975; Chapple and Tullis 1977) demonstrate that the gravitational instability of the plates plays an important role in mantle dynamics.

Little is known about the flow at depth. For continuity some large scale circulation must occur. Concentration of flow is evident in the descending slabs above 700 km depth. Plumes or jets (Morgan 1971; Artyushkov 1968; McKenzie et al. 1980) and small-scale longitudinal rolls (Richter 1973a; Richter and Parsons 1975) shall be disregarded, also such processes as phase changes and chemical differentiation, possibly stabilizing or destabilizing the mantle (Schubert and Turcotte 1971; Richter 1973b; Gebrande 1975; Schubert et al. 1975; Artyushkov 1968).

We assume that the convection is driven by thermally generated density differences. This assumption is convenient as thermal convection has been well studied since Bénard (1900) and Rayleigh (1916); it is also plausible (Tozer 1967a, b), the overcritical Rayleigh number Ra/Ra_c being very large

$$\left(Ra = \frac{g \alpha \Delta T L^3}{\nu \kappa}; \alpha = \text{thermal expansivity}; \Delta T = \text{temperature} \right.$$

drop from bottom to top; L = vertical extent; g = gravity; κ and ν = diffusivities of temperature and momentum (kinematic viscosity); Ra_c = critical Rayleigh number of order 10^3 for the mantle: Jeffreys 1926; Chandrasekhar 1953; 1961).

Even in the simplest case of constant material properties in an infinite horizontal layer heated from below or within and with any conceivable boundary conditions, a Rayleigh number as large as $100 \cdot Ra_c$ implies vigorous convection in rather flat non-steady cells (Busse 1967; Roberts 1967; Krishnamurti 1970a, b; Busse and Whitehead 1971). An important feature of such high-Rayleigh-number convection is the development of pronounced thermal boundary layers which particularly concern us here.

* Present address: Institute of Geophysics and Planetary Physics University of California Los Angeles, Los Angeles, CA 90024, USA

The mantle is, however, more complicated than that. It is not homogeneous; its properties, particularly viscosity, are temperature and pressure dependent. Thermal boundary layers thus also become mechanical ones. Stratified models (including those with a rigid lithosphere) can concentrate the flow at shallow depth (Foster 1969; Takeuchi and Sakata 1970) with large aspect ratios (Gebrande 1975; Houston and DeBremaecker 1975; Richter and Daly 1978; Daly 1980), but they do not exclude convection cells of great depth extent (Davies 1977). A step forward in modelling is to include the lithospheric descent by simply assuming the desired geometrical structure of "rigid" slabs (Richter 1973a; Richter and McKenzie 1978) or high-viscosity regions (Kopitzke 1979) or lateral thermal boundary conditions (Rabinowicz et al. 1980); the flow is then highly organized by the assumed structure. In a more realistic model such a structure should develop by itself as a result of temperature (T), pressure (p), and stress (τ) dependence of rheology. The problem becomes complex and highly non-linear because flow and temperature interact with the mantle properties. Studies of *sub*-lithospheric convection with T, p dependent linear rheology (Torrance and Turcotte 1971; Turcotte et al. 1973; Rabinowicz et al. 1980) and with only τ or with T, p, τ dependent non-linear rheology (Parmentier et al. 1976; Rabinowicz et al. 1980) suggest that the interaction of the material properties with the flow is not too critical. However, inclusion of the lithosphere complicates matters. Houston and DeBremaecker (1975) and DeBremaecker (1977) used T, p or T, p, τ dependent rheologies, while Daly (1980) used T -dependent rheologies, to simulate the lithosphere, but it was either immobile, if too thick and/or too viscous, or it deformed non-rigidly, if too thin and/or too soft.

Approach to Convection with Lithosphere Participating

We have followed a strategy of proceeding from simple stratified models to similar ones with built-in weakness zones, then to models with T, p dependent linear rheologies, and finally to models with non-linear T, p, τ dependent rheologies. These, too, were modified locally by weakness zones necessary to mobilize the lithosphere (see also Kopitzke 1979). In addition, we tested the effects of adiabatic heating/cooling and of dissipation.

The physics of the problem is described by the conservation of mass, i.e., continuity (ρ =density; $\mathbf{v}=(v_1, v_2, v_3)$ = velocity; $\partial_i = \partial/\partial x_i$, x_i =coordinate, $i=1, 2, 3$):

$$\partial \rho / \partial t + \partial_i(\rho v_i) = 0 \quad (1)$$

conservation of momentum - Navier-Stokes equation (D/Dt = material derivative with time t ; p =pressure; σ_{ij} =deviatoric viscous stress tensor; $\mathbf{g}=(0, 0, g)$ =gravity; $\hat{\mathbf{z}}$ =unit vector in x_3 direction):

$$D(\rho v_i)/Dt = -\partial_i p + \partial_j \sigma_{ij} + \rho g \hat{\mathbf{z}} \quad (2)$$

and conservation of energy (T =temperature; c_p =isobaric heat capacity; k =thermal conductivity; \mathcal{H} =volumetric heating; α =thermal expansivity)

$$\rho c_p DT/Dt - \alpha TDp/Dt = \partial_i(k \partial_i T) + \sigma_{ij} \partial_j v_i + \mathcal{H} \quad (3)$$

as well as the constitutive equations to be discussed later.

To obtain the dynamic pressure P , we subtract from p the hydrostatic pressure $\rho g x_3$ (x_3 positive downward); we in-

roduce the equation of state $\rho = \rho_0(1 + p/K - \alpha(T - T_0))$ where K is incompressibility; the viscous stress tensor has the form

$$\sigma_{ij} = \eta((\partial_j v_i + \partial_i v_j) - \frac{2}{3} \delta_{ij} \partial_i v_i) + \frac{1}{3} \eta^* \delta_{ij} \partial_i v_i$$

with η =dynamic viscosity and η^* =volume viscosity, neglected; δ_{ij} =Kronecker delta. We non-dimensionalise the equations following Turcotte et al. (1973).

All mantle properties except viscosity were assumed constant; their values (including reference viscosity ν_0) are given in Table 1. Table 2 lists typical values of the other variables. With these values the non-dimensionalisation allows us to neglect all terms which contribute no more than a few percent. We are left with the following dimensionless equations, with $\theta = T/T_c$ dimensionless, T_c characteristic temperature:

$$\partial_i v_i = 0 \quad (1a)$$

$$\partial_i P - \partial_j \sigma_{ij} - Ra_0 \theta \cdot \hat{\mathbf{z}} = 0 \quad (2a)$$

$$D\theta/Dt + Di \cdot \theta \cdot v_3 = \nabla^2 \theta + Di/Ra_0 \cdot \sigma_{ij} \partial_j v_i + H. \quad (3a)$$

H is the dimensionless internal heating per volume. Dissipation $Di/Ra_0 \cdot \sigma_{ij} \partial_j v_i$ and adiabatic heating $Di \cdot \theta \cdot v_3$ are in-

Table 1. Assumed mantle properties and other model parameters

Symbol	Value	Meaning
η	= variable	dynamic viscosity
ν_0	$= \eta_0/\rho_0 = 3 \cdot 10^{18} \text{ m}^2/\text{s}$	kinematic reference viscosity
K	$= 10^{13} \text{ Pa}$	incompressibility
α	$= 3.7 \cdot 10^{-5} \text{ K}^{-1}$	thermal expansivity
κ	$= 1.01 \cdot 10^{-6} \text{ m}^2/\text{s}$	thermal diffusivity
c_p	$= 10^3 \text{ J/kg} \cdot \text{K}$	isobaric heat capacity
\mathcal{H}	$= 6.65 \cdot 10^{-8} \text{ W/m}^3$	volumetric heating rate
ρ_0	$= 3.5 \cdot 10^3 \text{ kg/m}^3$	reference density
<i>Constants appearing in rheological equations</i>		
R	$= 8.31 \text{ J/K mol}$	gas constant
k	$= 1.38 \cdot 10^{-23} \text{ J/K}$	Boltzmann's constant
f (Eq. 4)	$= 147$ (or 21)	factor for superplasticity (or Nabarro-Herring creep)
<i>Parameters assumed for olivine</i>		
$n; A$	$= 2; 4 \cdot 10^{-3}$ or 3; $3 \cdot 10^{-6}$	Dorn's parameters
b	$= 6.98 \cdot 10^{-10} \text{ m}$	Burgers vector
$D_{v_0}; D_{B_0}$	$= 10^{-1} \text{ m}^2/\text{s}$	diffusivities: lattice; grain boundary
E_v, E_B	$= (5.4; 3.6) \cdot 10^5 \text{ J/mol}$	activation energies (as above)
V_v	$= 1.1 \cdot 10^{-5} \text{ m}^3/\text{mol}$	activation volume
$\Omega(\text{O}^{--})$	$= 1.15 \cdot 10^{-29} \text{ m}^3$	volume of diffusing species (O^{--})
μ	$= \mu_0 \left(1 + \frac{1}{\mu_0} \frac{\partial \mu}{\partial T} (T - T_0) + \frac{1}{\mu_0} \frac{\partial \mu}{\partial p} (p - p_0) \right)$	rigidity
μ_0	$= 7.91 \cdot 10^{10} \text{ Pa}$	rigidity at surface conditions
$\frac{1}{\mu_0} \frac{\partial \mu}{\partial T}$	$= -1.4 \cdot 10^{-4} \text{ K}^{-1}$	relative temperature derivative of rigidity
$\frac{1}{\mu_0} \frac{\partial \mu}{\partial p}$	$= 2.2 \cdot 10^{-11} \text{ Pa}^{-1}$	relative pressure derivative
d	$= 10^{-3}$ to 10^{-1} m	average grain size
δ	$= 1.4 \cdot 10^{-9} \text{ m}$	grain boundary width

Table 2. Typical values of the variables in mantle convection

Variable	Sym- bol	Assumed	
		Dimensional	Non-dimen- sional value
Depth of convection	L	700 km	1
Aspect ratio (assumed)	λ	3 to 4	3 to 4
Time	t	> 1 Ma	> 6 10^{-5}
Velocity	v	≤ 10 cm/a	≤ 2000
Hydrostatic pressure	p_h	$\leq 2.3 \cdot 10^{10}$ Pa	$\leq 3 \cdot 10^6$
Dynamic pressure	P	$\leq 10^8$ Pa	$\leq 10^5$
Shear stress	τ	$\leq 10^8$ Pa	$\leq 10^5$
Temperature	T	≤ 2000 K	≤ 2
Characteristic temperature	T_c	1 000 K	1
Gravity	g	10 m/s ²	3.4 10^{30}

(with constants of Table 1:)

Rayleigh number (scaling)	$Ra_0 = \frac{\alpha g T_c L^3}{\kappa_0 v_0}$	= 123 653
Dissipation number	$Di = \frac{\alpha g L}{c_p}$	= 0.199
Prandtl number	$Pr = \frac{v_0}{\kappa_0}$	= 10^{24}

cluded in some, but not all, models in order to investigate their influence. $Ra_0 = \frac{g \cdot \alpha \cdot T_c \cdot L^3}{v_0 \kappa}$ is the non-dimensionalisation Rayleigh number. To compute the effective Rayleigh number of a particular model, we have used the averaging procedure proposed by Parmentier et al. (1976). The Boussinesq approximation is implied, involving incompressibility in the above equations, but we deviate from this in the adiabatic heating term.

We assume two-dimensionality and introduce the stream function $\psi(x, z)$ with $\mathbf{v} = (\partial\psi/\partial z, -\partial\psi/\partial x)$ satisfying Eq. (1a) and obtain the final form of the Navier-Stokes equation ($v = \eta/(\rho_0 v_0)$, dimensionless kinematic viscosity)

$$4 \frac{\partial^2}{\partial x \partial z} \left(v \frac{\partial^2 \psi}{\partial x \partial z} \right) + \left(\frac{\partial^2}{\partial z^2} - \frac{\partial^2}{\partial x^2} \right) \left(v \cdot \left(\frac{\partial^2}{\partial z^2} - \frac{\partial^2}{\partial x^2} \right) \psi \right) = Ra_0 \frac{\partial \theta}{\partial x} \quad (2b)$$

and the energy equation:

$$\frac{\partial \theta}{\partial t} + \frac{\partial \psi}{\partial z} \frac{\partial \theta}{\partial x} - \frac{\partial \psi}{\partial x} \frac{\partial \theta}{\partial z} - Di \theta \frac{\partial \psi}{\partial x} = \frac{\partial^2 \theta}{\partial x^2} + \frac{\partial^2 \theta}{\partial z^2} + \frac{Di}{Ra_0} \cdot v \cdot \left(4 \left(\frac{\partial^2 \psi}{\partial x \partial z} \right)^2 + \left(\frac{\partial^2 \psi}{\partial x^2} - \frac{\partial^2 \psi}{\partial z^2} \right)^2 \right) + H. \quad (3b)$$

These equations were solved numerically on a 21×21 mesh in square model boxes 700 km deep and $\lambda \cdot 700$ km wide (λ = aspect ratio, variable). Earth curvature was neglected. The boundary conditions were: *sides* closed, free slip, no heat transfer (implying symmetry about the lateral boundaries and periodic repetition of the cells); *top* closed, free slip, temperature 0°C; *bottom* closed, no slip or free slip, heat flow into the box 16 mW/m². Internal heat generation was assumed 6.65 10^{-8} W/m³; the total heat input thus provides for a steady-state average loss of 60 mW/m² (Chapman and Pollack 1975) through the surface. In most cases of temperature

dependent rheology there was little difference between the no-slip and the free-slip condition at the bottom because a hot fluid lower boundary layer developed. The artificial assumptions, such as two-dimensionality, closed square model boxes, coarse grid, were dictated by our computational means and economy. They do present a serious limitation, but they may still help us in uncovering principal systematics in the role of the lithosphere. Even though details may not be quantitatively exact, we believe that our major conclusions are correct. We shall discuss this where appropriate.

Equation (2b) was approximated by a finite difference scheme similar to that of Andrews (1972) and was solved with the Gauss algorithm of elimination (Zienkewicz 1971, p. 462; see also Schmeling 1979; 1980). Equation (3) was solved with the alternating direct implicit (ADI) scheme used by Houston and DeBremaecker (1974). To ensure numerical stability, the time step occasionally had to be adjusted in model runs with T , p , τ dependent rheologies, as regions of low effective viscosity developed. This was done by hand with the aid of an interactive program (Schmeling 1979; 1980), facilitating the compromise between model progress and stability. It also lends credibility to the results.

Accuracy, stability, and convergence of the computations were tested against simple models (homogeneous material, free-slip boundaries, fixed temperature drop ΔT) with known analytical solutions. The results were unaffected by the initial conditions, and within limits, by the time step and grid spacing. The theoretical stability fields, as a function of aspect ratio, were exactly reproduced. Comparisons were run with published, more complex models (Houston and DeBremaecker 1975) and their amplitudes were reproduced to within 10–20%. Flow patterns were reproduced better. Only in cases of extreme rheology do we expect errors of flow amplitude and temperature greater than 20%.

Discussing our results, we shall focus on the comparison with aspects of mantle convection on which we have some information. Such aspects are surface velocities, rigid plate structure, and heat flow. Aspects of special interest include the stability or instability of cellular flow and the mantle temperatures. Observables such as topography, gravity (McKenzie 1977) and lithospheric stress (Richardson et al. 1979) will not be discussed here.

Results

The initial models investigated have a layered viscosity $\eta(z)$. In the first example (Fig. 1) the viscosity-depth distribution is that of McConnell (1968; model 62-12) who based his analysis on data from glacio-isostatic rebound of Precambrian shields. This viscosity is probably appropriate to such regions. It is characterized by a rather thick (120 km) high-viscosity lithosphere overlying an asthenosphere of 10^{17} m²/s kinematic viscosity ($3 \cdot 10^{20}$ Pa s). The viscosities are higher than those estimated from laboratory experiments (AvéLallemant and Carter 1970; Goetze and Brace 1972; Ross and Nielsen 1978; Vetter 1978; Jacoby and Ranalli 1979). Note that we shall use the terms “lithosphere” and “asthenosphere” without quotes even if we refer to nothing but the high and low viscosity layers in our models.

Figure 1 shows, as might have been expected, that the highly viscous and thick lithosphere keeps the circulation from reaching the surface and concentrates it in the softer regions below. The lithosphere is caught motionless between

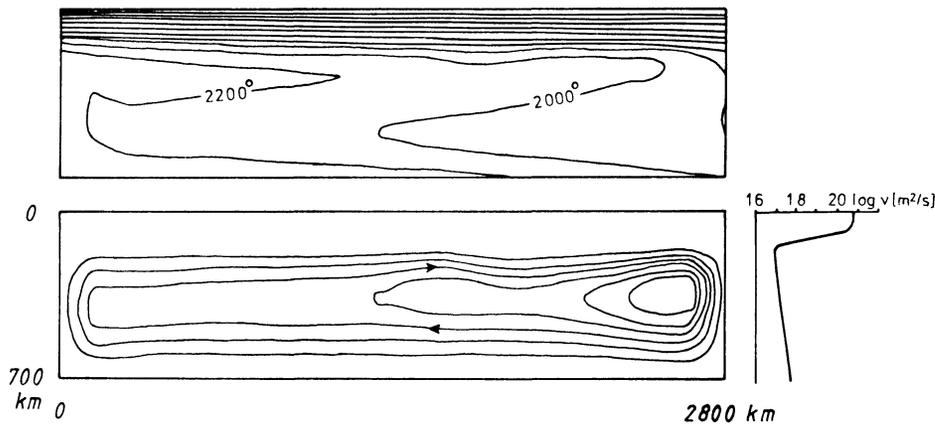


Fig. 1. Convection model with depth dependent viscosity $\eta(z)$ after McConnell (1968; model 62-12). $Ra \approx 3 \cdot 10^5$; heating from within $6.65 \cdot 10^{-8} \text{ W/m}^3$ and from below 12.56 mW/m^2 . *Top:* isotherms in $^{\circ}\text{C}$; *bottom:* stream lines; *right-hand side:* viscosity-depth function. Note that the high-viscosity lid does not participate in the circulation

the lateral boundaries, or better, between the neighbouring plates (because of the assumed mirror-image boundary condition). The lithosphere acts as a thermal insulator so that the mantle underneath heats up to unrealistically high temperatures of more than 2000°C in the solution presented.

This model, with an effective Rayleigh number of 2.6×10^5 did not reach a steady state; instead the circulation oscillated irregularly with an approximate period of $8 \cdot 10^7$ years about a state which is presented in Fig. 1, with a relative amplitude of 20%. Oscillations are characteristic for convection at high Ra in the laboratory (Krishnamurti 1970a; b). There they are, however, a typical three-dimensional phenomenon which cannot be modelled with a two-dimensional scheme. Since the above period is that of particle revolution of the inner cell region, the oscillatory behaviour at least resembles that of real physical systems and does not appear to be an artefact of the numerical method.

The aspect ratio of the circulation (Fig. 1) is 4 or even ≥ 5 if the lid is subtracted. One must ask whether this high aspect ratio might be a numerical effect of the high-aspect-ratio mesh. Our contention is that this is not so. Heating largely from within stabilizes large-aspect ratio flow (Tozer 1967a; Tritton and Zarraga 1967; DeLaCruz 1970; Thirlby 1970) and so does viscosity layering below the lid (Richter and Daly 1978; Daly 1980). Our contention is supported by the following: A test with the same model, but heated entirely from below, failed to produce a stable single cell; after 500 million years of unsteady variation a steady three-cell pattern developed. Similar tests with constant, instead of layered, viscosity below the lid also failed in this way. Finally the circulation of Fig. 1 is itself probably on the brink of instability, as suggested by the bumps of the stream lines and isotherms. We do not however claim that our limited computational procedure allows us to define the point of cell instability precisely.

The above model is obviously not satisfactory. Neither were experiments with lithospheres of reduced viscosity and/or thickness. As found by Houston and DeBreaecker (1975), the less rigid plates participated in the convection currents, but not as rigid plates; rather they deformed internally.

A better way of simulating the lithosphere is to introduce a yield stress in the model or at least in critical regions, i.e. weakness zones (Kopitzke 1979). If the yield stress is passed

during the computation, the viscosity at that location is reduced so that the material can yield and the shear stress is not exceeded in the next step. The yield stress is thus the upper limit of sustainable stress. It may not be a yield stress in the true physical sense but is meant to represent an average property of the weakness zone. In two lithospheric regions (stippled in Fig. 2) of the previous model the "yield stress" was set to 500 bar. The effect of this modification is seen clearly in Fig. 2. The lithosphere now partakes, without being deformed, in the convective circulation with a velocity of 2 cm/a. As demonstrated by the velocity depth profile (taken in the middle of the cell) the plate moves against resistance from the slower currents in the asthenosphere; it is not dragged by faster currents. The asthenosphere has become the zone of maximum shear.

The mean mantle temperature has dropped in the steady-state solution from the previous model to much lower values, about 830°C . This is low in the light of present estimates as e.g. Solomon (1976: $1100 \pm 100^{\circ}\text{C}$ in the low-velocity channel and $1300 \pm 150^{\circ}\text{C}$ near the olivine-spinel transition at 400 km depth) and Anderson (1980: $1100\text{--}1200^{\circ}\text{C}$, 100–400 km depth). It is also lower than most of Tozer's (1967a; b; personal communication, 1981) estimates based on convection theory and rheology. The lateral temperature variation computed is $200\text{--}300^{\circ}\text{C}$, characteristic of all our models with horizontal flow velocities of 1–2 cm/a and the assumed heat input. A more thorough discussion of these aspects follows after our results. The solution of Fig. 2 has a temperature inversion in the depth interval 200–400 km extending across nearly the whole length of the convection cell. This may preclude the small-scale Richter convection rolls (Richter and Parsons 1975), as argued by Kopitzke (1979), but it would not prevent instabilities of the boundary layers. Such a temperature inversion has not yet been resolved, except for an indication given by Anderson (1980) for the case that the seismic low-velocity zone consisted of dry lherzolite or eclogite.

Next, a viscosity-depth model was assumed with generally two orders of magnitude lower values than previously (Fig. 3). It was taken from a one-dimensional flow model with non-linear rheology based on laboratory data (Jacoby and Ranalli 1979), but with one order of magnitude further reduction. The strain-related viscosity in this case cannot be directly transferred from one model to another; it just served us as a test case for a low-viscosity model. As above, we introduced weakness

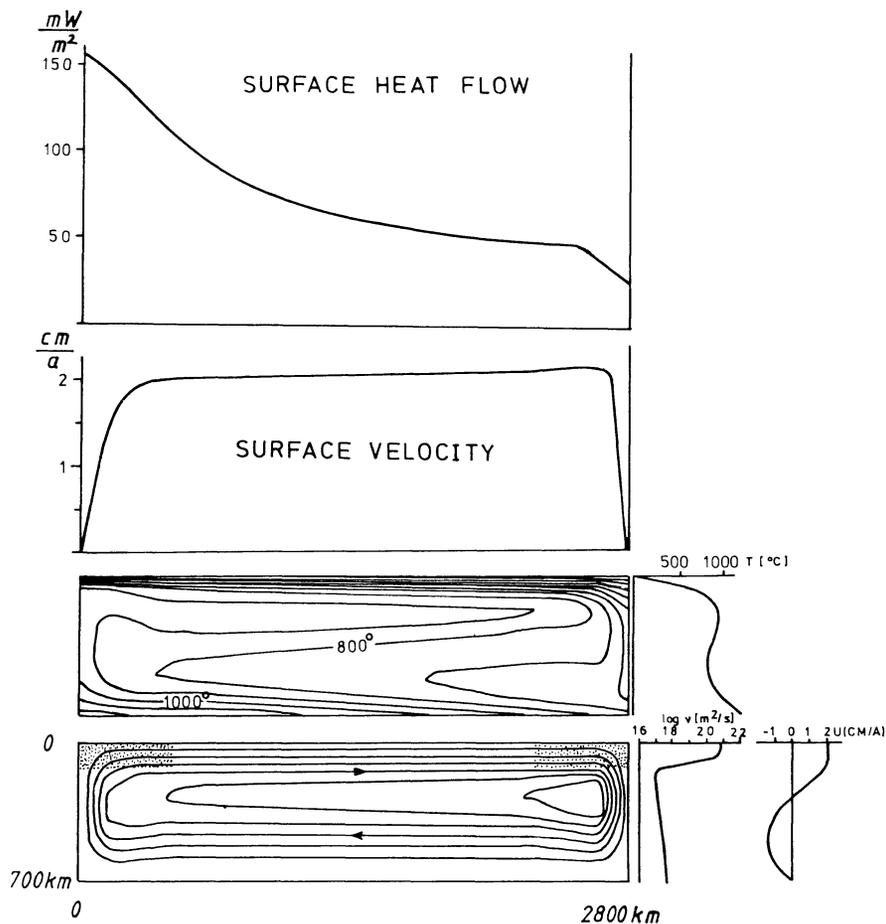


Fig. 2. Same as Fig. 1, but with yield stress (500 bar) assumed for stippled regions of lithosphere. $Ra \approx 1.5 \cdot 10^5$. *Center*: isotherms in $^{\circ}\text{C}$ with temperature-depth profile taken at center; *bottom*: stream lines with viscosity-depth profile and horizontal-velocity depth profile at center; *top*: surface velocity plotted versus x-axis and surface heat flow. Note that high-viscosity lid participates in circulation with nearly constant velocity

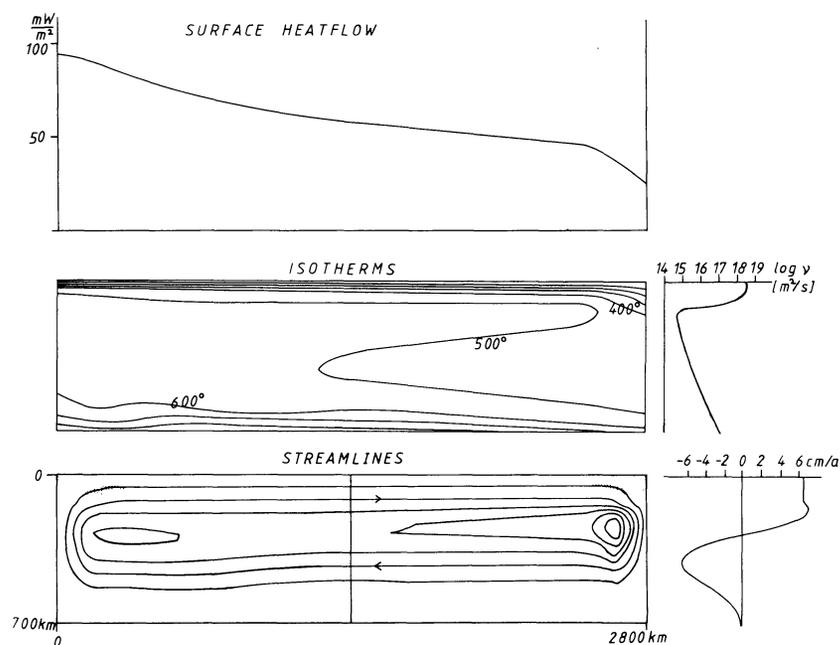


Fig. 3. Convection model with depth dependent viscosity $\eta(z)$ similar to one computed by Jacoby and Ranalli (1979) with weakened zones (stippled: $v = 10^{17} \text{ m}^2/\text{s}$); heating as in Fig. 1; $Ra \approx 1.25 \cdot 10^7$. *Bottom*: stream lines, but not steady state; vertical profile of horizontal velocity at right-hand side; *center*: isotherms; viscosity-depth profile at right-hand side; *top*: surface heat flow

zones (stippled in Fig. 3) in which the viscosity was simply reduced to $10^{17} \text{ m}^2/\text{s}$ (instead of by yield stress). In nature such a reduction of effective viscosity is achieved by fracturing, transient creep, plastic creep in narrow weakness zones, localised frictional heating, partial melting and thermal run-aways (Schubert and Yuen 1978; Kaula 1980).

As the model of Fig. 1, this also did not reach the steady state. The circulation oscillated with a relative amplitude of 3% and a constant period of $4 \cdot 10^7$ years about the state shown. Again the period is about that of particle revolution in the inner cell region. Changing the length of the time steps by factors of 2 and 4 during execution did not

affect the period or amplitude of oscillation, supporting the view that the behaviour reflects physical, not numerical instability.

The low viscosities lead to relatively high velocities, about 6 cm/a horizontally in the lithosphere and asthenosphere and up to 25 cm/a in the descending current. This is nowhere excessive: a viscosity decrease of two orders of magnitude has resulted in a velocity increase of only a factor of 3. Remember that the two models differ only in viscosity and have the same heat input.

The lower viscosities have led to much more efficient convection so that the driving temperature and density disturbance is strongly reduced. Correspondingly the average cell temperature has dropped to about 500°C and the lateral temperature difference to about 100°C. The temperature inversion has nearly vanished to only 10°C.

Comparison of the above models (Figs. 1, 2, 3) shows that the circulation pattern of the last one is somewhat intermediate between the former two: that of Figure 1 is characterized by asthenospheric currents driving, though ineffectively, the lithosphere; that of Fig. 2 is reversed, the lithosphere leading and thus driving the asthenospheric currents; in Fig. 3 the top asthenospheric currents are again faster, by a margin (0.5 cm/a), than the lithosphere (see velocity-depth profile taken at cell center). The lithosphere in this model is thus less "active" in the sense of being partly dragged by asthenospheric flow.

These models give us a clue to what determines the role of the lithosphere in the whole circulation. In the models the role is largely the result of the relative strength of coupling between the lithosphere and the asthenosphere, or the ratio of viscosity in the plate-decoupling zones and viscosity of the asthenosphere. Had we assumed a lower value than $10^{1.7} \text{ m}^2/\text{s}$ for the stippled zones of Fig. 3, the lithosphere would again have moved faster than the asthenosphere. A systematic study of this aspect is deferred to a later time.

The above solutions for different kinds of coupling of the plates demonstrate the broad capability of our numerical method. The major conclusions are not affected by the crude treatment. We must, however, be cautious as to quantitative details, e.g. the exact transition points from one mode to another, the stability of cells, the precise relative velocities of plates and asthenosphere. A reason for caution is the poor, smeared-out representation of the boundary layers in a mesh of 35 km vertical spacing. These points will be discussed again below.

The following experiments were conducted with materials having viscosities that depend on temperature and pressure (Newtonian) or on temperature, pressure and shear stress (non-Newtonian). We wanted to investigate whether such rheologies would automatically give convection involving the lithosphere. Based on the review by Stocker and Ashby (1973) of linear and non-linear creep mechanisms, one may consider Nabarro-Herring creep, Coble creep, and diffusion-accommodated grain boundary sliding as linear processes and dislocation creep as a non-linear mechanism; they all contribute to the effective viscosity (Jacoby and Ranalli 1979):

$$\eta = 1 \left/ \left(2 f \cdot \frac{D_{\text{eff}} \Omega}{k \cdot T \cdot d^2} + \frac{3^{n+1}}{2} \frac{A \cdot D_v \cdot b}{k T \mu^{n-1}} \tau^{n-1} \right) \right. \quad (4)$$

The meaning of the quantities and the numerical values assumed are presented in Tables 1 and 2. The first term combines the linear creep mechanisms and the second term de-

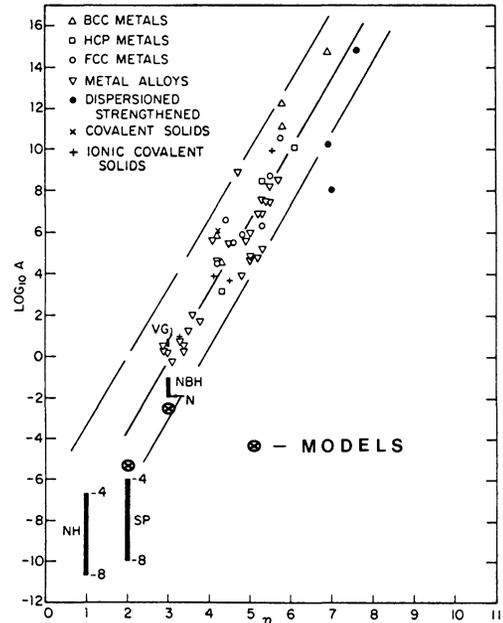


Fig. 4. Empirical relationship between Dorn's parameters A and n for various classes of metals (after Stocker and Ashby 1973) with the preferred values of our models added

scribes dislocation creep. Temperature and pressure dependence is mainly in the diffusivities:

$$D_v = D_{v0} \cdot \exp \left(- \frac{E_v + p V_v}{RT} \right);$$

$$D_{\text{eff}} = D_v (1 + \pi \delta / d \cdot D_B / D_v);$$

$$D_B = D_{v0} \exp \left(- \frac{E_B + p V_v}{RT} \right). \quad (5)$$

Similar rheologies have been investigated in boundary layer convection models of the upper mantle (Froidevaux and Schubert 1975; Schubert et al. 1976; Froidevaux et al. 1977; Schubert et al. 1978) and in convection models similar to ours (Rabinowicz et al. 1980; Daly 1980). Some of their results, e.g. concerning the effective viscosity, bear on our work and can be used as an additional test of our numerical method (see discussion, below).

The strong effect of temperature and stress in some of our models occasionally resulted in numerical, or indeed physical, instabilities. Since some of the rheological parameters responsible are in part not accurately known it seemed justified to vary them within plausible limits in order to achieve acceptable models: this mainly concerns grain size d , activation energy and volume E_v and V_v , and Dorn's parameters A and n . Our favoured values were $d = 10^{-3} \text{ m}$, $E_v = 5.4 \cdot 10^5 \text{ J/mol}$, $V_v = 10^{-5} \text{ m}^3/\text{mol}$, and the value of n was varied between 1 and 3; A was adjusted to give plausible flow velocities. Our favoured values are plotted on the experimental data for n , A as summarised by Stocker and Ashby (1973) (Fig. 4) and fall within the experimental scatter.

First we tested models having no weakened lithospheric zones, hoping that stress concentrations combined with power-law creep ($n=3$) might reduce the effective viscosity locally and decouple the surface plate. None of the many models investigated showed this effect, partly because the low temperatures dominated the rheology, but mainly because the expected instabilities cannot be represented in the coarse mesh. We

therefore again introduced weak zones. Now we located them 500 km from the lateral boundary so that subduction would be induced, but not at the boundary, i.e. not symmetrical (which it otherwise would be for reasons of computational symmetry). The crude method does not permit us to model detail, but the gross features can be studied. Furthermore, a series of tests demonstrated the need for some more decoupling of the descending flow by two narrow low-viscosity zones (stippled in Figs. 5-7) because otherwise the cold highly viscous material tended to stagnate. A viscosity of $5 \cdot 10^{16}$ – $4 \cdot 10^{17}$ m²/s in these weak zones did usually suffice for mobility to be regained. We are not happy about this manipulation, made in order to overcome the shortcomings of the crude mesh, but we also believe that nature, too, may provide some decoupling by narrow shear zones, local heating and melting, and at shallow depth, fracturing.

Figure 5 shows a model with linear T, p dependent rheology; the viscosity in the stippled regions is reduced to the constant value of 10^{17} m²/s. The initial condition was constant temperature at the bottom, with a small perturbation to initiate the flow in the direction wanted. For 150 million years model time, the lithosphere participated actively in the currents, but the return flow progressively concentrated at greater depth. Although the plate velocity was rather slow (maximum 1 cm/a heating of the subducted material was even slower and viscosity became so high that the return flow of plate material became inhibited. Drift and subduction of the lithosphere choked (the velocity decreased to 1–2 mm/a). Convection shifted to sublithospheric depths and split up into four cells. The results looked satisfactory for only 45 million years running time, afterwards they differed from our picture of real Earth behaviour; but admittedly we know little about the long-term aspect of plate motion.

We do not want to overemphasise the particular time behaviour described. The ADI method used cannot resolve rapid temporary changes. If, however, the change during a time step is small, as is the case here, the computation should not be grossly wrong. The important point we wish to make is that the model with the rheology chosen does not convect in a single stable cell; the flow breaks up to a new steady state with several cells. A similar evolutionary behaviour was observed with essentially the same computational method by Torrance and Turcotte (1971) and Rabinowicz et al. (1980).

The above model with mainly temperature dependent rheology is apparently incapable of describing steady-state convection with subduction during long time spans. If such a behaviour, of which we cannot be entirely certain, is favoured, it is not simulated by the model, particularly for fast-moving plates. A possible escape from the dilemma is non-linear, i.e. shear stress dependent rheology by which the effective viscosity in cold subduction regions can be sufficiently reduced to permit a steady-state return flow.

Equations (4) and (5) with Dorn's parameters $n=2$, $A=3 \cdot 10^{-6}$ describe the creep laws assumed for the next model (Fig. 6); the viscosity in the stippled regions was set to $4 \cdot 10^{17}$ m²/s. The result was the development of a steady-state convection cell with the lithosphere participating. It was about 80 km thick and thickening toward the region of descent. Below it a low-viscosity asthenosphere developed. Both drifted together at somewhat less than 2 cm/a. In the descending current, material from the lithosphere dropped at 0.5 cm/a while material from the asthenosphere flowed faster (0.9 cm/a). The figures quoted may be inaccurate by, say, 10%, but the general behaviour described seems significant.

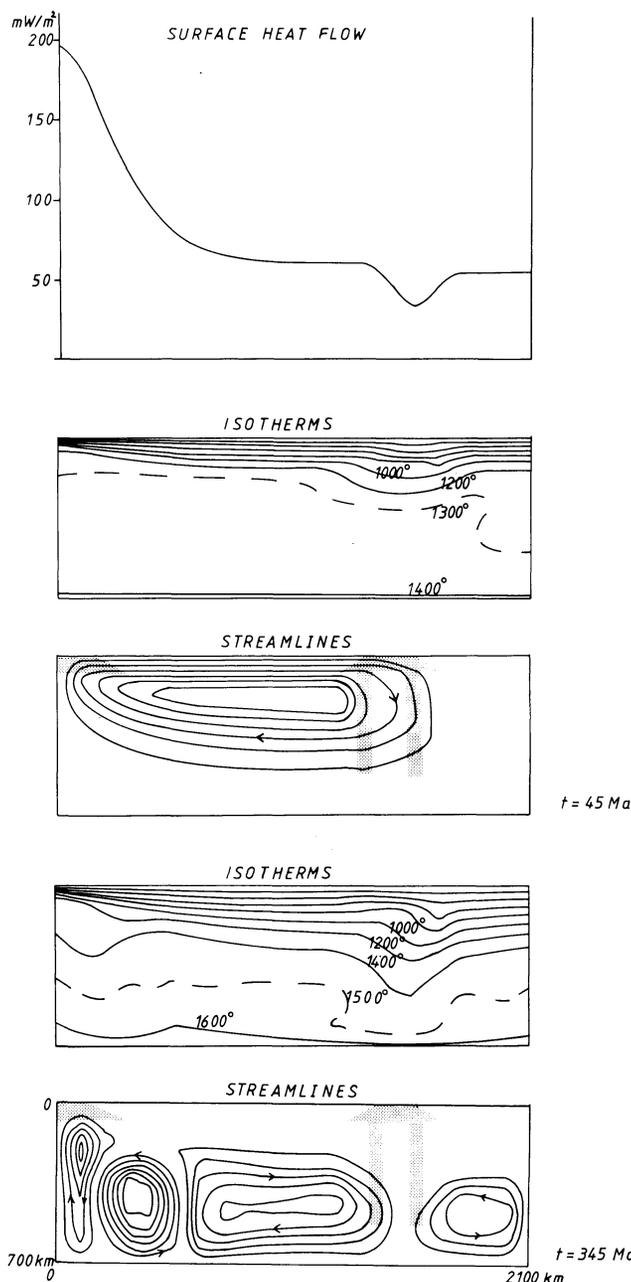


Fig. 5. Convection in a model with T, p dependent linear rheology with additional weakened zones (stippled) at two moments of time. At 45 million years the circulation encompasses the lithosphere (upper part: surface heat flow (top), isotherms (center), stream lines (bottom)); after 345 million years the surface is stagnant and four cells have developed in the asthenosphere (lower part: isotherms (top) and stream lines bottom))

A comparison of the above two models demonstrates that the reduction of the effective viscosity by shear stress prevents freezing of the descending current which can then turn into a return flow. Power-law creep dominates in the lower part of the descending current as can be seen by comparing the weakly and the strongly deformed regions in the field of computed viscosities (Fig. 6).

Note that a secondary cell developed behind the subduction region. Such cells are expected to exist beneath the marginal basins, e.g. of the western Pacific, and may provide an explanation for the basins and their high heat flow and

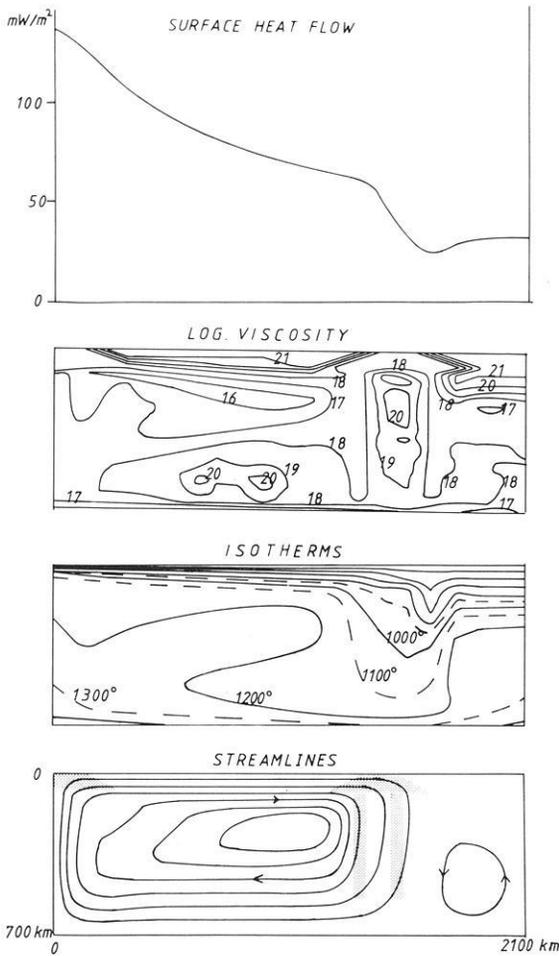


Fig. 6. Convection model with T, p, τ dependent non-linear rheology; Dorn's parameters $n=2, A=3 \cdot 10^{-6}$; weakened zones $v=4 \cdot 10^{17} \text{ m}^2/\text{s}$; surface heat flow (top), isotherms (second from bottom), stream lines with surface velocity $\sim 2 \text{ cm/a}$ (bottom), computed viscosity (second from top) with weakened zones visible. Dissipation and adiabatic heating included; if not, lithosphere descent becomes stagnant after 100 million years (in a model not shown here)

volcanism (Andrews and Sleep 1974; Rabinowicz et al. 1980). In the previous model with linear viscosity (Fig. 5) such a cell did not develop until subduction had choked and the asthenospheric convection had broken up into several cells.

It must be remarked that the features of model Fig. 6 were the result not only of non-linear stress dependent rheology; the thermal effects of adiabatic compression and decompression as well as of dissipation helped to keep the lithosphere going. This was different from previous models where these terms were also included without success. For example, one model (not shown) differed from that of Fig. 6 only by not including these effects; during the first 100 million years a plate drifted actively but then slowed down progressively to 0.5 cm/a while asthenospheric velocities were 10 cm/a . Obviously, in the model of Fig. 6, the heating of the descending flow by adiabatic compression was an important mechanism reducing the effective viscosity. Dissipation acted in the same direction, being proportional to the product of shear stress and strain rate. It is interesting to note that such non-linear processes as dissipation and power-law creep appear to stabilize the flow in the models.

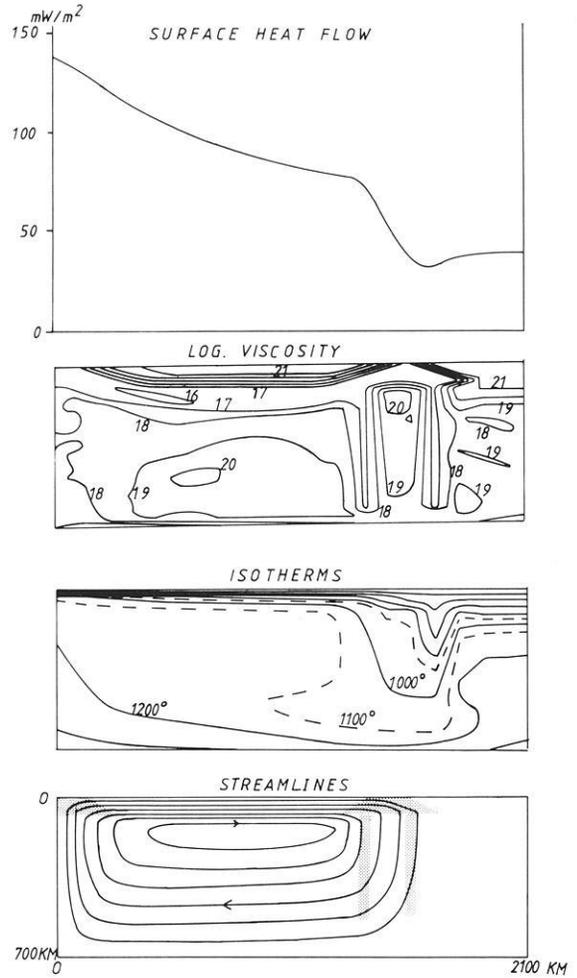


Fig. 7. Convection with T, p, τ dependent rheology; Dorn's parameters $n=3, A=4 \cdot 10^{-3}$; weakened zones $v=5 \cdot 10^{16} \text{ m}^2/\text{s}$. The model was quasi-steady-state after 150 million years; surface velocity was $\sim 3 \text{ cm/a}$. Arrangement of results as in Fig. 6. Note that the relatively higher temperatures were partly the result of a numerical heating excess (see text); the following model (Fig. 8) was computed to test whether similar results can be obtained with correct "physical" heating

The final model we present (Fig. 7) differs from the previous one by the choice of Dorn's parameters: $n=3, A=4 \cdot 10^{-3}$, and by the viscosity in the (stippled) decoupling zones: $v=5 \cdot 10^{16} \text{ m}^2/\text{s}$. In this model, in contrast to the previous one, the flow that developed in one direction was nearly restricted to the lithosphere while the return flow occurred at least partly in the asthenosphere. One of the reasons for this feature is probably that the strong non-linearity of the rheology promotes currents in which the zones of high shear rate and bending of the stream lines are spatially concentrated (into the asthenospheric shear zone). Another reason may be the more efficient decoupling in the stippled areas. Apart from the artificial aspects, the model generally simulates the lithosphere as an active part of the circulation rather well. Its velocity was 3 cm/a .

The temperatures appear to be quite in accord with what we know about the mantle, and this is coupled with realistic plate velocities. A close look reveals, however, an apparent heating excess of 35% resulting in an average surface heat flow of about 80 mW/m^2 (caused by inaccurate finite difference approximation of the partials in regions of highly vari-

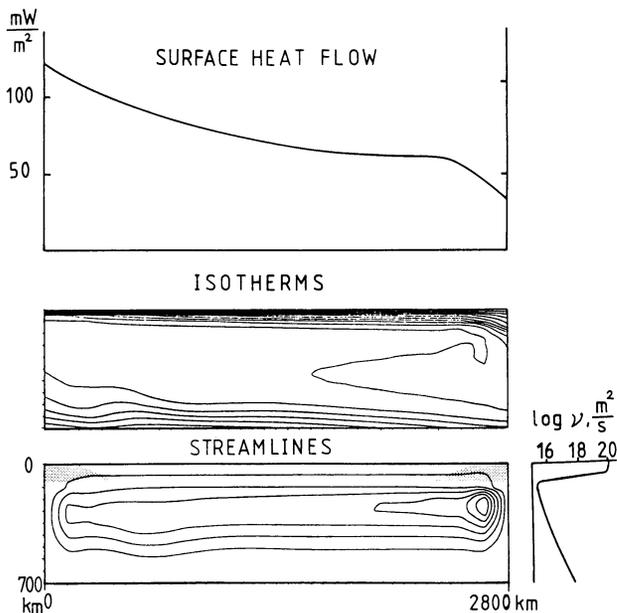


Fig. 8. Convection with depth-dependent viscosity $\eta(z)$, similar to that of Jacoby and Ranalli (1979); heat flow through bottom as in Fig. 1, internal heating raised to $9.4 \cdot 10^{-8} \text{ W/m}^3$ in order to add up to an average steady-state surface heat flow of 80 mW/m^2 . *Bottom*: streamlines, viscosity-depth profile at righthand side; *center*: isotherms; *top*: surface heat flow

able rheology and strain rate). This happens to be closer to the oceanic average of 78 mW/m^2 (Sclater et al. 1980) than the 60 mW/m^2 assumed above. It was this heating excess that led to the higher temperatures.

Similar temperatures result if the total “physical” heating corresponds to a steady-state average heat loss of 80 mW/m^2 and dissipation is computed correctly. This is shown by the model of Fig. 8, which has simply a depth-dependent viscosity and weak zones ($4 \cdot 10^{16} \text{ m}^2/\text{s}$). The solution oscillated (relative amplitude $< 3\%$, period $8 \cdot 10^7 \text{ yr}$); the lithospheric velocity of 2.7 cm/a and the mean temperature of $1,012^\circ$ are close to those of Fig. 7 (2.9 cm/a and $1,090^\circ$); there the low “physical” heat input and the high “numerical” dissipation seem thus to have cancelled to give, in a sense, a realistic model. By “realistic in a gross sense” we mean: large aspect ratio; nearly steady state; active participation of the lithosphere at plausible velocity, crudely subducting; acceptable mantle temperatures: all this obtained from a set of plausible assumptions (physics of the problem, material). Details, e.g. of viscosity distribution (Fig. 7), temperature and flow fields should be taken with a grain of salt; they are affected by the crude mesh, the numerics, and the artificially assumed weakness zones. We are therefore limited to drawing conclusions about the mantle only from the gross features of the results.

Discussion

In this section we wish to discuss more generally some aspects of temperature, rheology and viscosity, and flow patterns.

In most of our models (except Figs. 6–8) the total heat input was rather small by current knowledge. Correspondingly, the plate velocities and/or the average cell temperatures came out relatively low: $< 10 \text{ cm/a}$ and less than 1100° – 1300°C , respectively, which are believed to be realistic (So-

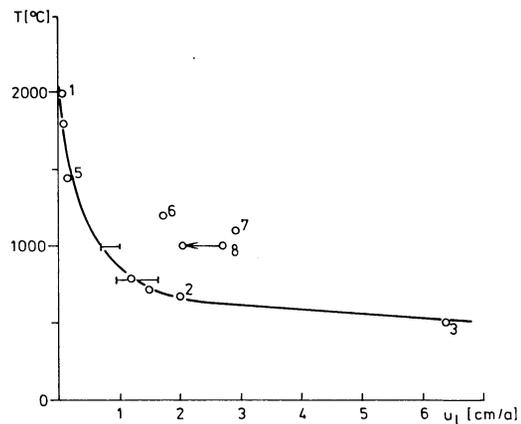


Fig. 9. Average model temperatures versus plate velocity u_l of models with aspect ratios 3 and 4. Not all models shown on diagram are discussed in paper (model numbers refer to Figures). *Arrow* points to position of model if scaling length L is increased (see text)

lomon 1976; Anderson 1980). Only as a result of insulation by a stagnant, strong lid (Fig. 1) did we observe high cell temperatures. As soon as plate motion makes convective heat transfer efficient, cell temperatures drop, in many models below the plausible range. In models with non-linear rheology, temperatures were closer to expectation, due to the heating excess which, as discussed above, had the right magnitude. If the temperatures are still considered too low, it should be remembered that the adiabatic gradient (or Di) is not well known and may be higher than assumed here: with the rheologies assumed, the viscosity would then drop, convection would become more efficient and temperatures would again drop in a self-regulating manner. The exact effect of varying the adiabatic gradient on self-regulation needs to be studied systematically.

Another reason for the relatively low temperatures and/or velocities may be our assumption of a shallow bottom to convection (700 km depth), as discussed by Schmeling (1980). The argument is as follows. All models (shown and not shown) computed with the same heat input and the same aspect ratio, but with different rheologies follow an empirical relationship of the form (Fig. 9):

$$u_l \cdot (\bar{T} - T_{num}) = c \quad (6)$$

where u_l is the lithosphere velocity, \bar{T} is the average model temperature, T_{num} a constant related to the particular choice of the grid, and c is a constant depending only on the total heat input (and the aspect ratio), but not on rheology or viscosity. Low plate velocity and/or temperature means a low value of c ; it does, however, increase with the dimensionless heat input $H = \mathcal{H} \cdot L^2 / (\rho \cdot c_p \cdot \kappa_0 \cdot T_c)$. The models of Fig. 6 (20% increase) and Fig. 7 (35%) clearly lie above the curve of Fig. 9; increasing \mathcal{H} in Fig. 8 by a factor of $4/3$ (80 mW/m^2 over 60 mW/m^2) is equivalent to an equal increase of L^2 with heating unchanged. In that case the dimensional velocity drops to about 2 cm/a , as indicated by the arrow on Fig. 9. The model would still clearly lie above the curve for constant heat input. Thus empirically c (velocity \times mean temperature) grows with L in the case of “mixed” heating from within and below, as has been found also by Kopitzke (1979).

The theoretical difficulties of treating the case of mixed heating do not arise in the case heated from below only. Here, however, mean velocity v and temperature fluctuation θ

do not depend on L to first approximation, if the superadiabatic temperature difference ΔT is fixed. The heat input is then (F. Busse, personal communication, 1981): $h = Nu k \Delta T$ (k = thermal conductivity); with $Nu \approx (Ra - Ra_c)^{\frac{1}{2}} / Ra_c^{\frac{1}{2}}$, $v \approx (Ra - Ra_c)^{\frac{1}{2}} \kappa / L \approx Ra_c^{\frac{1}{2}} h / (\Delta T \rho c_p)$, and $\theta \approx \Delta T (Ra - Ra_c) / Ra$, both independent from L . The fixed- ΔT case is, however, not relevant to our problem.

If higher mantle temperatures and plate velocities than in our shallow models are favored, e.g. by reliable inference of temperatures from observations, then, on the basis of the above discussion, deep-mantle convection must be favoured over shallow. This would also require that a considerable portion of the heating be internal, by radioactive decay and/or by the heat lost in cooling of the Earth, i.e. by stored heat.

On the other hand, we may accept the relatively low temperatures we found. Tozer (1967a, b) pointed out that a self-regulating convection mechanism through T (and p , τ) dependent rheology without fixed temperature bounds may indeed adjust to temperatures as low as computed in our models. Some of the temperature estimates (Solomon 1976) may be biased toward high values as they sample volcanism. This cannot be said about Anderson's (1980) estimate based on average seismic Earth models and laboratory data on minerals; the suggested temperature range is correspondingly lower.

Because of the low numerical resolution, it would be valuable to compare our results (temperature, effective viscosity, flow patterns, Figs. 6, 7) with physically similar models, obtained with different techniques. The most relevant models seem to be those based on boundary layer theory, giving more accurate results in the boundary layer region, with similar rheological assumptions (Froidevaux and Schubert 1975; Schubert et al. 1976; Froidevaux et al. 1977; Schubert et al. 1978). The closest resemblance in physical assumptions seems to be between the model of our Fig. 7 and that of fig. 6 of Schubert et al. (1976); in both cases rheology is nonlinear with Dorn's parameter $n=3$ and activation energy $E=5.4 \cdot 10^5$ J/mol; see Table 1. We found a minimum viscosity increasing with age of the overlying lithosphere from $\lesssim 3 \cdot 10^{19}$ Pa s at ~ 90 km depth to $\lesssim 3 \cdot 10^{20}$ Pa s at ~ 150 km depth; their equivalent values are: $5 \cdot 10^{19}$ Pa s at 70 km depth (10 million years) and $\sim 10^{20}$ Pa s at 180 km (150 million years). The coarse mesh cannot, of course, resolve accurately the boundary layer structure, however, the comparison with the results of Schubert et al. (1976, Fig. 7) shows that the resolution does not lead to grossly wrong results. Crudely, we find the high-gradient zones of temperature and horizontal velocity to thicken with age from 50–80 km, and from 80–110 km, respectively; their equivalent values were 50–100 and 80–160 km, respectively. The smearing-out of the crude mesh seems mainly to give less variation, not so much to give a greater thickness. Part of the discrepancy in the horizontal velocities also comes from the different model assumptions for the bottom: in our model the return flow must be accommodated within 700 km depth: their assumption is that horizontal velocity approaches zero "at great depth". Apart from this the flow patterns are not dissimilar.

The above comparison does not prove our results correct, neither generally nor in specific details, but we find it encouraging that even the boundary layer, and that in the most non-linear case, was represented rather satisfactorily in our models. It thus appears that the gross features and the major conclusions we draw from them are quite reliable. This is not

to be said about quantitative details generally, as discussed above, but we did not in any case ask for such details.

On the basis of our results, we cannot argue strongly for whole versus shallow mantle convection. The argument of Richter and McKenzie (1978) for shallow convection is mainly that the stress distribution in the descending slabs cannot easily be explained without a bottom at about 700 km depth (Isacks and Molnar 1971). Plausibly, however, a compressive resistance could also develop simply by a viscosity increase above and below 700 km without necessarily preventing the currents from extending deeper (Davies 1977; Jacoby 1978; Kopitzke 1979).

Conclusions

Our main conclusions are listed below.

1. Only by the introduction of weak zones into the lithosphere was it possible to simulate its drift as a rigid, i.e. undeformed, body. It thus seems that plate velocity is essentially controlled by the strength or weakness of the mechanical coupling with neighbouring plates as well as with the lower, more viscous parts of the upper mantle.

The coupling near the surface and in the mesosphere may, however, behave differently. If in the subduction zones some kind of shear instability (which cannot be modelled with our techniques) develops, the coupling will remain weak, maybe approximately steady-state (unless other processes, such as collision, interfere – see McKenzie 1969). In contrast, plate descent into the mesosphere increases the coupling progressively until it may inhibit the motion. Non-linear processes do, however, lessen this effect. Little can be said at present about such a phenomenon in nature, nor did our models that choked run for long enough, but the time lapse until a reactivation of plate motion would have been at least 10^9 a. Furthermore, our method is not applicable to the study of the time behaviour. It may be added, that our quasi-steady state models (e.g. Figs. 6, 7) remained so as long as the computation continued (500, 270 million years).

2. While the rheology of the plates themselves has turned out to be of major importance for the plate motions, that of the material below, apart from its generally low viscosity, is less important for the plate velocities. For viscosity variations of several orders of magnitude and for both linear and non-linear viscosity, the plate velocities varied by no more than 1–6 cm/a.

3. Mantle rheology is, however, important for the form of the return flow. Mostly we have obtained broad maxima of this flow at depths not shallower than 400 km, suggesting that in the earth it is not the asthenosphere which mainly carries the return flow. The asthenosphere is probably dragged with the plates to a considerable extent (see Fig. 6) or it is the zone of concentrated shear flow, an effect which is enhanced by non-linear rheology (see Figs. 5, 7). Besides the rheology assumed by us, other non-linear effects may further concentrate the shear; such effects include dissipation, partial melting and upward concentration of melt (which is assisted by the strong shear), and thermal runaways.

4. As discussed in the previous chapter, we obtained low temperatures and/or plate velocities, even when the heat input was realistic. The average temperature and velocity are coupled through the heat input and cell geometry (fixed in our models or only slightly varied); the non-dimensional heat input can be increased by increasing the depth or the general

size of the cell; dimensions of $6\,000 \times 2\,000 \text{ km}^2$ or even $9\,000 \times 3\,000 \text{ km}^2$ (instead of $2\,100 \times 700 \text{ km}^2$) may thus be preferred. For this it would not be critical if the viscosity in the lower mantle increases by one to even three orders of magnitude.

The low temperatures may, however, be realistic estimates of the average upper-mantle temperatures. Our experiments suggest that the mantle will convect strongly at such temperatures, if the thermal and mechanical upper boundary layers are decoupled piecewise at their margins, and are thus free to move. If the surface becomes stagnant, the interior convects at much higher temperature because the heat loss is inhibited. In the earth's upper mantle temperature may rise locally, but not to as high values as in the lidded case; hot spots or volcanic spots may be witnesses of local temperature anomalies and should therefore be taken with caution in estimating upper-mantle temperatures.

5. The importance of decoupling for the whole dynamical problem, plate motions, temperature and thermal evolution, makes it urgent to study the decoupling mechanisms thoroughly. It appears possible that besides the mechanisms mentioned earlier (yield stress, fracturing, transient creep, plastic creep on faults, frictional heating, local melting, thermal runaways), and for some of these mechanisms, the presence or absence of minor volatile constituents, particularly water, may be of paramount importance. The fact that the earth's atmosphere and lithosphere are moist, in contrast to Venus that is dry, may explain the great difference of the surfaces of these two, otherwise rather similar planets (H. Spetzler, personal communication, 1981).

6. A considerable temperature inversion usually developed in the convection cells, amounting to 100–200 K if the effective viscosity was of the order of $10^{17} \text{ m}^2/\text{s}$, but only to some 10 K for $10^{15} \text{ m}^2/\text{s}$. The inversion increases as the plate motions are facilitated by reduced coupling. This feature, if relevant to the earth, may stabilise the large-scale mantle flow (Kopitzke 1979) and inhibit smaller-scale convection rolls as proposed by Richter (1973a), but it would enhance boundary layer instabilities, such as rising plumes. The large-scale flow may also become destabilised if plates grow to an excessive size, which seems possible if plate size is largely determined by its growth and destruction rates at existing ridges and trenches, leading to a multiplicity of horizontal scales (Richter and Daly 1978). Destabilisation would be the consequence of the inefficient heat loss, since only the generation and destruction of the plates transfer heat convectively and elsewhere heat loss is by conduction through the boundary layer. There is evidence from bathymetry, gravity, and the geoid for three-dimensional flow below the plates (McKenzie et al. 1980). Three-dimensional instabilities below the Pacific may be enhanced by the circumstance that the Pacific is shrinking; hence the upper mantle may be more stagnant (Hager and O'Connell 1978; Grohmann 1980) than a simple two-dimensional model of return flow would suggest.

Acknowledgements. Helpful critical comments on earlier drafts of this paper were provided by F. Busse, D.A. Tozer, and an anonymous reviewer. U. Christensen (formerly: Kopitzke) gave us an earlier version of the computer program. We have profited from discussions with U. Christensen, J.C. DeBremaecker, S.F. Daly, H. Spetzler, D.P. McKenzie, G. Schubert, T. Spohn, and others. We like to thank W. Mahler for drafting the figures and I. Hörnchen for typing the manuscript. The computations were done at Hochschulrechenzentrum, University of Frankfurt. Deutsche Forschungsgemeinschaft gave us financial support (grant Ja 258/6).

References

- Anderson, O.L.: The temperature profile of the upper mantle. *J. Geophys. Res.* **85**, 7003–7010, 1980
- Andrews, D.J.: Numerical simulation of sea-floor spreading. *J. Geophys. Res.* **77**, 6470–6481, 1972
- Andrews, D.J., Sleep, N.H.: Numerical modeling of tectonic flow behind island arcs. *Geophys. J. R. Astron. Soc.* **38**, 237–251, 1974
- Artyushkov, E.V.: Gravitational convection in the interior of the earth. *Izv. Earth Phys.* **9**, 3–17, 1968 (trans. J. Büchner)
- AvéLallemant, H.G., Carter, N.L.: Syntectonic recrystallisation of olivine and modes of flow in the upper mantle. *Geol. Soc. Am. Bull.* **81**, 2203–2220, 1970
- Bénard, H.: Les tourbillons cellulaires dans une nappe liquide. *Rev. Gen. Sci.* **11**, 1261–1271, 1900
- Busse, F.H.: On the stability of two-dimensional convection in a layer heated from below. *J. Math. Phys.* **46**, 140–150, 1967
- Busse, F.H., Whitehead, J.A.: Instabilities of convection rolls in high Prandtl number fluid. *J. Fluid Mech.* **47**, 305–320, 1971
- Chandrasekhar, S.: The thermal instability in spherical shells. *Phil. Mag., ser. 7*, **44**, 233–241, 1953
- Chandrasekhar, S.: *Hydrodynamic and Hydromagnetic Stability*. Oxford, Clarendon Pr.: 1961
- Chapman, D.S., Pollack, H.N.: Global heat flow: A new look. *Earth Planet. Sci. Lett.* **28**, 23–32, 1975
- Chapple, W.M., Tullis, T.E.: Evaluation of the forces that drive the plates. *J. Geophys. Res.* **82**, 1967–1984, 1977
- Daly, S.F.: The vagaries of variable viscosity convection. *Geophys. Res. Lett.* **7**, 841–844, 1980
- Davies, G.F.: Whole mantle convection and plate tectonics. *Geophys. J. R. Astron. Soc.* **49**, 459–486, 1977
- DeBremaecker, J.C.: Convection in the earth's mantle. *Tectonophysics* **41**, 195–208, 1977
- DeLaCruz, S.: Asymmetric convection in the upper mantle. *Rev. Union. Geofis. Mexico.* **10**, 49–56, 1970
- Forsyth, D., Uyeda, S.: On the relative importance of the driving forces of plate motion. *Geophys. J. R. Astron. Soc.* **43**, 163–200, 1975
- Foster, T.D.: Convection in a variable viscosity fluid heated from within. *J. Geophys. Res.* **74**, 685–693, 1969
- Froidevaux, C., Schubert, G.: Plate motion and structure of the continental asthenosphere: a realistic model of the upper mantle. *J. Geophys. Res.* **80**, 2553–2564, 1975
- Froidevaux, C., Schubert, G., Yuen, D.A.: Thermal and mechanical structure of the upper mantle: a comparison between continental and oceanic models. *Tectonophysics* **37**, 233–246, 1977
- Gebrande, H.: Ein Beitrag zur Theorie thermischer Konvektion im Erdmantel mit besonderer Berücksichtigung der Möglichkeit eines Nachweises mit Methoden der Seismologie. *Diss. München*, 1975
- Goetze, C., Brace, W.F.: Laboratory observations of high-temperature rheology of rocks. *Tectonophysics* **13**, 583–600, 1972
- Grohmann, N.: Die 2-Scale Zellularkonvektion. Untersuchung über die Zusammenhänge zwischen Orogenese, Kontinentaldrift und Magmatismus. *Diss. München*, 1980
- Hager, B.H., O'Connell, R.J.: Subduction zone dip angles and flow driven by plate motion. *Tectonophysics* **50**, 111–133, 1978
- Harper, J.F.: On the driving forces of plate tectonics. *Geophys. J. R. Astron. Soc.* **40**, 465–474, 1975
- Houston, M.H., DeBremaecker, J.C.: ADI solution of free convection in variable viscosity fluid. *J. Comp. Phys.* **16**, 221–239, 1974
- Houston, M.H., DeBremaecker, J.C.: Numerical models of convection in the upper mantle. *J. Geophys. Res.* **80**, 742–751, 1975
- Isacks, B., Molnar, P.: Distribution of stresses in the descending lithosphere from a global survey of focal mechanism solutions of mantle earthquakes. *Rev. Geophys. Space Phys.* **9**, 103–174, 1971
- Jacoby, W.R.: Instability in the upper mantle and global plate movements. *J. Geophys. Res.* **75**, 5671–5680, 1970
- Jacoby, W.R.: One-dimensional modelling of mantle flow. *Pure Appl. Geophys.* **116**, 1231–1249, 1978

- Jacoby, W.R., Ranalli, G.: Non-linear rheology and return flow in the mantle. *J. Geophys.* **45**, 299-317, 1979
- Jeffreys, H.: The stability of a layer of fluid heated below. *Philos. Mag.* **57** (2), 833-844, 1926
- Kaula, W.M.: Material properties for mantle convection consistent with observed surface fields. *J. Geophys. Res.* **85**, 7031-7044, 1980
- Kopitzke, U.: Finite element convection models: Comparison of shallow and deep mantle convection, and temperatures in the mantle. *J. Geophys.* **46**, 97-121, 1979
- Krishnamurti, R.: On the transition to turbulent convection. 1. The transition from two to three dimensional flow. *J. Fluid Mech.* **42**, 295-307, 1970a
- Krishnamurti, R.: On the transition to turbulent convection. 2. The transition to time dependent flow. *J. Fluid Mech.* **42**, 309-320, 1970b
- McConnell, R.K.: Viscosity of the mantle from relaxation time spectra of isostatic adjustment. *J. Geophys. Res.* **73**, 7089-7105, 1968
- McKenzie, D.P.: Speculations on the consequences and causes of plate motions. *Geophys. J. R. Astron. Soc.* **18**, 1-32, 1969
- McKenzie, D.P.: Surface deformation, gravity anomalies and convection. *Geophys. J. R. Astron. Soc.* **48**, 211-238, 1977
- McKenzie, D.P., Watts, A., Parsons, B., Roufousse, M.: Planform of mantle convection beneath the Pacific Ocean. *Nature* **288**, 442-446, 1980
- Minster, J.B., Jordan, T.H., Molnar, P., Haines, E.: Numerical modeling of instantaneous plate tectonics. *Geophys. J. R. Astron. Soc.* **36**, 541-576, 1974
- Minster, J.B., Jordan, T.H.: Present-day plate motions. *J. Geophys. Res.* **83**, 5331-5354, 1978
- Morgan, W.J.: Convection plumes in the lower mantle. *Nature* **230**, 42-43, 1971
- Parmentier, E.M., Turcotte, D.L., Torrance, K.E.: Studies of finite amplitude non-Newtonian thermal convection with application to convection in the earth's mantle. *J. Geophys. Res.* **81**, 1839-1846, 1976
- Rabinowicz, M., Lago, B., Froidevaux, C.: Thermal transfer between the continental asthenosphere and the oceanic subducting lithosphere: its effect on subcontinental convection. *J. Geophys. Res.* **85**, 1839-1853, 1980
- Rayleigh, J.W.S.: On convection currents in a horizontal layer of fluid when the higher temperature is on the under side. *Philos. Mag.* **VI**, ser. 32, 529-546, 1916
- Richardson, R.M., Solomon, S.C., Sleep, N.H.: Tectonic stress in the plates. *Rev. Geophys. Space Phys.* **17**, 981-1019, 1979
- Richter, F.M.: Dynamical models for sea floor spreading. *Rev. Geophys. Space Phys.* **11**, 223-287, 1973a
- Richter, F.M.: Finite amplitude convection through a phase boundary. *Geophys. J. R. Astron. Soc.* **35**, 265-276, 1973b
- Richter, F.M., Daly, S.F.: Convection models having a multiplicity of large horizontal scales. *J. Geophys. Res.* **83**, 4951-4956, 1978
- Richter, F.M., McKenzie, D.P.: Simple plate models of mantle convection. *J. Geophys.* **44**, 441-471, 1978
- Richter, F.M., Parsons, B.: On the interaction of two scales of convection in the mantle. *J. Geophys. Res.* **80**, 2529-2541, 1975
- Roberts, P.H.: Convection in horizontal layers with internal heat generation. *Theory. J. Fluid Mech.* **30**, 33-49, 1967
- Ross, J.V., Nielsen, K.C.: High temperature flow of wet polycrystalline enstatite. *Tectonophysics* **44**, 233-261, 1978
- Schmeling, H.: Numerische Konvektionsrechnungen unter Annahme verschiedener Viskositätsverteilungen und Rheologien im Mantel. Diplomarb. Frankfurt, 1979 - also: Ber. Inst. Met. Geophys. Un. Frankfurt, Nr. **41**, 1980
- Schubert, G., Turcotte, D.L.: Phase changes and mantle convection. *J. Geophys. Res.* **76**, 1424-1432, 1971
- Schubert, G., Yuen, D.A., Turcotte, D.L.: Role of phase transitions in a dynamic mantle. *Geophys. J. R. Astron. Soc.* **42**, 705-735, 1975
- Schubert, G., Yuen, D.A.: Shear heating instability in the earth's upper mantle. *Tectonophysics* **50**, 197-205, 1978
- Schubert, G., Froidevaux, C., Yuen, D.A.: Oceanic lithosphere and asthenosphere: thermal and mechanical structure. *J. Geophys. Res.* **81**, 3525-3540, 1976
- Schubert, G., Yuen, D.A., Froidevaux, C., Fleitout, L., Souriau, M.: Mantle circulation with partial shallow return flow: effects on stresses in oceanic plates and topography of the sea floor. *J. Geophys. Res.* **83**, 745-758, 1978
- Slater, J.G., Jaupart, C., Galson, D.: The heat flow through the oceanic and continental crust and the heat loss of the earth. *Rev. Geophys. Space Phys.* **18**, 269-311, 1980
- Solomon, S.C.: Geophysical constraints on radial and lateral temperature variations in the upper mantle. *Am. Mineral.* **61**, 788-803, 1976
- Solomon, S.C., Sleep, N.H., Richardson, R.M.: On the forces driving plate tectonics: Inferences from absolute plate velocities and intra-plate stress. *Geophys. J. R. Astron. Soc.* **42**, 769-801, 1975
- Stocker, R.L., Ashby, M.F.: On the rheology of the upper mantle. *Rev. Geophys. Space Phys.* **11**, 391-426, 1973
- Takeuchi, H., Sakata, S.: Convection in a mantle with variable viscosity. *J. Geophys. Res.* **75**, 921-927, 1970
- Thirlby, R.: Convection in an internally heated layer. *J. Fluid Mech.* **44**, 673-693, 1970
- Torrance, K.E., Turcotte, D.L.: Structure of convection cells in the mantle. *J. Geophys. Res.* **76**, 1154-1161, 1971
- Tozer, D.C.: Some aspects of thermal convection theory for the earth's mantle. *Geophys. J. R. Astron. Soc.* **14**, 395-402, 1967a
- Tozer, D.C.: Towards a theory of thermal convection in the mantle. In: *The Earth's Mantle*, T.F. Gaskell, ed.: pp. 327-353. New York: Academic Press 1967b
- Tritton, D.J., Zarraga, M.N.: Convection in horizontal layers with internal heat generation, Experiments. *J. Fluid Mech.* **30**, 21-31, 1967
- Turcotte, D.L., Torrance, K.E., Hsu, A.T.: Convection in the earth's mantle. *Meth. Comp. Phys.* **13**, 431-454, 1973
- Vetter, U.R.: Stress and viscosity in the asthenosphere. *J. Geophys.* **44**, 231-244, 1978
- Zienkiewicz, O.L.: *The Finite Element Method in Engineering Science*. London: McGraw-Hill 1971

Received February 2, 1981, Revised version July 15, 1981
Accepted September 16, 1981

The 3 January 1978 Interplanetary Shock Event as Observed by Energetic Particle, Plasma and Magnetic Field Devices on Board of HELIOS-1, HELIOS-2 and PROGNOZ-6*

A.K. Richter¹, M.I. Verigin², V.G. Kurt³, V.G. Stolpovsky³, K.I. Gringauz², E. Keppler¹, H. Rosenbauer¹, F.M. Neubauer⁴, T. Gombosi⁵, and A. Somogyi⁵

¹ Max-Planck-Institut für Aeronomie Lindau, Katlenburg-Lindau, Federal Republic of Germany

² Space Research Institute, USSR Academy of Sciences, Moscow, USSR

³ Institute of Nuclear Physics, Moscow State University, Moscow, USSR

⁴ Institut für Geophysik und Meteorologie der Technischen Universität, Braunschweig, Federal Republik of Germany

⁵ Central Research Institute for Physics, Hungarian Academy of Sciences, Budapest, Hungary

Abstract. We study the different characteristics of the interplanetary medium and of the secondary enhancements in the energetic particle fluxes in the regions before, at and after a flare-generated interplanetary shock wave, as was observed by the solar wind plasma, magnetic field, and energetic particle experiments on board HELIOS-1, HELIOS-2 and PROGNOZ-6 near the Earth's orbit but at different helio-longitudinal positions about 40° apart on 3 January 1978, in the trailing edge of a solar particle event. By intercomparison of the different features in the interplanetary medium, in the energetic particle fluxes, and in the overall structure of the shock wave observed by these three spacecraft, respectively, we find: First, that this shock exhibits the properties of both an *R*- and *F*-type shock event, depending on the point of its observation in space. Second, that the pre-shock intensity enhancements of energetic particles can be explained by a cumulative first-order Fermi acceleration process of successive reflections of these particles at the shock in the entire region between the sun and the orbit of earth. Third, that the post-shock intensity enhancements of energetic particles are due to an acceleration of these particles by the shock wave and/or a trapping of these particles in the downstream region, and a second order Fermi acceleration of these particles by interactions with random Alfvénic and sound wave turbulences. Fourth, that the overall and rapid decrease in the energetic particle intensity observed by all three spacecraft in the downstream region of the shock wave is due to the occurrence of a magnetically closed region (magnetic cloud). Fifth, that the energetic particle enhancements at the shock cannot, at least in case of HELIOS-2, be explained uniquely by the shock drift acceleration mechanism only, as the characteristic angle between the upstream magnetic field direction and the shock surface is not small enough to allow this acceleration mechanism to operate effectively. Thus, highly oblique shocks can also be accompanied by energetic particle intensity increases.

Key words: Particle acceleration – Interplanetary shock waves – Structure of interplanetary shock – Multispacecraft observation

Introduction

Ever since particle and field measurements in interplanetary space with high sensitivity and sufficient high time, energy, and angular resolution have become available, many attempts have

been made to study in some detail the frequently observed phenomenon that the flux of charged energetic particles increases in association with propagating interplanetary shocks. Attempts have also been made to compare these observations with current theories on particle acceleration by shock waves (Armstrong et al. 1977, and references therein). These investigations were greatly improved in cases when observations of one and the same event by two or more spacecraft well separated in radial distance from the sun and/or in longitudinal distance were available (Burlaga et al. 1980, and references therein).

Following this latter line of investigation, we shall study in some detail the energetic particle intensity enhancements associated with an interplanetary shock wave observed by the two spaceprobes HELIOS-1 and -2 and the earth-orbiting satellite PROGNOZ-6 on 3 January 1978. Though at that time these three spacecraft were separated by a maximum radial distance of only 0.05 AU and a maximum longitudinal distance of only 40°, HELIOS-1 observed very different features in the fluxes of energetic particles in comparison with HELIOS-2 and PROGNOZ-6, though associated with the same shock wave. In order to understand this phenomenon within the framework of common theories on shock acceleration processes, we will first discuss briefly the overall particle event of 1–4 January 1978, as observed by these three spacecraft and in connection with solar observations. We will then study in greater detail the solar wind plasma and interplanetary magnetic field observations at the time the flare-associated shock wave passes the three spacecraft in question. These observations will provide some insight into the different processes of modulation and acceleration of energetic particles that will occur in the regions before, at and after the shock wave at the positions of these three satellites. Finally, we shall discuss the observed increases of energetic particle fluxes and their variations over these regions, in connection with model calculations on the interaction of these particles with propagating interplanetary shock waves.

The interplanetary plasma, magnetic field and energetic particle data used for this study have been provided in case of HELIOS-1 and -2 by the MPE/MPAE plasma experiment, by the fluxgate magnetometer of the University of Braunschweig, by the MPAE energetic particle spectrometer, and by the energetic particle experiment of the University of Kiel (for further details see the special HELIOS issue of J. Geophys. **42**, 1977), and in case of PROGNOZ-6 by the IKI plasma experiment (Gringauz et al. 1974), by the Izmiran fluxgate magnetometer, and by the energetic particle experiment of the Institute for Nuclear Physics of the Moscow State University, IKI and the Nuclear Research Center of Saclay (Kurt et al. 1979). The solar wind and energetic

* Paper presented at the Workshop on Acceleration of Particles by Shock Waves, October 7–9, 1980, Max-Planck-Institut für Aeronomie Lindau, Katlenburg-Lindau, Federal Republic of Germany

particle data from IMP-H and -J have been taken from the Solar Geophysical Data, Boulder, U.S.A. Other results used from PROGNOZ-6 observations can be found in the paper by Kurt et al. (1980), which discussed the 1-4 January 1978, event using energetic particle, plasma and magnetic field measurements provided by PROGNOZ-6.

The 1-4 January 1978 Particle Event

At the end of January 1, 1978, the energetic particle detectors on board HELIOS-1, HELIOS-2 and PROGNOZ-6 simultaneously detected an increase in the intensity of energetic protons. In Fig. 1 the overall time-intensity profiles as observed by these spacecraft are shown for protons in the energy range 4-13 MeV (HELIOS-1/-2) and 1.4-5.8 MeV (PROGNOZ-6) together with the times of the occurrence of a flare (F) on the sun and the observation of the flare-induced interplanetary shock wave (S). The profiles for HELIOS-1 and -2 are taken from the paper by Gombosi et al. (1981). At this time HELIOS-1 is located at 0.95 AU, 169.5° Carrington longitude and 1.7° Carrington latitude, while HELIOS-2 and PROGNOZ-6 are at 0.93 AU, 203.9° and -2.6° , and at 0.98 AU, 209° and -3.1° , respectively. Thus the maximum distances in interplanetary space covered by these three spacecraft are 0.05 AU in radial direction, about 40° in Carrington longitude and about 4.8° in Carrington latitude. These locations, in the ecliptic plane and relative to the site of the flare on the sun from which this event originated, together with the observed solar wind velocities, are shown in Fig. 2. Though HELIOS-1 and PROGNOZ-6 are separated by only 40° in longitude, the maximum intensity observed by HELIOS-1 is already reached about 6 h after the onset of the event, while PROGNOZ-6 observes a first maximum more than a day later after the onset. This result is at first sight in contradiction to the concept of a fast initial coronal propagation introduced by Reinhard and Wibberenz (1974).

We show in Fig. 3 the time profiles of the solar wind velocities for the same time period, as observed by the three spacecraft in question. In order to obtain the time histories of the solar wind observations at the Earth's orbit as completely as possible, we have added the IMP-J measurements to the PROGNOZ-6 observations by approximating them at those time intervals for which measurements of both spacecraft were available.

The corresponding time scales have been shifted in such a way that the times of the arrival of the flare-associated interplanetary shock wave (S) coincides for all three spacecraft. Note from this figure that during the onset of this event all three satellites are within a rather uniform, low-speed solar wind regime of about 380-430 km/s and, according to the magnetic field observations, of positive interplanetary magnetic field polarity. Taking these velocity values and the actual positions of the spacecraft into account, one can determine their relative locations on the sun by mapping them back along the average interplanetary magnetic field lines (Nolte and Roelof 1973). In Fig. 4 we have marked these three positions on part of the H_α synoptic chart of the Sun (Solar Geophysical Data, Number 407, Part II, 1978). This chart indicates the large-scale polarity regions of the solar magnetic field, the neutral lines, and the active regions. From the H_α solar flare observations one finds that the only active region that is located close enough to the three spacecraft' positions and that, at the same time, was active at the end of 1 January 1978, is located in the plage region 15081. On January 1 at about 21:53 UT, this region produced an H_α flare of importance 2N, which lasted from 21:45-22:16 UT and which was accompanied by solar II, III and IV radio emission,

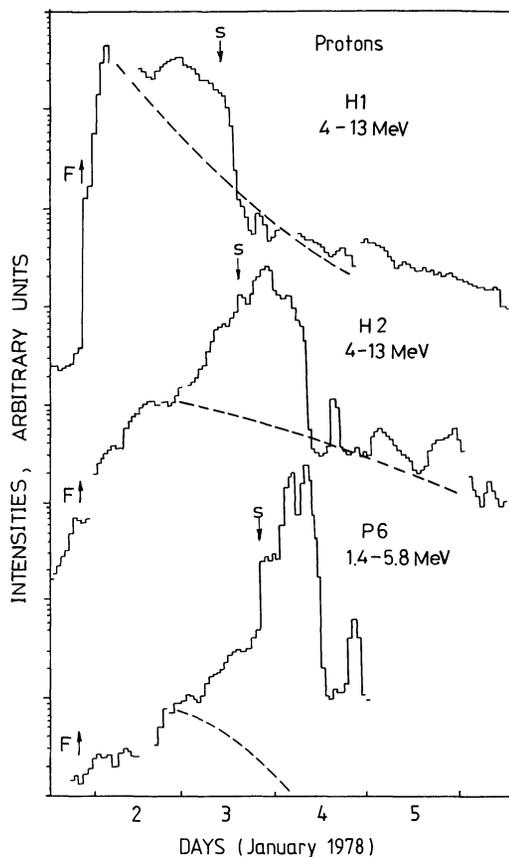


Fig. 1. Time-intensity profiles of the 1 January 1978, particle event occurring at time F on the sun, as observed by HELIOS-1 and -2 and PROGNOZ-6 in the energy intervals indicated (one-hour averages). The times of the arrival of the flare generated shock is marked by S . The result of Gombosi et al. (1981), who have modelled this event numerically, is shown as *dashed curves*

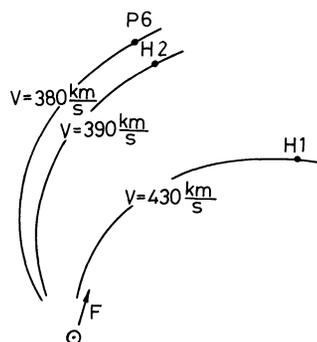


Fig. 2. Positions of HELIOS-1, HELIOS-2 and PROGNOZ-6 in the ecliptic plane and relative to the flare site F at the time the flare occurred on the sun. The values of the solar wind velocity, as measured by these three spacecraft at that time, are included

micro wave bursts and X-ray enhancements. Thus we may conclude that both the energetic particle events and the interplanetary shock waves observed by HELIOS-1/-2 and PROGNOZ-6 originated from this region. In addition we find from Fig. 4; first, that the large-scale solar polarity at the positions of the spacecraft coincides with the polarity of the interplanetary magnetic field, as observed by them; second, that at the time the 2N flare occurs HELIOS-1, HELIOS-2 and PROGNOZ-6 are located west relative to this flare, and that HELIOS-1 is well

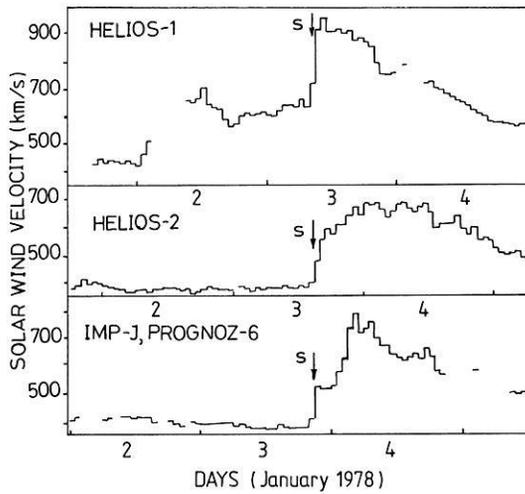


Fig. 3. Time profiles of one-hour averaged solar wind velocity values as observed by HELIOS-1, HELIOS-2 and PROGNOZ-6 during 2-5 January 1978. The passage of the flare generated shock is marked by S

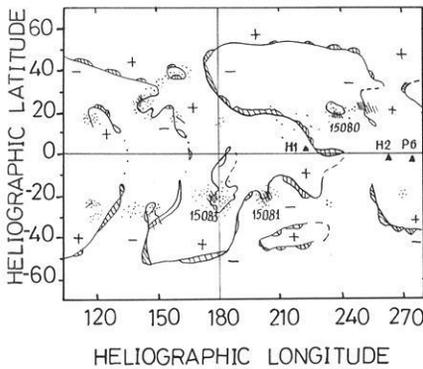


Fig. 4. Part of the H_{α} synoptic chart of the sun during December 1977-January 1978 showing the large-scale solar magnetic field polarity regions, the neutral lines, and the active regions. The projections of HELIOS-1 and -2 and PROGNOZ-6 at the time of flare onset in region 15081 are shown by triangles

separated in longitude from the positions of HELIOS-2 and PROGNOZ-6 and located much closer to the flare-site, even if one takes the approximate error of $\pm 10^{\circ}$ in Carrington longitude of the Nolte and Roelof technique into account; and third, that by a large, stable neutral line running from the northern solar hemisphere towards the solar equator (about 230° Carrington longitude) the region containing the flare site and the position of HELIOS-1 is separated in some way from the region containing the positions of HELIOS-2 and PROGNOZ-6. Taking these two latter results and accepted theories on coronal propagation of solar cosmic rays into account, one may well understand why HELIOS-1 observes such a fast increase and large maximum intensity, while HELIOS-2 and PROGNOZ-6 at the same time observe a longer-lasting increase and a smaller first maximum intensity.

At the beginning of 2 January the intensity of the 3-13 MeV protons, observed by HELIOS-1, starts to decrease again, while at HELIOS-2 and PROGNOZ-6 the corresponding intensities are still increasing. Despite the coronal effects mentioned above, we believe that the following interplanetary effect could in addition explain this feature: From Fig. 3 it follows that at this time HELIOS-1, in contrast to HELIOS-2 and PROGNOZ-6,

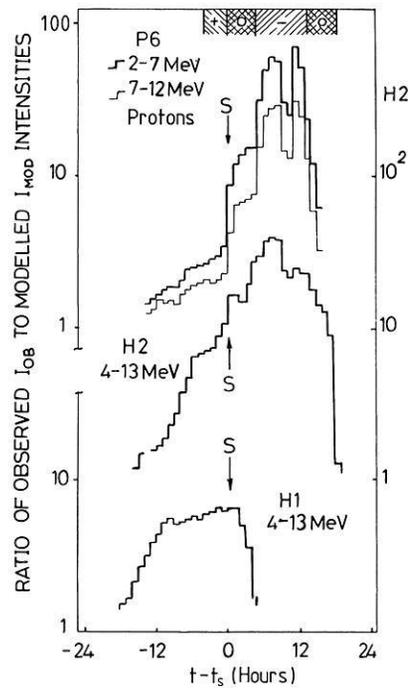


Fig. 5. Time profiles of the ratios of the observed energetic particle intensities to the modelled ones (see Fig. 1). Time in hours refers to the time t_s of the passage of the interplanetary shock S. In the upper part the angular direction of the energetic particle fluxes are shown, where + indicates a net streaming away from the sun, - towards the sun, and o an isotropic distribution

starts to move into a new and different solar wind regime of high-speed solar wind of more than 600 km/s and of negative interplanetary magnetic field polarity, while HELIOS-2 and PROGNOZ-6 remain in the low-speed solar wind of positive polarity. According to Fig. 4 this could mean that HELIOS-1 becomes disconnected from its previous region on the sun and connected by this high-speed solar wind to the large-scale solar region of negative polarity $+20^{\circ}$ north of the solar equator. For the following analysis it is important to note that the only active region in this new regime, plage region 15080, is not active between the end of 1 January and about 2:30 UT on 3 January, so that the second maximum observed by HELIOS-1 at the end of 2 January must be associated with the shock wave observed on 3 January, 8:39 UT, rather than with any solar active regions.

Taking coronal and interplanetary propagation effects into account, Gombosi et al. (1981) modelled, using a particular set of coronal and interplanetary propagation parameters, the time-intensity profiles of the 1 January event. Their results are shown as dashed curves in Fig. 1. By comparing the observed and modelled profiles one finds that there is an excess of energetic particles more or less strongly associated with the interplanetary shock waves observed by the three spaceprobes in question. In Fig. 5 we show these excess intensities where we plot the ratios of the observed (I_{ob}) and modelled (I_{mod}) intensities, respectively. In the following sections we shall concentrate on those parts of the event characterized by the following unique features: An increase of the particle intensity in front of the shock (pre-shock energetic storm particle (ESP) event) observed by all three spacecraft, an increase at the arrival of the shock (shock-spike event) observed by HELIOS-2 and PROGNOZ-6, an increase in the intensity after the shock wave (post-shock ESP event) again observed by HELIOS-2 and PROGNOZ-6,

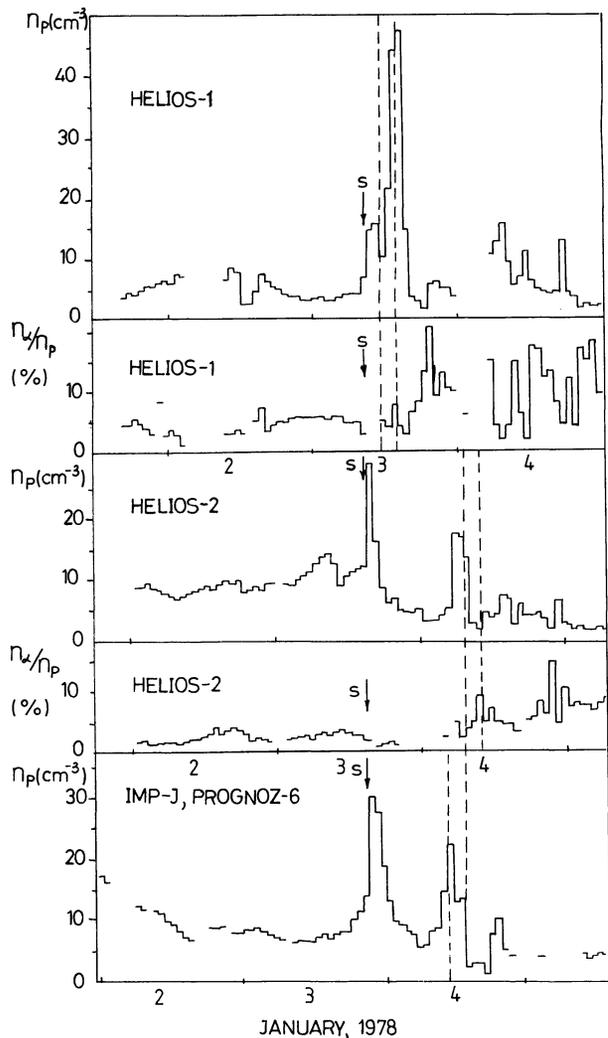


Fig. 6. Time profiles of one-hour averaged values of the solar wind number density N_p (cm^{-3}) and the alpha particle-to-proton ratio N_α/N_p (%), as observed by HELIOS-1 and -2 and by PROGNOZ-6 and IMP-J. The passage of the shock is indicated by *S*, and the times when the energetic particle intensity drops suddenly (see Fig. 1) by dashed lines

and a sharp decrease in the intensity observed by all three space-probes in the downstream region.

The 3 January 1978 Shock Event

The arrival of the interplanetary shock wave associated with the 2N flare on 1 January at 21:53 UT was observed by HELIOS-1, HELIOS-2 and PROGNOZ-6 on 3 January at 8:39 UT, 14:49 UT and 20:48 UT, respectively. According to Figure 3 this shock wave arrived at HELIOS-1 while this spacecraft was located in a high-speed solar wind regime, while it was observed by HELIOS-2 and PROGNOZ-6 at the time when both satellites were still in the low-speed solar wind region. Though HELIOS-1 is located by 0.02 AU further away from the sun than HELIOS-2, the most likely reason for the earlier detection of the shock by HELIOS-1 is most probably due to its propagation through a faster solar wind (Hirshberg et al. 1974).

In order to start with the large-scale features of this shock event we show in Fig. 6 one-hour-averages of the alpha particle-

to-proton ratios, N_α/N_p (%), and of the proton number densities, N_p (cm^{-3}), as observed by the plasma experiments on board of HELIOS-1 and 2, PROGNOZ-6 and IMP-J.

Note from this figure that at the shock, *S*, the density N_p increases rapidly to a first maximum in all three cases. In the post-shock regions, however, we find a second and overall maximum increase in N_p directly behind the shock in the case of HELIOS-1, whereas in case of PROGNOZ-6 and HELIOS-2 this second, yet smaller increase in N_p occurs about 15–16 h after the shock, respectively. In the case of HELIOS-1 and -2 these second increases in N_p are followed by an increase in the N_α/N_p ratio. In addition these increases are directly accompanied by large directional fluctuations in the interplanetary magnetic field, indicating some kind of ‘boundary’ between the plasma flow directly behind the shock wave and this alpha particle enriched solar wind region. Analyzing the magnetic field topology in the post-shock regime as observed by HELIOS-2 and IMP for this period of time, Burlaga et al. (1981) found that this ‘boundary’ is the front part of a magnetic cloud or magnetic loop, in which the observed magnetic field vectors rotate nearly parallel to a plane. The radial dimension of this cloud is approximately 0.5 AU. It should be noted that the occurrence of the front boundary of this magnetic cloud is directly associated with the rapid decreases in the intensities of energetic particles as depicted in Fig. 1 and as indicated by dashed lines in Fig. 6. This observation suggests that this cloud is a magnetically closed region preventing energetic particles being behind the shock to propagate further down into this post-shock solar wind region. An alternative suggestion concerning the dynamics of the flow behind the shock waves may be obtained from recent results of magnetohydrodynamic (MHD) models discussing the propagation of shocks in the interplanetary medium: D’Uston et al. (1981) have shown that this flow is often characterized by overexpansion marked by depressed densities and temperatures. Dryer et al. (1980) have demonstrated the closure of the magnetic field in at least one plane across these regions.

It has been proposed by Burlaga et al. (1981) that the shock wave might have been driven by this magnetic cloud closer to the sun, where this cloud may have moved faster. If this would hold, one could argue in view of Fig. 6 that the shock at the position of Helios-1 is still influenced by this magnetic cloud, as it is still attached to the shock (piston-driven shock wave?), whereas at the positions of HELIOS-2 and PROGNOZ-6 the same shock wave is well separated from this magnetic cloud by about 0.15 AU in radial distance and thus no longer influenced by this magnetic loop region (blast-wave type shock?). In order to test this hypothesis we applied the actual solar wind observations in the various regions downstream of the shock to the classification scheme set up by Hundhausen (1972). We find that in the case of HELIOS-1 the shock wave is an *R*-type shock, whereas in case of HELIOS-2 and PROGNOZ-6 it is an *F*-type shock wave. Thus, the 3 January 1978, shock event is another example to be added to the ones given by Ogilvie and Burlaga (1974), that a shock wave can exhibit the properties of blast and piston-driven shock waves at the same time, depending on the longitudinal position of its observation relative to the site of origin on the sun. In turn, we may conclude that the proposed classification of transient shock waves into *R*- and *F*-type shocks cannot be regarded as a unique scheme (see also Rosenau 1979), and it can not therefore be used to draw any conclusions on the origin of these shock waves.

In Fig. 7a the large-scale directional fluctuations perpendicular to the ecliptic plane for both the interplanetary magnetic field (δ_B) and the solar wind velocity (ϵ_p) vectors are depicted

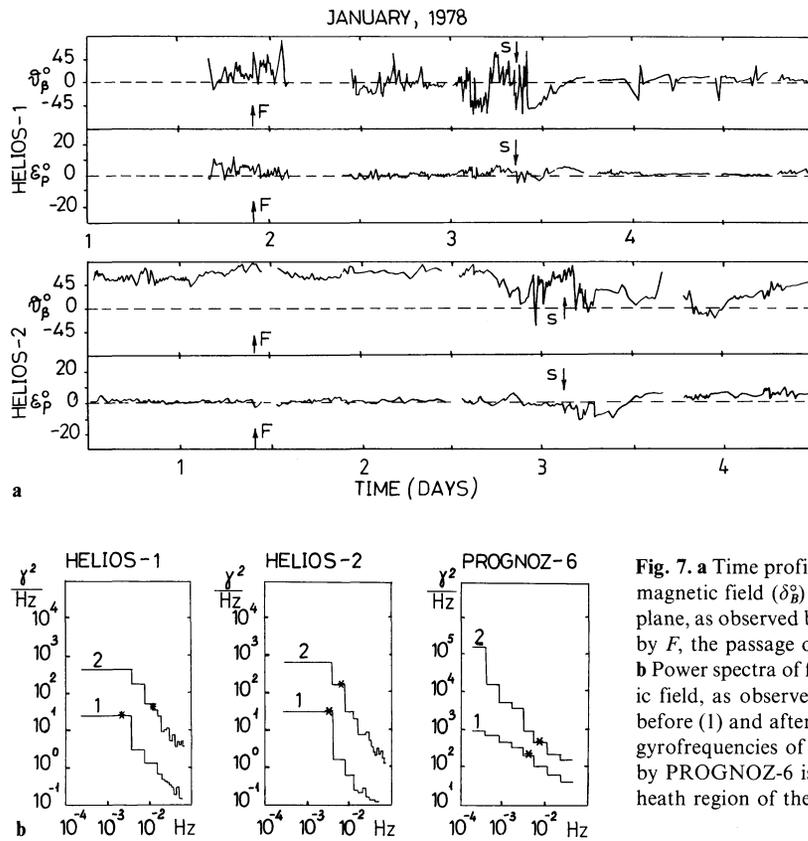


Fig. 7. a Time profiles of the fluctuations in the directions of the interplanetary magnetic field ($\delta\beta$) and of the solar wind bulk velocity (ϵ_p) out of the ecliptic plane, as observed by HELIOS-1 and -2. The occurrence of the flare is marked by *F*, the passage of the shock by *S*.

b Power spectra of fluctuations in the magnitude of the inter-planetary magnetic field, as observed by HELIOS-1 and -2 and PROGNOZ-6 in the regions before (1) and after (2) the shock wave. Asterisks indicate the doppler-shifted gyrofrequencies of 4 MeV protons in these two regions. The power observed by PROGNOZ-6 is in general higher, as the spacecraft was in the magnetosheath region of the earth at that time

for HELIOS-1 and HELIOS-2. From a systematic survey through more than 9 years of HELIOS' plasma and magnetic field data we found that the level of fluctuations of the angle ϵ_p clearly indicates the presence of transverse, Alfvénic type of fluctuations in the interplanetary medium. From Fig. 7a we find the following general features for these type of fluctuations: First, they are present in the pre-shock regimes for both HELIOS-1 and HELIOS-2. In case of HELIOS-1, however, they are much more pronounced and larger in amplitude. We found that this holds in general when comparing transverse fluctuations in high-speed solar wind streams with those in low-speed streams. Second, larger-amplitude fluctuations are observed throughout the entire pre-shock region (~ 1.5 days since the flare) in the case of HELIOS-1, but for only about 0.7 days before the shock in the case of HELIOS-2. Third, in the post-shock regions one now finds that these transverse fluctuations are of larger amplitude in the case of HELIOS-2 than in the case of HELIOS-1, where they vanish shortly after the shock.

In Fig. 7b the power-spectra of the fluctuations in the magnitude of the interplanetary magnetic field as observed by HELIOS-1 and -2 and PROGNOZ-6 are shown for the pre-shock (1) and the post-shock (2) regions, respectively. Note that the entire power-spectrum increases in all cases from the pre- to the post-shock region. However, taking the power at the doppler-shifted gyrofrequencies of 4 MeV protons into account (indicated by asterisks in Fig. 7b), we find that this increase in the case of HELIOS-1 is rather small, but nearly an order of magnitude in the case of HELIOS-2.

Finally, taking into account about 5 min averages of plasma and magnetic field parameters for HELIOS-1 and -2, directly in front of and behind the shock waves, and using them to optimize the jump conditions for MHD shocks by a kind of minimum difference technique (Lepping and Argentiero 1971),

Table 1. Some parameters for the shock waves observed by HELIOS-1 and -2

Parameter	HELIOS-1	HELIOS-2
Φ_{In}	-76.4°	33.6°
θ_n	-4.7°	3.7°
$\star (\mathbf{r}, \mathbf{n})$	76.4°	33.8°
$\star (\mathbf{B}, \mathbf{n})_1$	130.4°	5.6°
$\star (\mathbf{B}, \mathbf{n})_2$	-114.2°	-43.4°
$\star (\mathbf{V}, \mathbf{n})_1$	69.3°	3.9°
$\star (\mathbf{V}, \mathbf{n})_2$	56.0°	10.4°
$\star (\mathbf{B}_1, \mathbf{B}_2)$	115.4°	48.9°
ΔV_p (km/s)	157.0	64.1
ΔN_p (cm^{-3})	4.1	14.7
ΔB (nT)	5.1	2.7
V_s (km/s)	406.7	480.2

we can determine all parameters of the shocks and the changes of parameters across the shocks of interest. In Table 1 we have listed some of these values, where \mathbf{n} denotes the normal of the shock, \mathbf{r} the radial direction of the sun-spaceprobe line, \mathbf{B} the interplanetary magnetic field, V_p the solar wind velocity, N_p its number density, v_s the shock velocity, Δ the changes of solar wind parameters across the shock waves, and the index 1(2) the region before (after) the shock. Note, that for both HELIOS-1 and -2 the angles of the shock normals out of the ecliptic plane (θ_n) are rather small. In Fig. 8 we show the positions of the shocks of HELIOS-1 and -2, relative to the flare site on the sun, and the directions of their normals and surfaces

together with the overall stream structure and the post-shock boundary, as inferred from Fig. 3 and 6.

The Pre- and Post-Shock ESP Events

It has been shown by Scholer and Morfill (1975) that an increase in the flux of low energetic solar cosmic rays some hours before the arrival of interplanetary shock waves, the so called pre-shock energetic storm particle (ESP) events, can be explained by a gain of energy of these particles due to successive reflections at the moving interplanetary shock front by the first order Fermi process (shock acceleration with scattering in the pre-shock regime). As a result of their numerical model one finds: First, with a reasonable mean free path λ of ~ 0.07 AU for 1 MeV protons in front of the shock their intensity can increase by more than an order of magnitude in the pre-shock regime, followed by a pronounced drop of their intensity about an hour after the shock passage. Second, the number of reflections needed to account for this intensity increase is rather small (2–7 between the sun and 1 AU). Third, the maximum intensity of the ESP event depends roughly on λ in such a way that the smaller λ gets, the larger this intensity enhancement will be.

Comparing these theoretical results with the pre-shock intensity enhancements of the 4–13 MeV protons as observed by HELIOS-1 and -2 (see Fig. 1 and 5) we can develop the following general scenario: Modelling the overall intensity-time profiles of the 1–4 January event (Fig. 1), Gombosi et al. (1981) inferred an averaged interplanetary mean free path of the 4–13 MeV protons between the sun and about 1 AU of $\lambda = 0.10\text{--}0.15$ AU. As discussed above this value would be sufficient to account for an intensity enhancement in front of both the HELIOS-1 and -2 shock waves. In order to explain the differences, however, in these two ESP events, namely the location of their maxima relative to the shock fronts and the fact that the maximum intensity observed by HELIOS-2 is much larger than by HELIOS-1, we propose the following explanations: Taking into account the lengths in time (ΔL) of the occurrence of the larger-amplitude transverse fluctuations in the pre-shock regions between the flare onset and the shock arrivals at HELIOS-1 and -2 (Fig. 7a), we find a ratio of roughly $\Delta L(\text{HELIOS-1})/\Delta L(\text{HELIOS-2}) \approx 2.1$. On the other hand, taking the distances in time ($\Delta L'$) between the passages of the shock wave and the times when the intensities of the ESP events have decreased by a factor of about $1/e$ of the maximum intensities we find a ratio of roughly $\Delta L'(\text{HELIOS-1})/\Delta L'(\text{HELIOS-2}) \approx 2$. Comparing these two ratios, we think that the spread in time of the two pre-shock ESP events is controlled mainly by the spread of the observed larger-amplitude, transverse fluctuations in the interplanetary medium in the pre-shock regions. This would then mean that these local fluctuations have further reduced the mean-free path of the energetic particles in comparison with its averaged value given by Gombosi et al. (1981) at least in case of HELIOS-2.

Using the values given in Table 1, we can determine the angle θ between the upstream magnetic field and the shock surface to be $\theta \approx 40^\circ$ and $\theta \approx 84^\circ$ for HELIOS-1 and -2, respectively. As these shocks are oblique, the velocity increment of a particle through a single reflection at the shock will be of the order of $\sim 2 U_1 \sec \theta$ (Axford 1980, and references herein), where U_1 is the solar wind speed in the pre-shock regime in the rest frame of the shock. Taking the actual values for U_1 and θ into account we find that this increment is higher by a factor of about 4 in the case of HELIOS-2 than in the case of HELIOS-1, or, in other words, the shock at HELIOS-2 is four times more

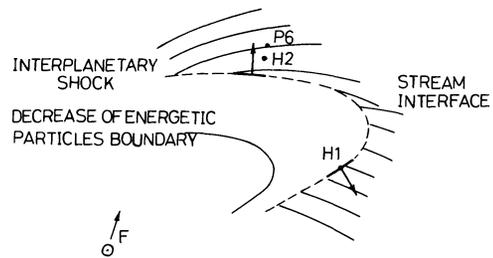


Fig. 8. Sketch of the positions and directions of the shock surfaces and of their normals at the time the interplanetary shock passes HELIOS-1 and -2, together with the overall stream structure of the interplanetary medium and the large-scale post-shock boundary, as inferred from Fig. 6

effective in accelerating particles by reflection than the shock at HELIOS-1. This would then mean, that locally the intensity of reflected particles at a certain energy would be larger at HELIOS-2 than at HELIOS-1, which we indeed observe according to Fig. 5. If this situation holds between the sun and the orbit of the HELIOS spaceprobes, then less than 4 (15) successive interactions of a solar wind particle with the shock front would increase its energy up to 4 MeV in the case of HELIOS-2 (HELIOS-1).

Studying now the post-shock regions, we are going to try to find a number of independent arguments, which could explain why, according to Fig. 1 or 5, the intensity of the 4–13 MeV protons in the post-shock regime of HELIOS-2 is larger than in case of HELIOS-1 by more than an order of magnitude:

- First, applying a very simple model of particle acceleration by plane shocks including convection and diffusion only (snowplough model) the change in the diffusion coefficient κ from the pre- (1) to the post-shock (2) region is given by $\kappa_2/\kappa_1 = (U_2/U_1)^2$, where U is the solar wind speed along the normal to the shock and in the rest frame of the shock (Vasiliev et al. 1980). Taking the actual observations into account, we find that in both cases κ will decrease from the upstream to the downstream region, but that this decrease is larger by a factor of about 2 in the case of HELIOS-2 than in case of HELIOS-1. Thus, particles leaving the shock to go into the post-shock regime can be trapped much more efficiently for HELIOS-2 than for HELIOS-1. According to Scholer and Morfill (1977), who have applied a full propagation model, including diffusion, convection and adiabatic deceleration, this trapping will lead to an increase in the particles' intensity roughly proportional to the decrease in κ .

- Second, still applying the snowplough model it has been shown (Axford, 1980, and references therein) that the increase in the number density behind the shock is directly proportional to the ratio U_1/U_2 . This ratio is larger by a factor 1.6 in the case of HELIOS-2 than in the case of HELIOS-1.

- Third, according to Fig. 7a, the amount of large-amplitude transverse fluctuations in the downstream regimes is by far larger in the case of HELIOS-2 than in the case of HELIOS-1. At the same time the level of longitudinal fluctuations at the doppler shifted gyrofrequency of 4 MeV particles is, according to Fig. 7b, again larger by nearly an order of magnitude for HELIOS-2 than for HELIOS-1. These two effects will therefore lead not only to a further decrease of κ_2 for HELIOS-2 and therefore to a further increase of the trapped particle population, but at the same time to an additional turbulent acceleration of particles due to their interaction with random Alfvén and sound waves (second order Fermi mechanism; Toptygin 1980). That this latter effect does indeed occur, can be seen directly

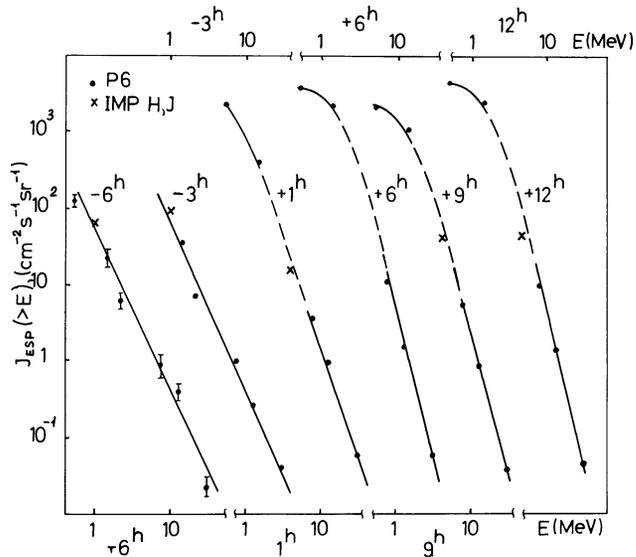


Fig. 9. Spectra of integral intensities vs energy as observed by PROGNOZ-6 and IMP-H, -J from -6 h before to +12 h after the shock

from Fig. 9, where we show the energy spectra of energetic particles ≥ 0.5 MeV, as observed by PROGNOZ-6 and IMP-J (which are very close to HELIOS-2, as seen from Fig. 8) from -6 h before the shock to +12 h after the shock. From this figure it follows readily that the power-law pre-shock spectrum steepens and is transformed into an exponential spectrum in the post-shock regime in accordance with the results of Toptygin (1980). This is a general feature often observed in cases where particles are accelerated by turbulent wave-fields.

The Shock-Spike Event

In addition to the intensity enhancements of energetic particles in the up- and downstream regions of the interplanetary shock wave, we have already noted from Figure 1 that there is an increase of energetic particle fluxes directly associated with the passage of this shock, the so-called shock-spike event. This situation becomes more obvious if we use a higher time resolution and a lower energy window. For ≥ 5 min averages, and for 1.4–5.8 MeV protons or ≥ 80 keV (≥ 87 keV) ions as observed by PROGNOZ-6 or HELIOS-1 (HELIOS-2), respectively, we can clearly identify this event in Fig. 10.

These observations seem to contradict those accepted theoretical models on the origin of shock-spike events that are based on the shock drift acceleration mechanism (Armstrong et al. 1977; Pesses et al. 1980, and references therein). Within these models the shock itself is regarded as the prime particle accelerator, and the fluctuations of the interplanetary magnetic field in the up- and down-stream regions of the shock are, to first order, ignored. The actual acceleration of low-energy particles to higher energies is then provided by the following, more or less instantaneously operating, non-Fermi mechanism. Positively charged particles interacting with fast-mode shock waves experience a guiding-center drift motion in the shock and are accelerated at the same time by the $\mathbf{V}_p \times \mathbf{B}$ induced electric field being due to the motion of the up- and down-stream solar wind. It has been shown that the efficiency of this mechanism depends crucially on the angle θ between the directions of the pre-shock magnetic field \mathbf{B}_1 and the shock surface. For angles $\theta > 10^\circ$ the energy gain should be so small that no noticeable shock-spike

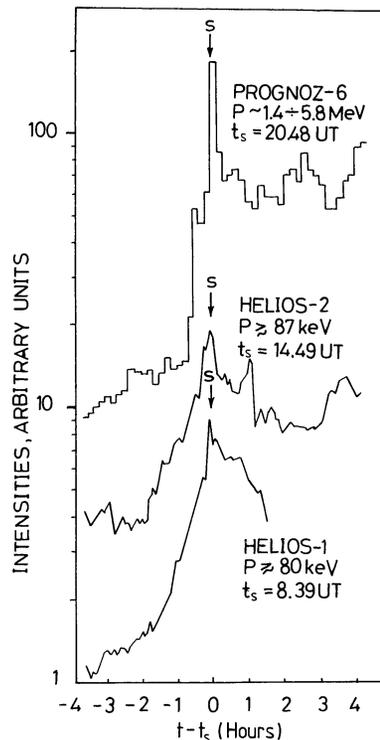


Fig. 10. Time profiles of the low-energetic particle enhancements associated with the passage of the interplanetary shock wave (S), as observed by HELIOS-1 and -2 and PROGNOZ-6 at the energies indicated. About ≥ 5 min averages are used. Time t_s refers to the arrival of the shock

intensity enhancement should occur. Yet, though $\theta \approx 40^\circ$ and 84° in the case of HELIOS-1 and -2, respectively, we do find rather pronounced intensity increases in association with the shock wave (Fig. 10).

However, as already mentioned by these authors their result, which was obtained for weak, planar and thin shocks without fluctuations in the up- and downstream medium, must be revised if real, observed interplanetary shock waves are considered: The drift of the pre-shock magnetic field lines, around which the particles gyrate, into the shock and the angular change between the upstream and the downstream magnetic field directions may be different to that considered in these model calculations. The shock may be thick and turbulent, and there may be large directional fluctuations of the interplanetary magnetic field just in front of and behind the shock. All these effects could prevent a particle once in the shock from scattering out of this region either into the pre-shock (reflection) or into the post-shock (transmission) regime very easily. Thus, as these processes would increase the possibility of a multiple crossing of the shock, a weaker shock-induced electric field could accelerate these particles to higher energies. Introducing only small angle fluctuations of $\Delta\theta \leq 10^\circ$ of the pre-shock magnetic field, Armstrong et al. (1977) showed that the maximum energy obtained by particle-shock interaction increases by a factor of roughly 2. In addition, high-frequency directional fluctuations just in front of the shock could temporarily change the angle θ to a value which is favorable for acceleration of particles via the shock drift mechanism.

In Fig. 11 we have plotted the distributions of the angles θ between the directions of the interplanetary magnetic field actually observed by HELIOS-1 and -2 in the pre- and post-shock regions (40.5 s averaged values) and the averaged directions of the shock surfaces. In addition we have marked the $\theta \leq 10^\circ$ line

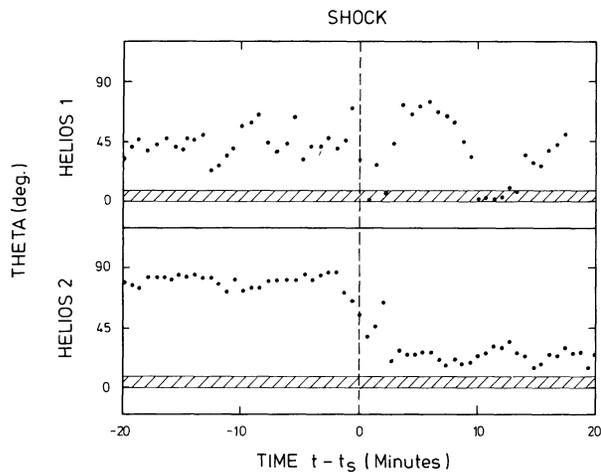


Fig. 11. Distributions of 40.5 s averaged values of the angle θ between the instantaneous direction of the interplanetary magnetic field and the shock surfaces, as observed by HELIOS-1 and -2 about 20 min before and after the passage of the shock wave (S). The shaded regions for θ are regions for which the acceleration mechanism according to Armstrong et al. (1977) is applicable

of preferred acceleration according to Armstrong et al. (1977). From this figure the following results can be obtained: First, in the regions of a few minutes before and after the shock we find, in the case of both HELIOS-1 and -2, rather large fluctuations in the values of the angle θ . For HELIOS-1 (HELIOS-2) θ ranges from about 30° – 65° (55° – 87°) in the upstream and from about 0° – 75° (20° – 64°) in the downstream region, or $\Delta\theta$ from about 10° – 25° (3° – 30°) or 35° – 40° (20° – 64°). Second, directly across the shock the angle θ decreases rather abruptly to a value of 0.06° in the case of HELIOS-1. Thus, this effect could explain the occurrence of a shock-spike intensity enhancement for at least the low energetic particles >80 keV. Third, in the case of HELIOS-2, however, the angle θ never decreases below about 35° , though a clear shock-spike intensity enhancement of not only the low (≥ 87 keV) but also of the higher energetic particles (≥ 4 MeV) is observed. Thus, although still in contradiction with theory we have to conclude that highly oblique shocks can also produce shock-spike events, and that new and refined models are needed to explain such a phenomenon which have to take into account large-amplitude directional fluctuations of the pre- and post-shock magnetic field and therefore enhanced scattering of particles (Fisk and Lee 1980, and references therein).

Conclusions

In this report we have studied the different characteristics of the interplanetary medium and of energetic particle fluxes at different energies in the regions before, at and after one interplanetary shock wave, observed on 3 January 1978 simultaneously by three different spacecraft, near the Earth's orbit but at different longitudinal positions about 40° apart. In order to explain the observed intensity increases of energetic particles in these three regions on the basis of current theories on particle acceleration by propagating shock waves, we found:

First, that really none of them can be explained uniquely on the basis of the shock drift acceleration mechanism only, where the particles interact with the shock, experience a grad \mathbf{B} drift in the shock, are accelerated at the same time by the shock-induced electric field, and then leave the shock to go

into the up- or downstream region, respectively (Armstrong et al. 1977, and references therein). Even taking the observed highly time-resolved fluctuations of the angle θ between the upstream magnetic field direction and the shock surface into account, it has been shown that θ never shows small enough values of below about 5° , that are necessary for this acceleration mechanism to work effectively. Thus, only those models taking also into account the observed large-amplitude fluctuations in direction and in magnitude of the interplanetary magnetic field in the up- and downstream regions may explain the observed features of the energetic particle fluxes in connection with this shock wave (acceleration with scattering).

Second, concentrating on the pre-shock regime we have shown that the shock wave observed by HELIOS-1 and -2 is able to accelerate particles rather effectively by reflection (especially in case of HELIOS-2). As the average mean free path in this region is rather small ($\lambda \leq 0.08$ AU), and as it may be further reduced locally by the larger-amplitude directional fluctuations observed directly in front of the shock (Fig. 7a), the pre-shock intensity enhancements of energetic particles can be explained by a cumulative first order Fermi process of successive reflections of particles at the shock in the entire region between the sun and the Earth (Scholer and Morfill 1975). This idea is well supported by the following two facts: First, only about four successive reflections over the range of 1 AU would be sufficient to accelerate solar wind particles up to 4 MeV in case of HELIOS-2, and second the widths in time of the intensity enhancements observed by HELIOS-1 and -2 correspond directly with the widths in time of the upstream larger-amplitude directional, Alfvénic type of fluctuations of the interplanetary medium.

Third, going then to the post-shock region and comparing the intensities of energetic particles as observed by HELIOS-1 and -2, we find a huge increase in the case of HELIOS-2 but nearly no effect in the case of HELIOS-1. By inter-comparison of the corresponding plasma and magnetic field observations we can explain this feature by different mechanisms working in the same direction: First, there is a further and larger reduction of the mean free path in the case of HELIOS-2 compared to HELIOS-1, so that transmitted solar flare and/or shock accelerated particles can be trapped more easily in the case of HELIOS-2. This will in general lead to an increase in the particles' intensity (Scholer and Morfill 1977). Second, there are large-amplitude directional fluctuations present in case of HELIOS-2, but almost none in the case of HELIOS-1. Thus in the case of HELIOS-2 an additional first order Fermi acceleration mechanism will work in the downstream medium. Third, there are longitudinal fluctuations present in the post-shock regime, which are more than an order of magnitude more powerful in the case of HELIOS-2 than in the case of HELIOS-1. Thus, a second order Fermi acceleration mechanism will work more effectively in the case of HELIOS-2, due to particle interactions with random Alfvén wave and sound wave turbulences (Toptygin 1980). The bending of the energy spectrum of energetic particles from the post- to the pre-shock region is regarded as evidence that this mechanism works in the case of HELIOS-2 or PROGNOZ-6.

Fourth, in the case of HELIOS-1 and -2 and PROGNOZ-6 there are increases in the intensities of energetic particles directly associated with the passage of the shock wave. In the case of HELIOS-2 and PROGNOZ-6 these increases can also be observed for particles ≥ 1 MeV. The observation by HELIOS-2 at least stands, at present, in some opposition to the theory on the origin of shock-spike events, if explained by the shock

drift mechanism alone. Thus, we can conclude that highly oblique shocks can also produce shock-spike intensity enhancements of energetic particles, and that new and refined models have to be developed to explain this feature.

Fifth, the classification of interplanetary shock waves into *R*- and *F*-events introduced by Hundhausen (1972), i.e., into events for which the post-shock energy flux of the solar wind rises or falls, is not a 'unique' scheme. In this report we have shown explicitly that one and the same shock wave can be observed as an *R*-event (HELIOS-1) and as an *F*-event (HELIOS-2) at the same time. The possibility of this effect was first mentioned by Dryer (1975).

Acknowledgements. Part of this work was done while one of us (A.K.R.) was a guest at the Space Research Institute in Moscow. He wants to thank the director and all colleagues of this institute for their great hospitality. V.G.K. and V.G.S. want to thank Yu. I. Logachev for useful discussions. Finally, all authors want to thank H. Kunow and G. Wibberenz for providing the energetic particle data from HELIOS-1 and -2, as depicted in Figure 1, prior to publication.

References

- Armstrong, T.P., Chen, G., Sarris, E.T., Krimigis, S.M.: Acceleration and modulation of electrons and ions by propagating interplanetary shocks. In: Study of Travelling Interplanetary Phenomena. Holland: D. Reidel Publishing Company 1977
- Axford, W.I.: The acceleration of galactic cosmic rays. IAU/IUPAP Symposium on Origin of Cosmic Rays, Bologna 1980
- Burlaga, L., Lepping, R., Weber, R., Armstrong, T., Goodrich, C., Sullivan, J., Gurnett, D., Kellogg, P., Keppler, E., Mariani, F., Neubauer, F., Rosenbauer, H., Schwenn, R.: Interplanetary particles and fields, November 22 to December 6, 1977: HELIOS, VOYAGER and IMP observations between 0.6 and 1.6 AU. *J. Geophys. Res.* **85**, 2227–2242, 1980
- Burlaga, L., Sittler, E., Mariani, F., Schwenn, R.: Magnetic loop behind an interplanetary shock: VOYAGER, HELIOS, and IMP-8 observations. *J. Geophys. Res.* **86**, 6673–6684, 1981
- Dryer, M.: Interplanetary shock waves: Recent developments. *Space Sci. Rev.* **17**, 277–325, 1975
- Dryer, M., Wu, S.T., Han, S.M.: Two-dimensional, time-dependent MHD simulation of the disturbed solar wind due to representative flare-generated and coronal hole-generated disturbances. *Geophysica Int.* **19**, 1–15, 1980
- D'Uston, C., Dryer, M., Han, S.M., Wu, S.T.: Spatial structure of flare-associated perturbations in the solar wind simulated by a two-dimensional numerical MHD model. *J. Geophys. Res.* **86**, 525–534, 1981
- Fisk, L.A., Lee, M.A.: Shock acceleration of energetic particles in corotating interaction regions in the solar wind. *Astrophys. J.* **237**, 620–626, 1980
- Gombosi, T.I., Green, G., Kecskemety, K., Kunow, H., Kurt, V.G., Logachev, Yu.L., Somogyi, A.J., Stolpovskii, V.G., Wibberenz, G.: Rigidity dependent coronal transport of energetic particles during the January 2–4, 1978 event. 17th Int. Cosmic Ray Conf., Paris, 1981
- Gringauz, K.I., Bezrukhikh, V.V., Volkov, G.I., Verigin, M.I., Davitaev, L.N., Kopylov, V.F., Musatov, L.S., Shoutchenkov, G.F.: Study of solar plasma near Mars and along the Earth to Mars path by means of charged particle traps aboard the Soviet spacecraft launched in 1971–1973. I. Techniques and devices. *Kosmicheskii Issledovaniya* **12**, 430–436, 1974
- Hirshberg, J., Nakagawa, Y., Welck, R.E.: Propagation of sudden disturbances through a nonhomogeneous solar wind. *J. Geophys. Res.* **79**, 3726–3730, 1974
- Hundhausen, A.J., Coronal Expansion and Solar Wind. Heidelberg: Springer-Verlag 1972
- Kurt, V.G., Logachev, Yu.I., Pisarenko, N.F., Stolpovsky, V.G.: Energetic spectra in solar flare measured by PROGNOZ-6. *Izvestiya Akademii Nauk SSSR* **43**, 3519–3523, 1979
- Kurt, V.G., Stolpovskii, V.G., Gombosi, T.I., Kecskemety, K., Somogyi, A.J., Gringauz, K.I., Kotova, G.A., Verigin, M.I., Styazhkin, V.A.: Energetic particle, solar wind plasma and magnetic field measurements on board PROGNOZ-6 during the large scale interplanetary disturbance of January 3–4, 1978, Preprint KFKI-1980-32, Central Research Institute for Physics, Hungarian Academy of Sciences, 1980
- Lepping, R.P., Argentiero, P.D.: Single spacecraft method of estimating shock normals. *J. Geophys. Res.* **76**, 4349–4359, 1971
- Nolte, J.T., Roelof, E.C.: Large-scale structure of the interplanetary medium. I. High coronal source longitude of the quiet-time solar wind. *Solar Phys.* **33**, 241–257, 1973
- Ogilvie, K.W., Burlaga, L.: A discussion of interplanetary postshock flows with two examples. *J. Geophys. Res.* **79**, 2324–2330, 1974
- Pesses, M.E., Decker, R.B., Armstrong, T.P.: The acceleration of charged particles in interplanetary shock waves. Johns Hopkins University Preprint JHU/APL 80-10, 1980
- Reinhard, R., Wibberenz, G.: Propagation of flare protons in the solar atmosphere. *Solar Phys.* **36**, 473–494, 1974
- Rosenau, P.: Propagation of two-fluid interplanetary shock waves. *J. Geophys. Res.* **84**, 5897–5906, 1979
- Scholer, M., Morfill, G.E.: Simulation of solar flare particle interaction with interplanetary shock waves. *Solar Phys.* **45**, 227–240, 1975
- Scholer, M., Morfill, G.E.: Modulation of energetic solar particle fluxes by interplanetary shock waves. Proceedings of COSPAR Symposium B, AFGL-TR-77-0309, 1977
- Toptyghin, I.N.: Acceleration of particles by shocks in a cosmic plasma. *Space Sci. Rev.* **26**, 157–213, 1980
- Vasiljev, V.N., Toptyghin, I.N., Chirkov, A.G.: Particle acceleration at shock waves in turbulent media. *Kosmicheskaya Issledovaniya* **18**, 556–564, 1980

Received March 6, 1981; Revised version September 29, 1981

Accepted September 30, 1981

Monte Carlo Simulation of Collisionless Shocks Showing Preferential Acceleration of High A/Z Particles*

Donald C. Ellison**¹, Frank C. Jones¹, and David Eichler***²

¹ Laboratory for High Energy Astrophysics, NASA Goddard Space Flight Center, Code 660, Greenbelt, Maryland 20071, USA

² Astronomy Program, University of Maryland, College Park, Maryland 20742, USA

Abstract. A collisionless, quasi-parallel shock is simulated using Monte Carlo techniques. In this kinetic approach, scattering of all velocity particles (from thermal to high energy) is assumed to occur such that the mean free path is directly proportional to velocity times the mass to charge ratio and inversely proportional to the plasma density. Within the constraints of this assumption, the shock profile and velocity spectra are obtained showing preferential acceleration of high A/Z particles relative to protons.

Key Words: Cosmic rays – Shock acceleration

Introduction

Analytic treatments of the acceleration of charged particles by collisionless shocks have been successful in describing some of the basic properties of this mechanism. The work of Krymsky (1977), Axford et al. (1977), Bell (1978), and Blandford and Ostriker (1978) shows that steady state shocks can yield high energy power law spectra similar to that observed in many astrophysical environments. Exponential spectra, as observed within the solar system, can also be derived when adiabatic losses (Fisk and Lee 1980) or finite shock size (Scholer et al. 1980; Terasawa 1981; Eichler 1981) are considered. Also, Axford et al. (1977) and Drury and Volk (1981) consider cosmic ray mediation and derive solutions in which sufficient cosmic ray pressure prevents a shock, with the bulk kinetic energy going into cosmic rays rather than heat.

In the above treatments, the number flux of cosmic rays is generally conserved through the shock; that is, the cosmic rays are treated as a second fluid component, with the question of their origin left open. Bell (1978) proposes that the shock picks up a few superthermal particles from the shocked fluid, but the rate of this process remains a free parameter in his discussion. Our own work, which models production of cosmic rays at the shock directly from the thermal plasma, is motivated by the following considerations. Observations suggest that shocks routinely put a significant fraction of their energy into cosmic rays in addition to shock heating the thermal fluid, and this requires that the number of “injected”

particles, if it is a free parameter, be chosen very carefully. If there are too few, their energy content subsequent to acceleration remains small. If there are too many, the thermal shock is washed out, and the thermal fluid is not shock heated, contrary to observations of supernova remnants etc. Other arguments against pre-injection, in the context of galactic cosmic rays, have been made on observational grounds by Eichler (1980), Cowsik (1980), and Fransson and Epstein (1980).

The physical details of injection, that is, if and how a shock creates new cosmic rays from the thermal plasma, depends on the shock structure. A model of cosmic ray production at shock fronts was proposed by Eichler (1979). The “injection rate” of thermal particles into the cosmic ray pool is fixed by the assumptions of the model, i.e. that thermal particles are themselves compressed by converging scattering centers, (e.g. as in the shock model of Parker (1961)), and that their acceleration is limited only by the presence of higher energy particles. The resulting injection rate is just what is needed to maintain a dynamically significant level of cosmic rays.

The premise that any particle in the upstream thermal plasma is about as likely as any other particle of the same ion species to end up as a cosmic ray suggests that the relative composition of cosmic rays can be calculated, in principle, by solving for the shock structure. Conversely, measurements of cosmic ray composition can (again, in principle) provide observational tests of shock models that feature cosmic ray production.

In treating first-order Fermi acceleration of cosmic rays (see, in particular, Axford et al. 1977 and Blandford and Ostriker 1978), the procedure generally followed is to solve a diffusion equation namely,

$$\frac{\partial}{\partial x} \left(uf - k \frac{\partial f}{\partial x} \right) = \frac{1}{3} \left(\frac{\partial u}{\partial x} \right) \frac{\partial}{\partial p} (pf) \quad (1)$$

where u is the plasma flow velocity, $f(x, |\vec{p}|)$ is the cosmic ray distribution function, x is the position measured in the shock frame, p is the momentum and k is the diffusion coefficient. This diffusion approximation holds only when the thermal velocity of the particles is large compared to the flow velocity. In addition, one or more of the following simplifying assumptions are made: (1) treating the shock as a discontinuity (that is, neglecting cosmic ray back pressure on the incoming plasma), (2) ignoring or simplifying energy dependence of the diffusion coefficient, and (3) assuming an infinite

* Presented, in part, at MPAE-Lindau Workshop on Acceleration of Particles by Shock Waves, October 1980

** Also at The Catholic University of America, Washington D.C.

*** Alfred P. Sloan Fellow

scale to the accelerating system. The difficulties in achieving an analytic solution and the inability of the diffusion approximation to describe the behavior of low energy particles have led us to attempt a Monte Carlo simulation of this problem. This kinetic approach can describe the behavior of thermal particles *if* the proper scattering law is known and, therefore, can model how high energy particles might be drawn from the thermal pool.

Description of Monte Carlo Method

In this simulation of a parallel shock we follow the trajectories of particles originating far upstream in the thermal plasma, allowing them to scatter *isotropically* and *elastically* in the frame of the massive scattering centers. The particles gain energy by scattering off converging flows and are observed far downstream as they leave the system. We assume that the scattering of particles of all velocities (including the thermal particles responsible for creating the collisionless shock) can be represented by a mean free path, λ , given by

$$(1/3) \lambda v = k_0 (v/c) R^\alpha \quad (2)$$

where $R = \text{rigidity} = pc/(Ze)$, v is the particle velocity measured in the plasma frame, α is a constant (Jones 1978), c is the velocity of light, e is the electronic charge, Z is the charge number and k_0 is assumed to be inversely proportional to the plasma density, ρ . Equation (2) can be written (for non-relativistic particles) as

$$\lambda = \lambda_0 (A/Z)^\alpha (v/u_2)^\alpha (\rho_2/\rho) \quad (3)$$

where A/Z is the mass to charge ratio, u_2 is the downstream flow velocity, and ρ_2 is the downstream plasma density. This assumption determines the shock structure as the bulk motion is randomized by isotropic scattering.

Our calculations are non-relativistic and for this preliminary work we consider only cases where $\alpha=1$. This choice simplifies computation since it implies a constant collision time and has some observational justification in that a diffusion coefficient of approximately this form has been deduced from observations of diffuse ions upstream of the earth's bow shock (see Scholer et al. 1980; Forman 1981; Ellison 1981).

The size of the system is defined by a free escape boundary located a distance d upstream from the shock (defined as $x=0$).

Results

Discontinuous Shock

As an illustration we present results where the shock is taken to be discontinuous. This is done for two injection velocities for a strong shock ($u_1=4u_2$, see Fig. 1) where the flow velocity, $u(x)$, is given by

$$u(x) = u_1, \quad x < 0 \\ u(x) = u_2, \quad x \geq 0.$$

In this first case, the velocity of the injected particles, v_i , is large compared to u_2 . Figure 2 shows the integral spectra obtained for $v_i=10u_2$ (the free escape distances are measured in λ_0 , the mean free path of a particle with $A/Z=1$, $v=u_2$, and $\rho=\rho_2$).

The expected slope for an integral, non-relativistic velocity spectrum is given by (Axford et al. 1977 and Bell 1978)

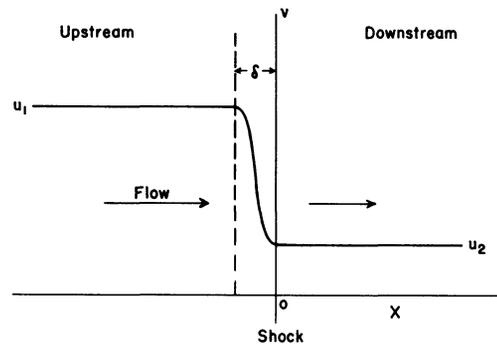


Fig. 1. Schematic velocity profile of a strong collisionless shock. In discontinuous shocks, δ is small compared to the mean free path of the accelerated particles.

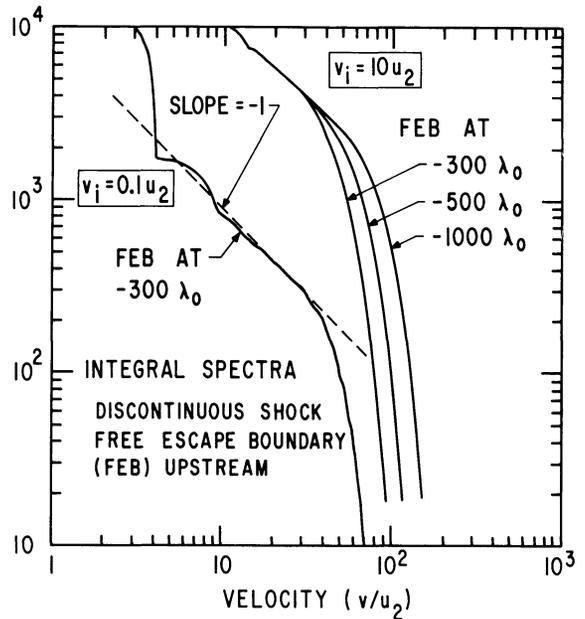


Fig. 2. Integral velocity spectra for a discontinuous shock with $v_i=0.1$ and $10u_2$. The dashed line shows the expected slope obtained from the diffusion approximation. The position of the free escape boundary is measured in units of λ_0 , the mean free path of a particle with $A/Z=1$, $v=u_2$, and $\rho=\rho_2$.

$$\text{slope} = -\frac{3u_2}{u_1 - u_2} \\ = -1 \quad \text{for } u_1/u_2 = 4. \quad (4)$$

The effect of the finite scale of the accelerating system is seen in the cutoff that occurs when

$$\frac{u_2 d}{k} \sim 1. \quad (5)$$

Significant differences from the predicted power law occur when the injected velocity is of the order or smaller than the flow velocities. Figure 2 shows the result for $v_i=0.1u_2$. Above about $10u_2$, the expected power law and cutoff occur but for low velocities we see the "heating" effect that results when all particles make their first crossing of the shock. After one crossing, the injected particles will have a velocity measured in the downstream plasma frame given by

$$v = \sqrt{[v_i \cos \theta + (u_1 - u_2)]^2 + (v_i \sin \theta)^2} \quad (6)$$

where θ is the angle the particle makes with the shock normal as it crosses the shock. It is seen that even if v_i is zero, particles will acquire a velocity boost of $u_1 - u_2$, consistent with the Rankin-Hugoniot relations for an initially cold plasma. This means there will always be particles downstream with velocities greater than u_2 and, therefore, a definite proportion will diffuse back upstream across the shock and participate in the first order acceleration. The high energy particles in the power law spectrum are drawn naturally from the thermal population.

Particle Smoothed Shock

The Rankin-Hugoniot relations for particle, momentum, and energy flux are respectively,

$$\rho(x)u(x) = \text{constant} = A$$

$$P(x) + \rho(x)u(x)^2 = \text{constant} = B$$

$$\rho(x)u(x) \left[\frac{1}{2}u(x)^2 + \frac{\gamma P(x)}{\rho(x)(\gamma - 1)} \right] = \text{constant} \quad (7)$$

where P is the x -component of pressure measured in the plasma frame, u the plasma flow velocity, and γ the ratio of specific heats (taken in what follows to be $5/3$). The first two of these equations can be combined to give

$$u(x) = [B - P(x)]/A. \quad (8)$$

In our simulation we choose an initial velocity profile $u(x)$ and then calculate A , B and $P(x)$. A new velocity profile is then calculated using Eq. (8) and the process repeated until the particle and momentum fluxes are approximately constant across the shock.

For any real shock acceleration process, the predicted power law slope (Eq. 4) will not be obtained at very high energies. That this must be the case is seen in the fact that Eq. 4, if it describes the spectrum to unlimited energies, places an infinite amount of energy in the cosmic ray tail. To avoid this, particles will either escape from the system by geometric effects from a finite shock or from a lack of scattering at some sufficiently high energy, the time of acceleration will become longer than the lifetime of the shock, or the shock will be smoothed to the point that it no longer produces a slope of -1 at the highest energies. If particles escape, this loss of energy will in effect soften the equation of state of the gas and allow a compression ratio greater than 4. Equations (7), when modified for particle escape give the compression ratio r ,

$$r = \frac{5 + \sqrt{25 - 16[1 - (2q/\rho_1 u_1^3)]}}{2[1 - (2q/\rho_1 u_1^3)]} \quad (9)$$

where q is the rate of energy loss from the high energy particles leaving the system and the subscripts indicate values far upstream. We have assumed a high Mach number shock and that the flux of particles leaving the system is small. In these preliminary calculations we include a velocity cutoff above which particles escape freely from the system but neglect any possible effect on the compression ratio.

Figure 3 shows the resultant velocity profile for $r=4$, an injection velocity of $v_i = 0.07 u_2$, $A/Z=1$, an infinite scale, and a velocity cutoff of $50 u_2$. As can be seen from the figure, the transition takes place on essentially two length scales. There is an abrupt transition caused by the large number of low velocity particles on a scale of order λ_0 , and a large scale

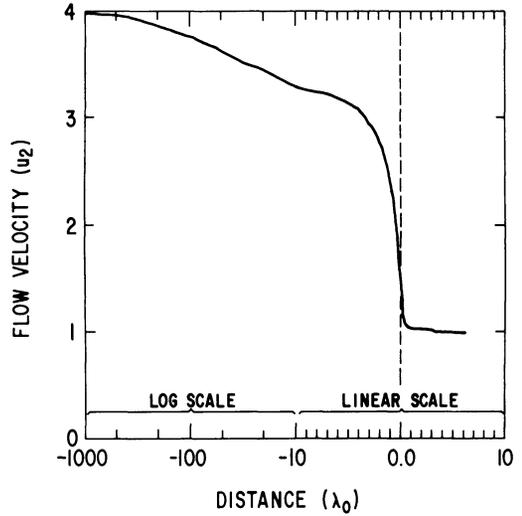


Fig. 3. Upstream velocity profile obtained by demanding that the particle and momentum fluxes be conserved (within 20% of the far upstream values). The free escape boundary is at infinity and a velocity cutoff at $v = 50 u_2$ is imposed.

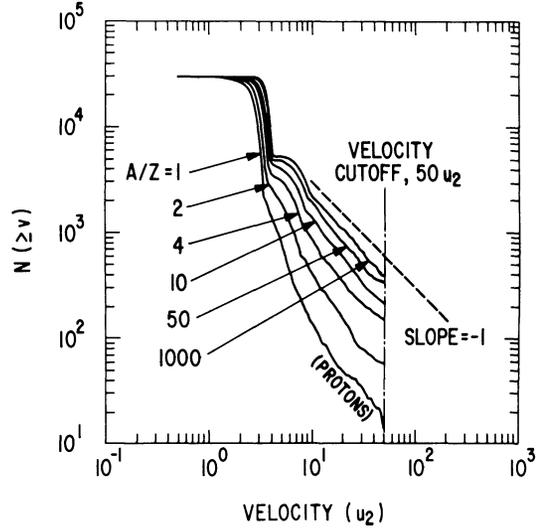


Fig. 4. Integral velocity spectra for different A/Z particles with $v_i = 0.07 u_2$ using the velocity profile shown in Fig. 3. The free escape boundary is at infinity and a velocity cutoff at $v = 50 u_2$ is imposed.

smoothing caused by the few high velocity particles that stream far out in front of the shock. The relative importance of this high velocity tail depends on the proportion of thermal particles that are accelerated to high energies which in turn depends on the shape and position of the shoulder in Fig. 2. It also depends on the velocity cutoff and indeed, the profile will be smoothed farther upstream as the cutoff is raised and higher energy particles are accelerated.

These results must be considered suggestive rather than quantitative, but it is clear that if the shock profile is smoothed as described above and the diffusion coefficient increases with energy, the spectrum will not be a strict power law but will be flatter at high energies than at low.

Enhancement of High A/Z Particles

The combination of a finite scale velocity transition and a mean free path dependent on rigidity naturally produces an

enhancement of high A/Z particles as suggested by Eichler (1979). To see this we assume the shock profile is determined by protons ($A/Z=1$) and we have a small admixture of high A/Z particles. The high A/Z species will have a longer mean free path, see a sharper shock, and therefore be accelerated more than the protons. This will result in an over abundance of high A/Z particles relative to protons at high velocities.

Figure 4 shows spectra for different A/Z using the velocity profile of Fig. 3.

Again it must be emphasized that the relative importance of this effect depends on the shape and height of the low velocity shoulder in Figs. 2 and 4; that is, on the details of how the thermal particles interact with the collisionless shock which is addressed in this calculation only through the very simple scattering law of Eq. (3).

The presence of this effect depends on two things. First, the shock must be smoothed by the back pressure from high velocity particles and second, the injection velocity must be low enough so that the particles do not see the shock as a discontinuity. In other words, a seed particle description with large injection velocities will not show an enhancement effect.

Conclusions

The kinetic simulation of shock acceleration using the Monte Carlo method offers an alternative with several advantages to solving the diffusion equation. It becomes simple to introduce a velocity dependent mean free path for scattering, a finite scale either by a velocity cutoff or a free escape boundary, and no restriction $v \gg u$ need be made. This allows a self-consistent calculation of the shock wherein the pressure of the accelerated particles produces a velocity profile that conserves the momentum flux (within 20%) of these particles across the shock. We obtain the following results:

- a) For discontinuous shocks, a power law with an exponential cutoff at approximately $u_2 d/k = 1$ results when $v_i \gg u_2$.
- b) For $v_i < u_2$, within the restriction of the simple scattering law that we use, the proportion of high energy particles drawn from the thermal population is obtained.
- c) The inclusion of the back pressure of the scattering particles on the inflowing plasma produces a smoothing of the shock profile. This implies that the spectra are steeper than for a discontinuous shock. This effect may be balanced to some degree by the increase in the compression ratio due to the energy loss from escaping particles.
- d) The combination of a finite scale shock profile and a mean free path dependent on A/Z implies that high A/Z particles will be enhanced at high velocities over protons. (It has been shown by Cesarsky (1981) that for certain source temperatures the enhancement of galactic cosmic rays over local galactic abundances may show a positive correlation with A/Z). This effect saturates as A/Z increases and, for the parameters used here, produces an enhancement in the helium to proton ratio by about a factor of 3.

Acknowledgements. The authors wish to thank L. Drury for helpful discussions and one of us (DCE) wishes to thank H. Volk and G. Morfill for hospitality provided at the Max-Planck-Institut für Kernphysik, Heidelberg, FRG where a portion of this work was completed. DCE was supported in part by the NASA Graduate Student Researchers Program and in part by the NASA Solar Terrestrial Theory Program. DE was supported in part by NSF Grant AST 800 2673, the NASA Solar Terrestrial Theory Program, and the Sloan Foundation.

References

- Axford, W.I., Leer, E., Skadron, G.: The acceleration of cosmic rays by shock waves. Presented at 15th International Cosmic Ray Conference, Plovdiv, Bulgaria, 1977
- Bell, A.R.: The acceleration of cosmic rays in shock fronts - I. *Monthly Notices R. Astron. Soc.* **182**, 147-156, 1978
- Blandford, R.D., Ostriker, J.P.: Particle acceleration by astrophysical shocks. *Astrophys. J.* **221**, L29-L32, 1978
- Cesarsky, C.J.: Mass per charge ratio in hot plasmas and Cosmic ray source composition. *Proc. 17th Int. Cosmic Ray Conf., Paris, France*, **2**, 269-272, 1981
- Cowsik, R.: Comments on stochastic acceleration of cosmic rays. *Astrophys. J.* **241**, 1195-1198, 1980
- Drury, L.O'C., Volk, H.J.: Hydromagnetic shock structure in the presence of cosmic rays. *Astrophys. J.* **248**, 344-351, 1981
- Eichler, D.: Particle acceleration in collisionless shocks: regulated injection and high efficiency. *Astrophys. J.* **229**, 419-423, 1979
- Eichler, D.: Basic inconsistencies in models of interstellar cosmic-ray acceleration. *Astrophys. J.* **237**, 809-813, 1980
- Eichler, D.: Energetic particle spectra in finite shocks: the earth's bow shock. *Astrophys. J.* **244**, 711-716, 1981
- Ellison, D.C.: Monte carlo simulation of charged particles upstream of the earth's bow shock. *Geophys. Res. Lett.* **9**, 991-995, 1981
- Fisk, L.A., Lee, M.A.: Shock acceleration of energetic particles in corotating interaction regions in the solar wind. *Astrophys. J.* **237**, 620-626, 1980
- Forman, M.: First-order Fermi acceleration of the diffuse ion population near the earth's bow shock. *Proc. 17th Int. Cosmic Ray Conf., Paris, France*, **3**, 467-470, 1981
- Fransson, C., Epstein, R.I.: Acceleration and propagation of cosmic rays. *Astrophys. J.* **242**, 411-415, 1980
- Jones, F.C.: The dynamical halo and the variation of cosmic-ray path length with energy. *Astrophys. J.* **229**, 747-752, 1978
- Krymsky, G.F.: *Dok. Akad. Nauk. S.S.R.* **234**, 1306, 1977
- Parker, E.N.: A quasi-linear model of plasma shock structure in a longitudinal magnetic field, *J. Nucl. Energy, Part C*, **2**, 146-153, 1961
- Scholer, M., Ipavich, F.M., Gloecker, G., Hovestadt, D.: Conditions for acceleration of energetic ions ≥ 30 keV associated with the earth's bow shock. *J. Geophys. Res.* **85**, 4602-4606, 1980
- Terasawa, T., Energy spectrum of ions accelerated through Fermi process at the terrestrial bow shock. *J. Geophys. Res.* **86**, 7595-7610, 1981

Received January 22, 1981; Revised Version September 4, 1981
Accepted September 21, 1981

Acceleration Processes at the Earth's Bow Shock *

V. Formisano¹, S. Orsini²

¹ Space Science Department, ESA-ESTEC, Noordwijk, The Netherlands

² Laboratorio Plasma Spazio, CNR-Frascati, Italy

Abstract. ISEE-2 solar wind data are used to study acceleration processes at the Earth's bow shock by looking at the details of particle velocity distribution functions on several crossings. In the magnetosheath the transmitted accelerated particles show, downstream of a quasi-perpendicular shock, energy spectra which are different from those observed downstream of quasi-parallel shocks. A clear cut-off in the energy spectrum can be identified in the first case, while pronounced high energy tails are observed in the second case. The reflected backstreaming particles also show a cut-off in the energy spectrum. We conclude that the same acceleration mechanism generates the transmitted accelerated particles and the reflected ions. This mechanism is located at the shock, and is due to the electric field component tangent to the shock. Some evidence for weaker acceleration can be found at quasi-parallel shocks, where 5–10 keV ion flux appears to be constant going from upstream to downstream-zones. We conclude that diffuse ions cannot be accelerated at the shock, but are the result of this interaction of reflected particles with the solar wind. Particles which are stationary in the shock frame have been observed for both quasi-parallel and quasi-perpendicular shocks.

Key words: Collisionless shocks – Acceleration at shocks – Earth's bow shock – Shock structure

Introduction

Particles accelerated and reflected at the Earth's bow shock interact with the solar wind in the foreshock region. (Fairfield 1969; Asbridge et al. 1968). The acceleration at the shock was modelled by Sonnerup (1969) and by Greenstadt (1975); evidence for ion acceleration at the shock was reported by Montgomery et al. (1970) and by Formisano and Hedgecock (1973a). The presence of a high energy tail, downstream of the quasi-perpendicular turbulent shocks observed by Formisano and Hedgecock (1973b) (see also the review by Formisano, 1977) is also an indication that strong acceleration takes place at the Earth's bow shock. A fraction of the particles accelerated at the shock escape upstream and interact with the solar wind. Recently a division into two groups has been suggested for this population of accelerated particles: "reflected" and "diffuse" accelerated particles,

characterized by different angular-energy distribution, the first being a population of cold particles, localised in energy and angle; the second being a population of hot particles and appearing "diffuse" in energy and angle (Gosling et al. 1978). Only diffuse backstreaming particles are observed simultaneously with the magnetohydrodynamic (MHD) magnetosonic waves (Paschmann et al. 1979; Bonifazi et al. 1980). Formisano et al. (1980) have shown how the solar wind proton bulk speed and direction of flow oscillate when diffuse particles are observed, and how the solar wind is practically uperturbed in the presence of reflected particles. Formisano and Amata (1976) and Bame et al. (1980) have shown that the solar wind slows down and deviates from its direction while interacting with the diffuse particles.

Although there is evidence for acceleration at the bow shock, a detailed experimental study of this process is not yet available in the literature and the relationship between the reflected particles and shock structure is not clear. Are the diffuse and reflected particles coming from the same region of the shock or from different regions? Are the diffuse particles produced by the interaction of the reflected particles with the solar wind or accelerated at the shock they produce the instabilities and associated waves, by intersetting with the solar wind?

It is likely that the same mechanism that is accelerating particles to produce "reflected" ions can also produce transmitted accelerated particles. So another interesting question is: what are the properties of the transmitted accelerated particles? Can they reach the same energy as the reflected particles?

In the following we will address some of these problems by analyzing plasma particle data from the solar wind plasma experiment on board ISEE-2. The experiment has been described by Bonifazi et al. (1978); here we mention only that for hot particle populations (such as magnetosheath, backstreaming particles and occasionally also the solar wind) we can increase the time resolution of the measurements from 96 s per sample to 12 s per sample because the 64 energy windows available were obtained by means of 8 fast scans of the energy range covered (55 eV–11 keV). Each of these 8 subcycles provides 8 energy windows with complete angular distribution. Therefore, for a population of hot particles they provide enough information to characterize plasma parameters.

In this paper we first present a qualitative study of some properties of the reflected and transmitted accelerated particles in order to identify their energy properties. We then study in detail two shock crossings where reflected particles are observed immediately upstream and two shock crossings where diffuse particles are observed immediately upstream. Conclusions and discussions will be presented at the end of the paper.

* Paper presented at the Workshop on Acceleration of Particles by Shock Waves, 7–9 October 1980, Max-Planck-Institut für Aeronomie, Lindau, Federal Republic of Germany

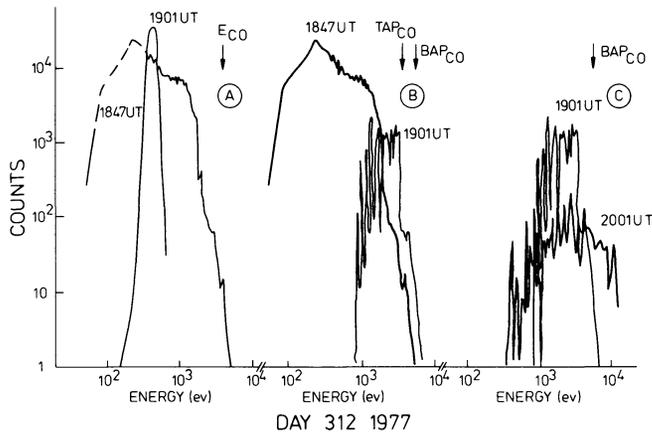


Fig. 1. Comparison of ion energy spectra for day 312, 1977. *A* Solar wind (1901 UT) and magnetosheath (1847 UT): transmitted accelerated ions are observed in the magnetosheath above solar wind energies. *B* Magnetosheath (1847 UT) and solar wind reflected particles (1901 UT): these particles can reach energies higher than those of the transmitted ones. *C* Reflected ions (1901 UT) and diffuse particles (2001 UT): the diffuse particles can have energies higher than reflected particles

Energy Spectra of Particle Populations near the Bow Shock

A first qualitative answer to some of the problems mentioned above can be obtained by means of a simple inspection of measured energy spectra in the bow shock region. Energy spectra observed in the vicinity of the bow shock, when reflected particles are present, are displayed in Fig. 1. Data refer to day 312, 1977, when the bow shock was crossed at 1,857 UT (Fig. 8 and Paschmann et al. 1979). The solar wind peak energy was found to be near 500 eV, and very few particles had energy above 600 eV (Fig. 1a). In the magnetosheath the energy spectrum is observed to extend to much higher energies, and a cut-off was actually observed at 3,850 eV. (The cut-off was located by identifying the energy where the count rate went below 3 ct/s.) The acceleration mechanism for the transmitted particles is, in this case, not able to provide many particles with energy larger than 3,850 eV. Reflected particles were observed in the solar wind, and their energy spectrum is compared with that of transmitted ions in Fig. 1b. A cut-off is also observed for the reflected ions, but in this case it is at 5,400 eV, i.e. at an energy level 1,500 eV higher. The particles that are able to escape upstream gain more energy than the ones transmitted downstream. One hour later the solar wind energy is still the same, but the magnetic field direction has changed slightly in such a way that MHD waves and diffuse particles are observed. A comparison with the reflected particles is presented in Fig. 1c: The diffuse ions are spread over a wider range of energies and do not show any cut-off. The important point we would like to stress here is that the spectrum of the diffuse population shows many particles with energy higher than the solar wind energy, higher than the energy of transmitted ions and higher than the reflected ion energy. It is also interesting to look at magnetosheath energy spectra downstream of a quasi-parallel shock. A typical case is shown in Fig. 2 for day 308, 1977. Solar wind protons were observed at 0300 UT, to have energy around 700 eV (Fig. 2a). In the magnetosheath, at 0256 UT, the energy spectrum had a shape different from that of day 312, 1847 UT: no cut-off was observed, and at 10 keV the flux was at least a factor 50 larger on day 308 than on day 312 although the total number density was lower. There may be fewer transmitted accelerated

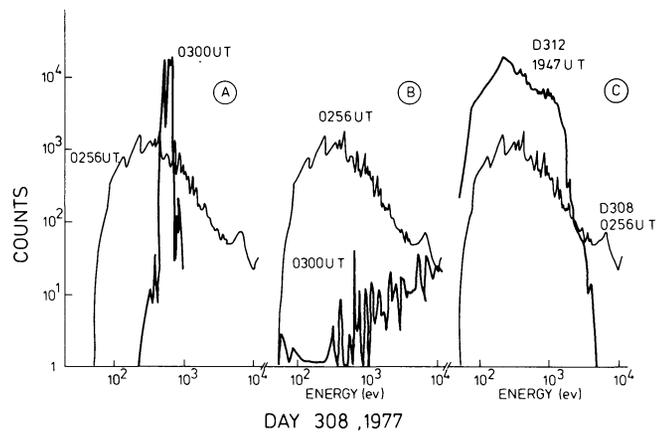


Fig. 2. Comparison of ion energy spectra for day 308, 1977. *A* Solar wind (0300 UT) and magnetosheath (0256 UT). *B* Magnetosheath (0256 UT) and solar wind diffuse ions (0300 UT). *C* Magnetosheath downstream of quasi-perpendicular shock (day 312, 1947 UT) and downstream of quasi-parallel shock (day 308, 0256 UT)

ions in the quasi-parallel shock, but they reach energies higher than downstream of a quasi-perpendicular shock (Fig. 2c). The magnetosheath energy spectrum in Fig. 2a actually has the shape of a power law: $E^{-3.8}$. The previous result is confirmed by looking at more than 100 magnetosheath passes. If we compare transmitted and reflected ion energy spectra for day 308 (Fig. 2b), we see that at high energies the flux is comparable inside and outside the shock. This is to be expected if the diffuse ions are the result of the interaction of the reflected ions with the solar wind. Conversely if the shock is the source, it must produce as many diffuse as transmitted ions for quasi-parallel conditions, which is different from what is observed in a quasi-perpendicular shock. In any case we are dealing with two acceleration mechanisms. As we shall see later we have evidence that the quasi-parallel shock is not producing the diffuse ions, although it is accelerating particles to lower energies than a quasi-perpendicular shock.

The cut-off energy (E_{co}) for the transmitted ions downstream of quasi-perpendicular shocks has been statistically studied in order to infer properties of the acceleration mechanisms. E_{co} has a very strong dependence on the energy of the solar wind (E_{sw}). In Fig. 3 we show a scatter plot of E_{co} versus E_{sw} with the best linear fit given by $E_{co} = 10.6 E_{sw} - 1,519$ eV, the correlation coefficient being 0.92. The three highest points in Fig. 3 are only lower limits because we observe within our energy range (up to 11 keV) only part of the decrease of the energy spectrum and, from extrapolation, 10 counts should be reached above 15 keV. We have also studied the scatter of the points present in Fig. 3 using the linear fit given above to bring the observed E_{co} to a fixed solar wind energy ($E_{sw} = 500$ eV). The cut-off energy for fixed energy input (E_{co} , $E_{sw} = 500$ eV) is shown to be correlated with the magnetosonic Mach number (M_{MS}) in Fig. 4 with correlation coefficient 0.49. Larger M_{MS} values occur together with lower E_{co} when E_{sw} is fixed. This result may be important when we extrapolate from the bow shock to astrophysical shocks, in order to obtain large acceleration factors. This needs to be further investigated in view of the low correlation coefficient found.

In order to study properties of acceleration processes at the shock, we note first that reflected ions have not been observed

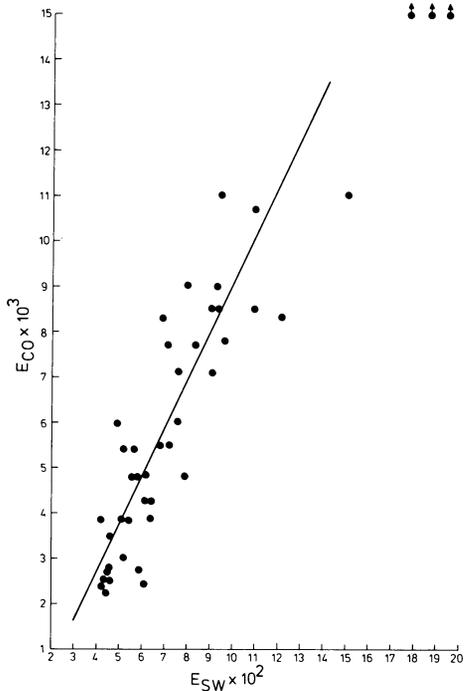


Fig. 3. Relationship between the cut-off energy (E_{co}) of the transmitted particles with the solar wind energy E_{sw} . The linear best fit through the points is given by $E_{co} = 10.6 E_{sw} - 1,519$

upstream of shocks with $\widehat{B}\widehat{n} \geq 70^\circ$, while diffuse ions have not been observed upstream of shocks with $\widehat{B}\widehat{n} \geq 60^\circ \sim 55^\circ$ (Bonifazi and Moreno 1981). It is clearly of interest to study particle properties at shocks with $\widehat{B}\widehat{n} > 70^\circ$, $55^\circ < \widehat{B}\widehat{n} < 70^\circ$ and $\widehat{B}\widehat{n} < 55^\circ$ in order to locate the source of the reflected and transmitted ions, and in order to identify the acceleration mechanism.

Quasi-Perpendicular Shocks

One shock crossing with $\widehat{B}\widehat{n} = 78^\circ$, and no backstreaming particles, was observed on day 335, 1977 at 2220 UT. Data for the period 2216–2222 UT are shown in Fig. 5. This figure deserves some comment because figures of this type will be used several times later. The angular distribution of one or two energy channels is shown at the top, for the solar wind proton peak and, if present, for the peak in the energy spectrum of the reflected ions. In the magnetosheath usually only one energy channel is shown, namely the peak in the energy spectrum. At the shock one or two significant energy channels are shown depending on the relevance of the information to be presented. Numbers like E21 indicate that the angular distribution of energy channel number 21 is shown. It should be noted that in the angular distribution there is a blind sector at all energies (Bonifazi et al. 1978) with no measurement between -34° and -56° indicated by a dashed sector in the figures. The angular distributions are always presented with the sun direction given by the arrow on the top of the figure. The radial distance from the centre gives, on a logarithmic scale, the particles counted when the instrument was looking in that direction. For the 90° sectors the number of particles counted has been divided by 16 in order to have measurements comparable with the 5.6° sectors, and should be considered as the average counts over 16 sectors 5.6° wide. Below the angular distributions there is, in Fig. 5, a strip showing energy

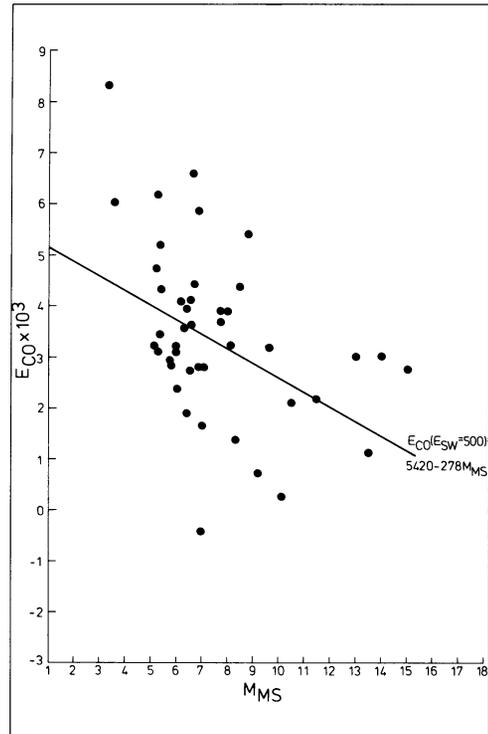


Fig. 4. Relationship between E_{co} of transmitted particles normalized to a fixed solar wind energy ($E_{sw} = 500$ eV), and the Mach number M_{MS} . The best linear fit through the points is also given.

spectra. Each energy spectrum is shown on a log scale, however the energy scale is shown in a large size when more than 8 energy channels are presented. It should be noted that the energy range covered is the same in both cases. The subcycle spectra (8 channels) are used in order to increase our time resolution without shrinking the energy range covered. One or two energy spectra are shown for each subcycle, taken looking in two different directions. The energy spectrum observed in the solar wind direction is represented by a dashed line, the one at 90° to it (or magnetosheath direction, 90° wide) is represented by a continuous line. Magnetic field data (one component B_z or the magnitude B) is shown at the bottom of the figure in order to locate the spacecraft with respect to the shock. The timing of the energy spectra is shown by open boxes. The parameters M_{MS} , $\widehat{B}\widehat{n}$ and β characterising the shock are shown on the left hand side of the magnetic field strip, and were computed using the solar wind hourly averages published by King (1980).

The shock observed on day 335 at 2220 UT was a quasi-perpendicular turbulent shock. In the solar wind (spectrum 1) a typical situation was observed with no backstreaming particles. The solar wind energy spectrum shows the multiple peaks due to ${}^4\text{He}^{2+}$ and heavy ions as is usually observed with electrostatic analyzer. The angular distribution also shows the cold collimated solar wind ion beam. As we move close to the shock (spectrum 2) we also observe, together with the solar wind, a consistent population of particles that gyrate around \widehat{B} and are able to go upstream for a distance $\simeq 1 R_j$ (gyroradius) (foot of the shock). Note that the energy spectrum of these reflected particle peaks at energies below the solar wind energy. The next spectrum is very interesting: we are still on the solar wind side of the shock, as it is confirmed by the cold beam revealed by the angular distribution. The solar wind has, however, been slowed down considerably: the peak is observed in channel 13

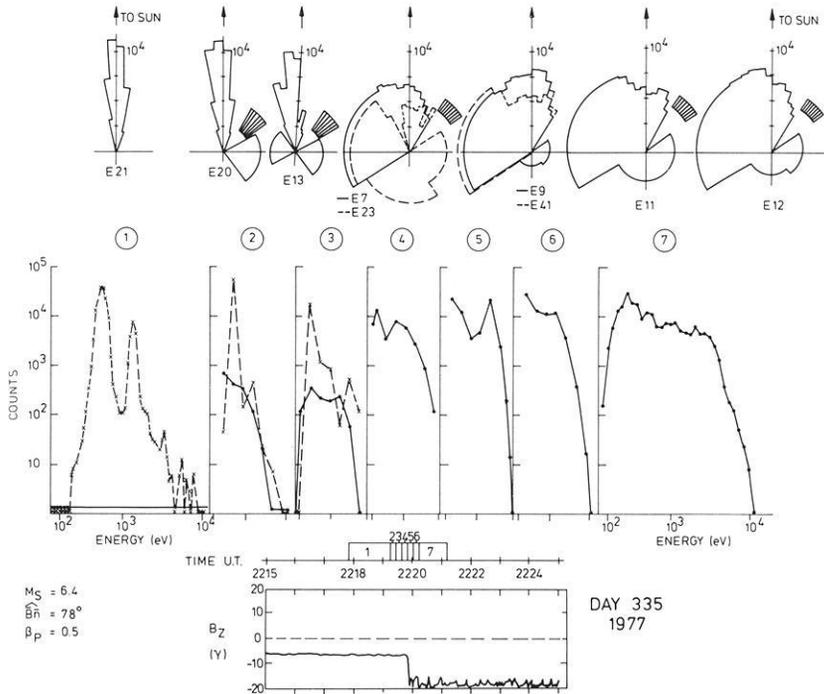


Fig. 5. Details of a quasi-perpendicular shock crossing with no reflected ions. On the top strip we have angular information for the peak proton energy channel. Two angular distributions (two energies indicated by energy channel number) are shown when two peaks (two particle populations) are observed. Note the “blind” sector at -34° to -56° , where no measurements are taken for technical reasons. On the middle strip we have energy spectra covering the energy range 55 eV–11 keV. Note that the energy scale is reduced when only 8 energy channels (subcycles) are used. On the *bottom* of the figure the shock parameters (magnetosonic Mach number, \bar{B}_n and β_p) are given on the *left hand side*; also, on the *bottom* of the figure, the time scale is given with indication of the time intervals during which the observed numbered spectra were taken. Note in this figure the increase of the cut-off energy from 3.5 keV (spectrum 2) to 3.6 (3), 5.5 (5), 6.5 (6), 8 keV (7). See text for further explanations

(442 eV) while before it was in channel 20 (641 eV). From the angular distribution it can be seen that the solar wind has already been somewhat deflected and heated. At the same time, at the high energy end of the observed spectrum, the flux has increased to almost the magnetosheath level. In the spectrum observed at 90° to the solar wind there is evidence for a first peak at 442 eV and for a second peak at 1,606 eV. The presence of the two peaks is confirmed by the next two energy spectra (4 and 5 in Fig 5) which clearly show magnetosheath plasma already heated (see the angular distributions). The first peak is at 236 eV and the second at 800–1,000 eV in spectrum 4 and around 2,000 eV in spectrum 5. The second peak becomes a plateau in spectrum 6 and a long high energy tail followed by a cut-off at $\sim 8,000$ eV in spectrum 7.

The cut-off was always present in the spectra at 90° to the solar wind direction (full lines in Fig. 5) and is found at lower energies in spectra 2 and 3 when no particles are observed at energies of 3,500 and 3,650 eV, while in spectrum 5 it is at 5,500, and in spectrum 7 at 8,000 eV. It is evident that the particles accelerated at the shock are not able to escape upstream in this situation ($\bar{B}_n = 78^\circ$). Only transmitted accelerated particles are observed, and their maximum energy increases going from upstream to downstream. A typical feature for this kind of shock is a bimodal ion velocity distribution immediately after the shock and just upstream of it. Both accelerated and decelerated particles are able to reach the immediate upstream region (foot) of the shock before being convected downstream.

Next we will deal with quasi-perpendicular shock structures with reflected particles observed upstream. Two cases with different solar wind energies will be described in order to be able to explore energies much larger (day 312, 1977) or lower (day 337, 1977) than those associated with the solar wind. On day 312, 1978 the bow shock was crossed at 1852:50 UT (Fig. 6), with $\bar{B}_n = 60^\circ$ and $M_{MS} = 4$. Due to rather low magnetic field intensity and high plasma density, the value of β (plasma pressure to magnetic pressure ratio) was rather high for this shock crossing: $\beta = 19$ (see also Russell and Greenstadt 1979). The high

value of β increases the magnetic turbulence as discussed by Formisano et al. (1976). Here we are interested in the behaviour of the plasma particles. In Fig. 6 the solar wind energy spectrum is shown on the right side (dashed curve), as this was an out-bound crossing. Upstream reflected ions are observed between 0.8 and 8 keV together with the cold solar wind (angular distribution of the spectrum taken at 1900 UT). Some fluctuations are present in both the solar wind and the reflected particle energy spectrum (see also the data for this crossing in the previous section). The magnetosheath energy spectrum is shown on the left hand side and is similar to the sheath spectrum 7 in Fig. 5. The cut-off energy is observed at 3,850 eV. Spectra 2 and 3, on the magnetosheath side of the shock, revealed the presence of two peaks, one at low energy and another at 1,100 eV. Spectrum 4 is very interesting for its shape and for the low energy angular distribution. In channel 9 (350 eV) the angular distribution shows two peaks, one being the cold solar wind, the other one being similar to the magnetosheath population of particles. At higher energy a peak is observed only in this direction. This peak is at 1 keV like the peak observed in the magnetosheath. Spectrum 4 identifies the shock in the ion plasma particles as it separates magnetosheath from cold solar wind energy spectra. The cut-off of the second peak, which is going to generate the reflected ions is at 5,370 eV (as it is also confirmed by spectrum 5). Spectrum 4 is observed between the first large pulse in the magnetic field intensity and the second final increase of intensity of \bar{B} . The solar wind is observed from spectrum 5 to spectrum 9 together with reflected ions. We note a decrease of the peak flux intensity of the reflected ions between spectra 5 and 8, and this decrease corresponds also to a change of the peak energy from 1 keV to 2 keV. At the same time, however, the cut-off does not change and is still at 5,400 eV. We conclude therefore that between spectra 5 and 8 there is no further acceleration, and the difference in the reflected ion spectrum can be explained by loss of the low energy part of the population. This conclusion is confirmed by a direct comparison (overlapping) of spectra 8 and 5: at high energy the two spectra overlap,

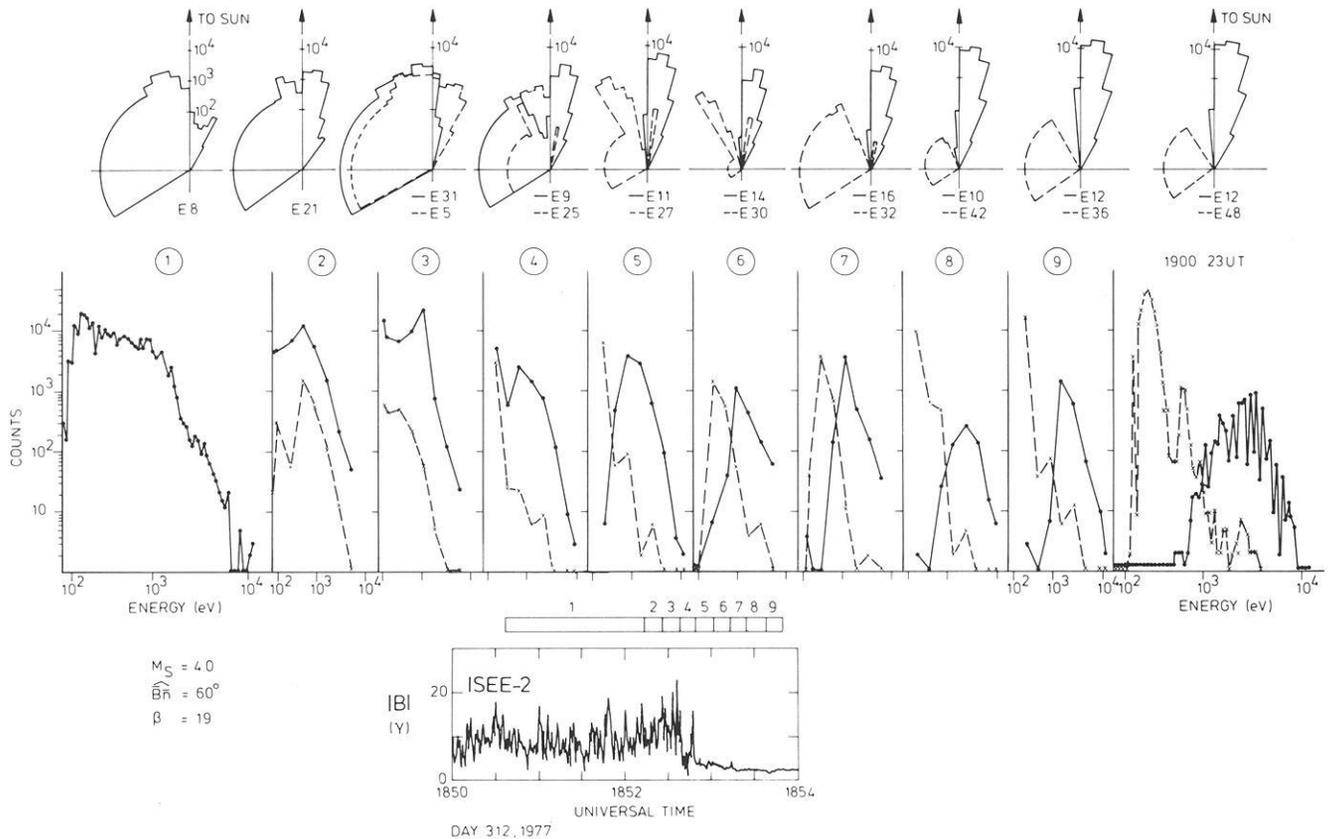


Fig. 6. Details of quasi-perpendicular shock crossing with reflected ions observed upstream (day 312, 1977, 1853 UT). Note the decrease of flux intensity and related increase of peak energy between spectrum 5 and spectrum 8 of the reflected ions (*continuous lines*). The format of the figure is the same as in Fig. 5; the magnetic field intensity from ISEE-2 is shown on the bottom. Note the two streams in the angular distribution (4) identifying the shock crossing

while at low energy part of the population present in 5 is missing in 8.

In conclusion we locate the acceleration region, by means of spectra 4 and 5, at the large magnetic intensity gradients, i.e. at the shock, because on the solar wind side the flux of reflected ions decreases as we move away from the shock, and on the sheath side the flux observed in the second peak (accelerated particles) also decreases going away from the shock.

The behaviour of the reflected ions at low energy is studied by means of the observations on day 337, 1977, when the bow shock was crossed at 0055 UT. Our data are shown in Fig. 7 together with one component (B_z) of the magnetic field data from the ISEE-1 spacecraft. As the magnetic field is shown just as a guide to locate the observations with respect to the shock, the complete magnetic information from the spacecraft ISEE-2 (from which the plasma data were obtained) is not needed. The shock was again a quasi-perpendicular shock with $\widehat{Bn} = 69^\circ$ and $M_{MS} = 4.5$, $\beta_p = 0.1$. Reflected ions were observed in spectrum 1, showing a cut-off similar to the one already presented in Fig. 1 b. The magnetosheath spectrum shows the same features we have seen for the previous two shocks (see spectrum 10 in Fig. 7). From spectra 2–9 in Fig. 7 we can study how the reflected ion energy spectrum changes at low energies (below the solar wind energy) when we approach the bow shock from upstream. From spectrum 1 we see that no backstreaming particles were observed below the solar wind energy (1,200 eV at the peak). In the solar wind energy range some fluctuations of the reflected particles appear in spectrum 1. In spectra 2, 3 and 4 the reflected particle

spectra show no fluctuations, but tell us rather that there are no reflected particles at energies below the solar wind energy. In this sense it is interesting to note in the angular distributions that in the solar wind peak energy channel no reflected particles are observed in spectra 1–4. In spectra 5–7 we see a modification of the reflected ion spectrum, in the sense that particles are now added at energies lower than the solar wind energy, and at very high energies. In spectra 8 and 9 we observe that the solar wind has been partially deflected and thermalized. From these spectra we conclude that we are at the shock, i.e. we see that at low energy we still have a cold population, while at high energy (the second peak in spectrum 9 is at 3,500 eV) the plasma is hot and appears from all directions. This mixed situation is observed for another complete cycle (now shown in the figure) before spectrum 10 is observed in the magnetosheath. Subcycles of these data not shown in the figure present bimodal energy spectra like spectrum 3 in Fig. 6.

In conclusion we have shown that quasi-perpendicular shocks do always accelerate particles. For $\widehat{Bn} > 70^\circ$ the accelerated particles can only be transmitted, although they are already observed in the foot of the shock. For $\widehat{Bn} < 70^\circ$ they can leave the shock, and are observed as reflected ions upstream of it. At the shock we have energy spectra with two peaks and angular distributions with two peaks. The transmitted particles always show a cut-off at an energy that depends on the solar wind velocity (see previous section). The backstreaming reflected particles also show a cut-off, but at an energy higher than the cut-off for the transmitted ones. For $\widehat{Bn} > 70^\circ$ the cut-off of the transmitted particles moves

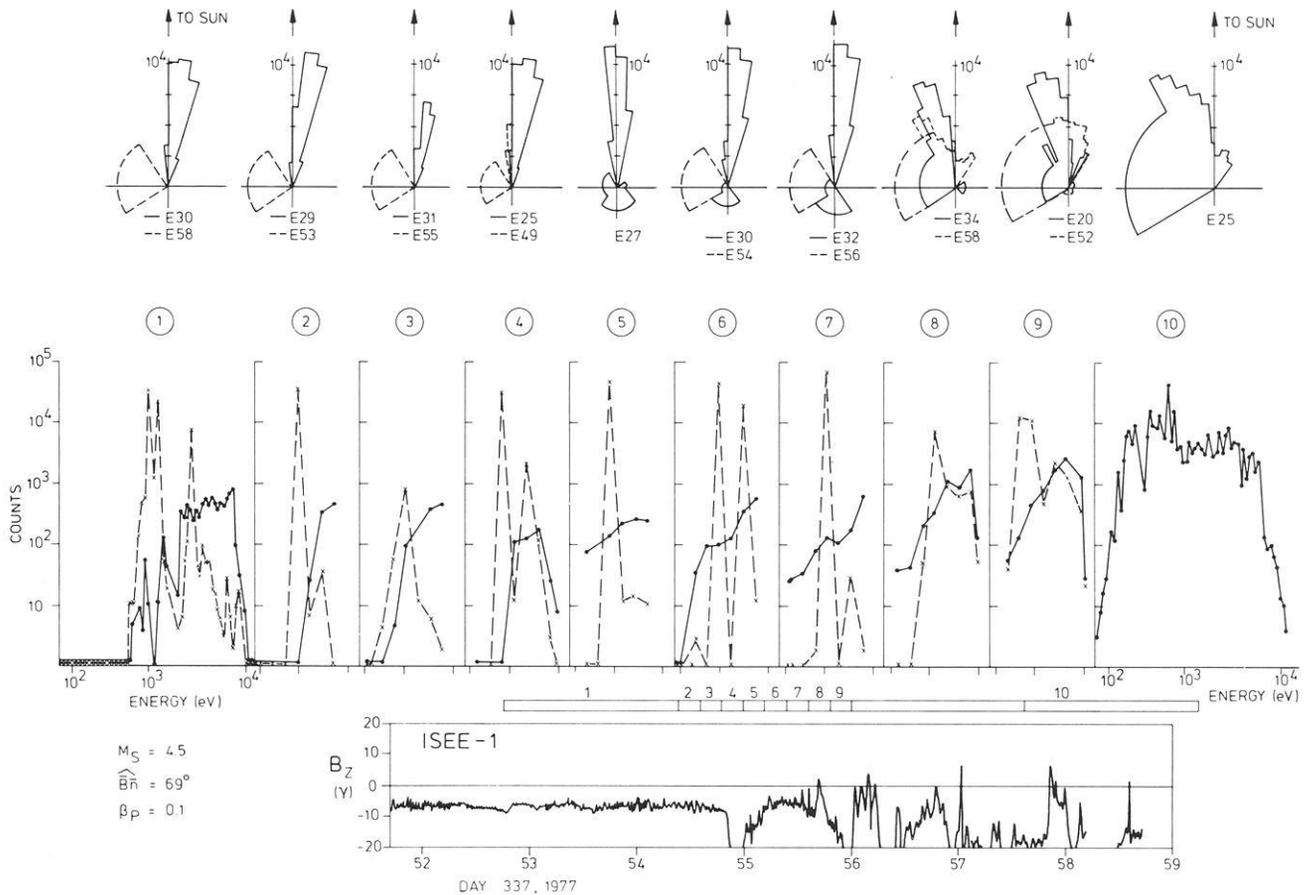


Fig. 7. Details of a quasi-perpendicular shock crossing with reflected particles observed upstream (day 337, 1977, 0055 UT). Note the increase of particles flux at energies below the solar wind energies as we approach the shock. On the bottom one component (B_z) of the magnetic field observed by ISEE-1 is shown as a reference. The format of the figure is the same as in Fig. 5

from lower to higher energies (from 3,500 eV to 8,000 eV) as we move from the foot of the shock to the sheath side of it. The flux intensity of the reflected ions increases as we move closer and closer to the shock. Close to the shock a population of low energy particles is added at energies lower than that associated with the solar wind. These particles are probably stationary in the shock frame, and may play an important role in the dissipation mechanism.

Quasi-Parallel Shocks

A quasi-parallel shock observed at 0259 UT on day 308, 1977 is shown in Fig. 8. As already mentioned, the upstream solar wind is very much perturbed when diffuse particles are observed, i.e. in front of quasi-parallel shock. Downstream of the shock there is no evidence for a cut-off in the proton energy spectrum and a long high energy tail is seen; fluxes of 10 keV particles comparable with the fluxes of 10 keV diffuse upstream ions are also observed (spectra 1 and 10 in Fig. 8). In Fig. 8, as we approach the shock from the magnetosheath side, the proton energy spectrum does not show the two peaks and the plateau observed in Fig. 5 (spectra 5 and 7), Fig. 6 (spectra 3 and 1), and Fig. 7 (spectra 9 and 10) for quasi-perpendicular shocks. Spectrum 4 in Fig. 8, which may be compared with spectrum 6 in Fig. 5, shows that the “plateau” due to transmitted particles, if present, has an ion flux a factor 20 lower than the first (main) peak. This plateau, on the other hand, can also be interpreted as the sum of the diffuse ions of upstream origin, plus ions locally

(at the shock) accelerated and/or heated and/or scattered. We note that spectra (and angular distributions) 4 and 5 in Fig. 8 are rather “intermediate” between the cold solar wind and the hot magnetosheath: spectrum 5 shows a decelerated warm solar wind, while the angular distributions in 3 and 4 show that the plasma has not yet been deflected. The spectrum – angular distribution that identifies the shock (spectrum 6 in Fig. 8) is also the one that most resembles the quasi-perpendicular shock observations (spectrum 3 in Fig. 5 and spectrum 4 in Fig. 6). In the angular distribution two streams are observed and at 90° to the solar wind the energy spectrum shows accelerated particles. With respect to the solar wind, however, the acceleration at this shock is much smaller than at quasi-perpendicular shocks: the ratio E_{accel}/E_{sw} is 2.9 on day 335, 1977 (680 eV (solar wind) to 2,000 eV (transmitted particles, second peak)) and only 1.4 on day 308, 1977 (680 eV (solar wind) to 950 eV (accelerated particles) at the shock – spectrum 6). These accelerated particles are not seen downstream, while upstream of the shock we see (going from spectrum 6, to 7, 8, 9 and 10) that their flux decreases rapidly, until they are practically absent even upstream in spectrum 10. Our interpretation of this data is that the acceleration mechanism valid for quasi-perpendicular shocks, also works here (with $\hat{B}\hat{n}=30^\circ$), although with much lower efficiency. In this case, the accelerated particles are quickly scattered in the phase space due to the presence of large amplitude waves carried at the shock by the solar wind itself. These waves have been produced by the interaction of the reflected particles with the solar wind far upstream of the shock. In this picture the diffuse ions,

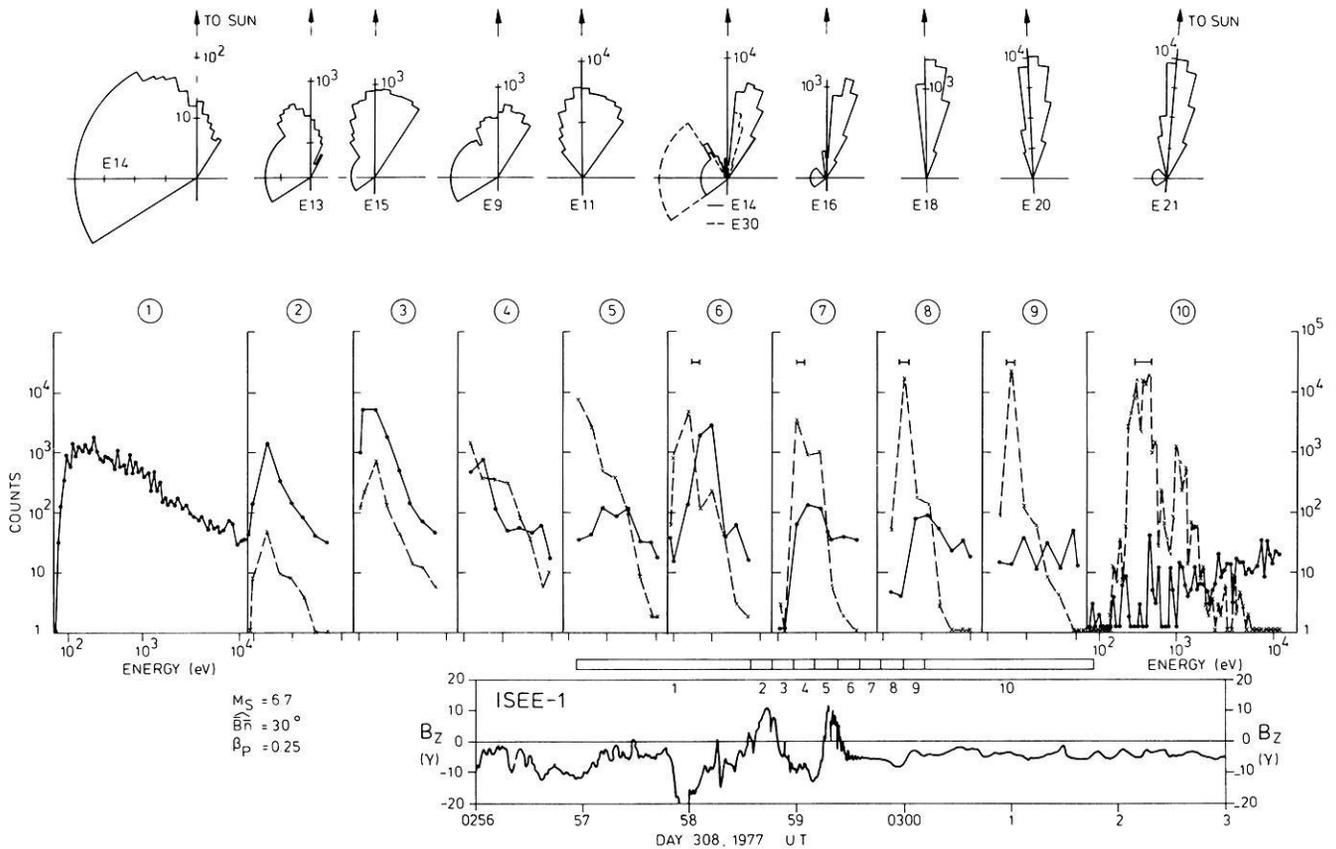


Fig. 8. Details of a quasi-parallel shock crossing with diffuse ions observed upstream (day 308, 1977, 0259 UT). The format of the figure is the same as in Fig. 5. Note the double ion streams in the angular distribution (6)

which show no clear direction of motion and a strongly oscillating angular distribution, (Formisano et al. 1980) are the results of some further acceleration mechanism due to wave particle interaction in the foreshock region.

This picture is confirmed by our second example of quasi-parallel shock crossing: day 316, 1977 at 1929 UT (Fig. 9). In this case $\bar{B}\hat{n}$ is still smaller than in the previous case: $\bar{B}\hat{n} = 24^\circ$ and we find hardly any evidence for acceleration at the shock. Our data are shown in Fig. 9 for the period 1927–1931 UT when a quasi-crossing of the shock was observed. The last spectrum on the right hand side was observed after the shock was finally crossed at 1940 UT. Upstream of the shock large amplitude MHD waves are observed in the magnetic component B_z together with diffuse ions (spectra 1 and 2). Downstream of the shock (spectra 7, 8 and 9) again we find no evidence for either a second peak or a plateau of transmitted particles. We do find, however, a long tail of energetic particles, with a power law-like energy spectrum. The shock is identified in spectrum 6 when the two population (solar wind and particles at 90° to it) reach similar fluxes, showing two beams in the angular distribution.

It is interesting to look at energy spectra 2–5. In spectrum 2 we have the same situation as in spectrum 1: solar wind plus diffuse ions at high energy. As we approach the shock the situation changes and we add a new population at low energies. Spectrum 3 shows that, at 90° to the solar wind, we observe an energy spectrum with two peaks: one below the solar wind energy at ≈ 100 eV and the second above 4 keV. The two populations are separated by a gap: no particles are observed between

1,200 and 1,800 eV. The low energy particles are probably stationary in the shock frame. Spectrum 4 shows that the flux of the low energy population has increased, and starts to fill the gap previously mentioned. The flux intensity of the high energy population, on the other hand, does not change. In spectrum 5 in the quadrant looking at the sun, the solar wind is deflected and slowed down, whereas at 90° to the solar wind a practically flat energy spectrum is observed between 50 eV and 10 keV. As on day 308, the energy spectrum observed at the shock (spectrum 6 in Figs. 8 and 9) is the only one that may perhaps give an indication of acceleration. This is, however, a rather weak indication, and the acceleration would be still smaller than for the case of Fig. 8.

In conclusion we may say that at quasi-parallel shocks there is less evidence (compared with quasi-perpendicular structures) for acceleration of particles. The acceleration is in any case smaller. Downstream of these shocks there is no evidence for a cut-off in the energy spectrum, i.e. no limit to the acceleration factor, and a long high energy tail (power law-like) is observed in ion energy spectra. When some small acceleration seems to be present at the shock, the accelerated particles are quickly scattered in phase space, so that no evidence of special features related to them is found immediately upstream or downstream of the shock. At very high energies (5–10 keV) the main feature observed is the continuity in the ion flux level from upstream, through the shock, downstream. Close to the shock a new population of particles is added, which is probably stationary in the shock frame of reference.

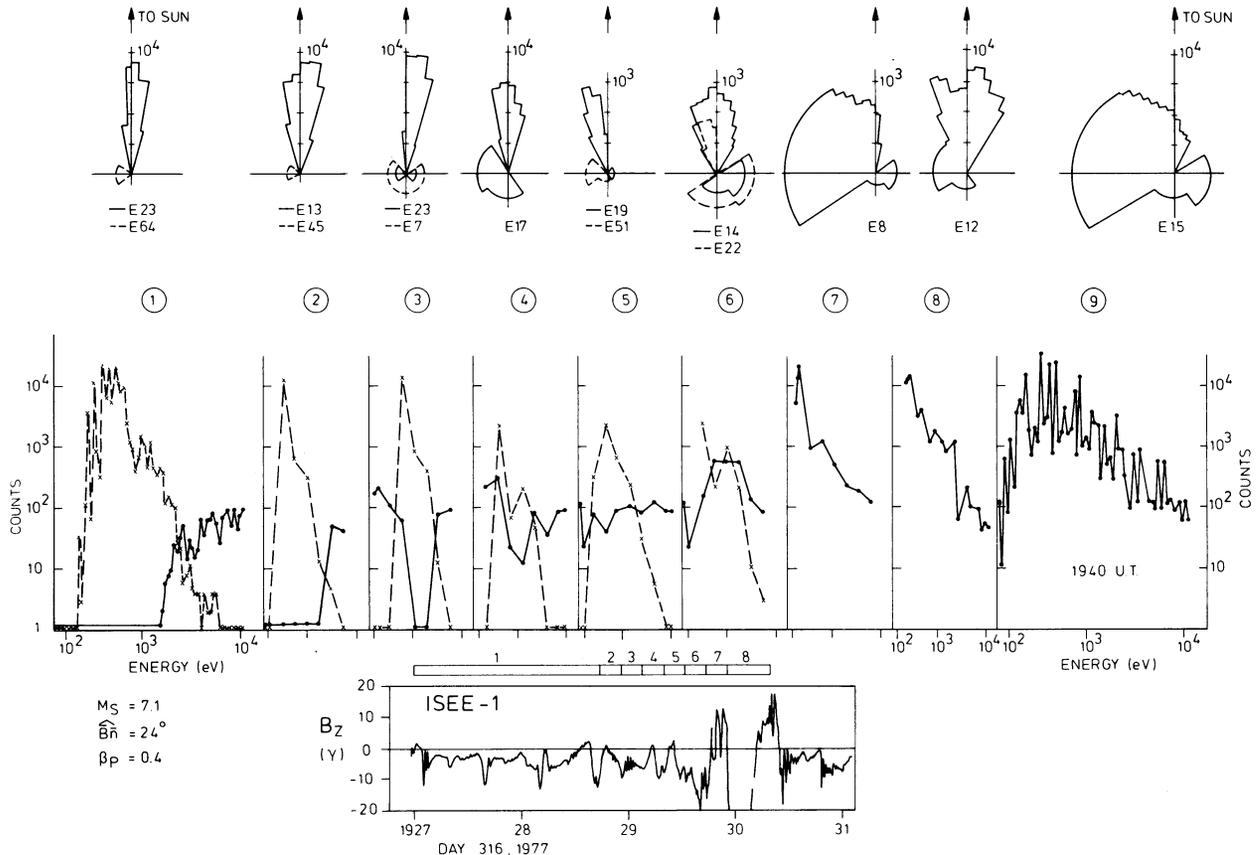


Fig. 9. Details of a quasi-parallel shock crossing with diffuse particles observed upstream (day 316, 1977, 1929 UT). The format of the figure is the same as in Fig. 8. Note the double ion streams in the angular distribution (6) and the low energy particles (stationary in the shock frame) in spectra 3 and 4

We interpret these observations as evidence for the bow shock being a source of decreasing power for reflections as \widehat{Bn} changes from 90° to 0° . The ions leaving the shock in the region with $\widehat{Bn} < 70^\circ$ interact with the solar wind at some distance from the shock, and generate all the MHD waves observed in the foreshock region. These waves also scatter the reflected ions in phase space, generating the diffuse particles which are then carried at the shock and downstream of it together with the waves. These diffuse ions are observed together with heated solar wind particles in the magnetosheath, where they give rise to the long, high energy tail.

Summary and Discussions

We have, in this study, presented a number of observations concerning the acceleration mechanisms present in the Earth's bow shock region. In summary it has been shown that:

a) Downstream of quasi-perpendicular shocks the transmitted particles do not show any cut-off in their energy spectrum, the maximum acceleration factor ranging between 5 and 11 times the solar wind energy.

b) Downstream of quasi-parallel shocks the transmitted particles do not show any cut-off in the energy spectrum, and few particles seem to have been accelerated to energies higher than downstream of quasi-perpendicular shocks (power law-like spectrum).

c) Upstream of the quasi-perpendicular shocks, the reflected ions also show a cut-off in the energy spectrum on the high energy side; this cut-off, however, is at higher energies than

the one for the transmitted accelerated particles downstream of the same shock crossings.

d) The cut-off energy of the transmitted particles appears to move toward higher energies for increasing solar wind speed.

e) The cut-off energy of the transmitted particles decreases with increasing Mach number once the dependence on the solar wind velocity has been eliminated.

f) The 5–10 keV ion flux level is constant across quasi-parallel shocks.

g) A detailed study of quasi-perpendicular shocks with $\widehat{Bn} > 70^\circ$ shows that

- acceleration is occurring at the shock as a bimodal energy spectrum is observed immediately downstream of it
- the accelerated particles cannot leave the shock on the upstream side (transmitted population)
- a two peaks angular distribution is observed at the shock
- the cut-off in the energy spectrum moves from “low” to high energies as we move from the foot of the shock to the magnetosheath
- at the shock, low energy particles are observed (pseudo trapped particles).

h) A detailed study of quasi-perpendicular shocks with $70^\circ > \widehat{Bn} \geq 60^\circ$ shows the same characteristics listed in g), except that the accelerated particles can also leave the shock moving upstream (reflected backstreaming population). For these particles we note that

- the flux decreases as we move away from the shock
- the apparent increase in the average energy of the backstreaming accelerated ions as we move away from the shock

is due to loss of the low energy population as these particles are carried downstream by the solar wind.

i) A detailed study of the quasi-parallel shock crossings with diffuse ions upstream shows that some features of the quasi-perpendicular shocks are kept, such as the double-peaked angular distribution at the shock. We note, however, that

– the acceleration of particles at the shock, if present at all, is less effective

– evidence for these accelerated particles is quickly lost as we move away from the shocked layer, probably because they are quickly scattered in the phase space

– way upstream of the shock there has to be another acceleration mechanism producing the diffuse ions, as these particles also have energies larger than the reflected ones and do not show any cut-off

– while at high energies (5–10 keV) the ion flux is constant, as we go from upstream through the shock downstream; at low energies, just upstream of the shock, a new population with energy comparable or even smaller than the solar wind energy is observed (pseudo trapped particles).

Concerning the problem of the origin of diffuse ions, we may say that there is enough evidence from the detailed study of crossings of quasi-parallel structures to show that they are not generated at the shock, unlike the reflected particles. On the other hand there are several indications that the diffuse ions are the result of the interaction between the reflected particles and the solar wind. The continuity of the ion flux with energy 5–10 keV across the shock excludes the shock itself as a source for these particles. The acceleration at quasi-parallel shocks is weaker, not stronger, than at quasi-perpendicular shocks, therefore these shocks cannot provide ions with energies larger than the energy of reflected ions. In our opinion the reflected particles do interact with the solar wind, but the interaction has a long growth time compared with the time the particles need to travel a few Earth radii. This may not be true if the unstable waves are already present, as it may be at quasi-parallel shocks with \tilde{B}_n not too small. In the interaction the reflected ions are scattered in phase space, therefore some of them will also be accelerated, and the resulting diffuse ions will be carried by the solar wind at the shock (where the newly generated population of reflected ions will be added to them and quickly scattered) and downstream (where they will form the long, high energy tail of the magnetosheath ion velocity distribution). Our data therefore support the interpretation of the solar wind reflected ions interaction given by Paschmann et al. (1979) and Bame et al. (1980) in contrast with the analysis by Greenstadt et al. (1980).

Concerning the problem of the acceleration mechanism, we consider two possible ways of producing transmitted and reflected ions: acceleration due to the solar wind electric field $\vec{E} = -\mathbf{V} \times \mathbf{B}$ tangent to the shock (Sonnerup 1969; Paschmann et al. 1980), or due to large amplitude waves present in the shock region.

We note that on the basis of our results, the acceleration mechanism has to produce a cut-off for the transmitted ions, this cut-off has to be at energies smaller than for the reflected ions. This cut-off has to move to higher energies as we go from upstream (foot of the shock) to downstream. The solar wind electric field acceleration could provide some of these features, but, on the other hand, it is difficult to understand the behaviour of the cut-off energy when going from the foot, through the shock, downstream, in terms of acceleration due to an electric field tangent to the shock. Acceleration due to waves must in

any case be present in the foreshock region, producing diffuse ions with energies higher than the reflected ones.

Acknowledgements. We are very grateful to C.T. Russell of UCLA for providing us with unpublished ISEE magnetic field data. We are also grateful for the involvement of a number of staff of IPS-CNR, Frascati in the EGD plasma experiment.

References

- Asbridge, J.R., Bame, S.J., Strong, I.B.: Outward flow of protons from Earth's bow shock. *J. Geophys. Res.* **73**, 5777–5789 (1968)
- Bame, S.J., Asbridge, J.R., Feldman, W.C., Cosling, J.T., Paschmann, G., Schopke, N.: Deceleration of the solar wind upstream from the Earth's bow shock and the origin of diffuse upstream ions. *J. Geophys. Res.* **85**, 2981–2990 (1980)
- Bonifazi, C., Egidi, A., Moreno, G., Orsini, S.: Backstreaming ions outside the Earth's bow shock and their interaction with the solar wind. *J. Geophys. Res.* **85**, 3461–3472 (1980)
- Bonifazi, G., Cerulli-Irelli, P., Egidi, A., Formisano, V., Moreno, G.: The EGD positive ion experiment on the ISEE-B satellite. *IEEE Trans. Geoscience Elect.* **GE-16** 243–247 (1978)
- Bonifazi, G., Moreno, G.: Reflected and Diffuse ions backstreaming from the Earth's bow shock: 2. Origin. *J. Geophys. Res.* **86**, 4405–4413 (1981)
- Fairfield, D.H.: Bow shock associated waves observed in the far upstream interplanetary medium. *J. Geophys. Res.* **74**, 3451–3465 (1969)
- Formisano, V.: The physics of the Earth's collisionless shock wave. *J. Phys.* **38**, C6, 65–88 (1977)
- Formisano, V., Hedgecock, P.C.: Solar wind interaction with the Earth's magnetic field. III) On the Earth's bow shock structure. *J. Geophys. Res.* **78**, 3745–3760 (1973a)
- Formisano, V., Hedgecock, P.C.: On the turbulent bow shock structure. *J. Geophys. Res.* **78**, 6522–6534 (1973b)
- Formisano, V., Orsini, S., Bonifazi, C., Egidi, A., Moreno, G.: High time resolution observations of the solar wind in the Earth's foreshock region. *Geophys. Res. Lett.* **7**, 385–388 (1980)
- Formisano, V., Amata, E.: Solar wind interaction with the Earth's magnetic field. 4-Preshock perturbation of the solar wind. *J. Geophys. Res.* **81**, 3907–3912 (1976)
- Gosling, J.T., Asbridge, J.R., Bame, S.J., Paschmann, G., Schopke, N.: Observations of two distinct populations of bow shock ions in the upstream solar wind. *Geophys. Res. Lett.* **5**, 957–961 (1978)
- Greenstadt, E.W.: The upstream escape of energized solar wind protons from the bow shock; in "The Magnetospheres of the Earth and Jupiter" ed. by V Formisano, Reidel Pub. Co. Dordrecht, Holland, 3–19 (1975)
- Greenstadt, E.W., Russell, C.T., Hoppe, M.: Magnetic field orientation and suprathermal ion streams in the Earth's foreshock. *J. Geophys. Res.* **85**, 3473–3479 (1980)
- King, J.H.: "Interplanetary medium data book Supplement 1" NASA NSSDC/WDC-A-R & S 79-08
- Montgomery, M.D., Asbridge, J.R., Bame, S.J.: "Vela 4 plasma measurement observations near the Earth's bow shock" *J. Geophys. Res.* **75**, 1217–1230 (1970)
- Paschmann, G., Schopke, N., Bame, S.J., Asbridge, J.R., Gosling, J.T., Russell, C.T., Greenstadt, E.W.: Association of low frequency waves with suprathermal ions in the upstream solar wind. *Geophys. Res. Lett.* **6**, 209–212 (1979)
- Paschmann, G., Schopke, N., Asbridge, J.R., Bame, S.J., Gosling, J.T.: Energization of solar wind ions by reflection from the Earth's bow shock. *J. Geophys. Res.* **85**, 4689–4694 (1980)
- Russell, C.T., Greenstadt, E.W.: Initial ISEE magnetometer results: shock observation. *Space Sci. Rev.* **23**, 3–37 (1979)
- Sonnerup, B.V.Ö.: Acceleration of particles reflected at a shock front. *J. Geophys. Res.* **74**, 1301–1303 (1969)

Received January 22, 1981; Revised Version September 9, 1981

Accepted September 22, 1981

In Memoriam

Georg Pfotzer

* 29. November 1909 † 24. Juli 1981



Am 24.7.1981 starb unerwartet der frühere Direktor des Max-Planck-Instituts für Aeronomie, Prof. Dr.-Ing. Georg Pfotzer.

Georg Pfotzer wurde am 29.11.1909 in Willstätt (Kreis Kehl) geboren. Er studierte Physik an der Technischen Hochschule Stuttgart. Für seine Dissertation zum Dr.-Ing. führte er erstmals Messungen der Richtungsabhängigkeit der Kosmischen Strahlung in der Atmosphäre aus. Er verwendete dazu eine von Ballonen getragene, automatisch registrierende Koinzidenz-Apparatur. Dabei erklärte er 1935 das nach ihm benannte „Pfotzer-Maximum“ der Strahlungsintensität in großer Höhe durch Sekundärprozesse in der Hochatmosphäre. Mit dieser Entdeckung hat sich G. Pfotzer in das Buch der Geschichte der Kosmischen Strahlung eingeschrieben. Dem Studium folgte von 1936 bis 1945 eine Tätigkeit im Forschungslaboratorium der Firma Siemens & Halske in Berlin. Er trug dort wesentlich zur Entwicklung

und Verbesserung der ersten Generation von großtechnisch einsetzbaren Halbleitern, vor allem auf der Basis von Kupferoxydul und Selen bei. Nach dem Kriege nahm er 1947 eine Stelle als wissenschaftlicher Mitarbeiter in der Forschungsstelle für Physik der Stratosphäre an, die sein ehemaliger Lehrer, Prof. E. Regener, in Weissenau bei Ravensburg gegründet hatte. Dort veröffentlichte er mehrere Arbeiten über die Kosmische Strahlung und entwickelte einen Neutronenmonitor für das Internationale Geophysikalische Jahr. Die Forschungsstelle wurde 1952 in ein Max-Planck-Institut umgewandelt, 1955 mit dem Max-Planck-Institut für Ionosphärenforschung in Lindau vereinigt und trägt seit 1957 den Namen Max-Planck-Institut für Aeronomie. Nach einem Aufenthalt als Gastprofessor an der University of California in Berkeley, USA, wurde G. Pfotzer 1960 zum wissenschaftlichen Mitglied des Max-Planck-Instituts für Aeronomie berufen. Im Jahre 1965 wurde er zum Direktor am Max-Planck-Institut für Aeronomie ernannt, ab 1967 war er Honorarprofessor an der Technischen Universität Braunschweig. Von 1975 bis zu seiner Emeritierung 1977 leitete er das Max-Planck-Institut für Aeronomie als Geschäftsführender Direktor.

Während seiner Tätigkeit in Lindau hat sich G. Pfotzer weitgehend der damals jungen Weltraumforschung zugewandt. Er baute eine Arbeitsgruppe zur Erforschung der Solaren Kosmischen Strahlung auf und begann in Europa als erster mit der Messung von Röntgenbremsstrahlung in der Polarlichtzone zur Untersuchung magnetosphärischer Vorgänge.

Später beteiligte sich die Arbeitsgruppe an den ersten Raketenkampagnen in Europa und leistete wesentliche Beiträge zum ersten deutschen Forschungssatelliten AZUR. G. Pfotzer und seine Arbeitsgruppe waren auch an vielen der nachfolgenden nationalen und internationalen Satelliten- und Raumsondenmissionen zur Untersuchung der Magnetosphäre der Erde und ihrer Wechselwirkung mit dem Sonnenwind maßgeblich beteiligt. Genannt seien hier nur die wissenschaftlich so erfolgreichen Projekte HELIOS und GEOS.

Die hohe Kompetenz und die allseits anerkannte Unparteilichkeit von G. Pfotzer führten dazu, daß er immer wieder in Beratungsgremien, vor allem des Bundes und der Europäischen Weltraumorganisation, berufen wurde. Trotz dieser vielfältigen beratenden und organisatorischen Aufgaben blieb er bis über seine Emeritierung hinaus wissenschaftlich tätig.

Georg Pfotzer hatte viele Freunde im In- und Ausland. Sie werden ihn als vorbildlichen Kollegen im Gedächtnis bewahren, bei dem sich hervorragende wissenschaftliche Begabung und Leistung mit beispielhafter Mitmenschlichkeit und Bescheidenheit verbanden.

H. Rosenbauer

Correction

J Geophys (1981) 49:192-197

**Non-Adiabatic Expansion of Low-Temperature Solar Wind
Radial Temperature Gradients**

A. Geranios

The addresses given under the name of the author and in the footnote respectively should be exchanged (as a matter of fact the work has been done at the Max-Planck-Institut für Aeronomie, D-3411 Katlenburg, Lindau 3, FRG).

Received September 8, 1981

Book Review

Applied Inverse Problems, edited by P.C. Sabatier. Lecture Notes in Physics 85, 425 S., Springer, Berlin Heidelberg New York, 1978

The book evolved from a workshop on applied inversion problems organized by the CNRS in Montpellier 1978. The lecture notes deal with quite different aspects of applied inverse problems in various scientific fields, whenever a number of local parameters of a physical system must be deduced from a set of measured data such as geophysics, optics, quantum mechanics, particle field theory.

After an introduction to general aspects of inverse theory and an overview of the methods of inversion and its fields of application about 20 lecture notes on applied but as well on theoretical inverse problems are reprinted. 6 of them deal with geophysical applications in seismology, gravity, magnetotelluric and electromagnetic prospection. These include mostly linear inversion methods, the properties of which are fairly well understood up to now but nonlinear aspects are treated as well, i.e., the determination of seismic foci and of the electric conductivity. As the lectures are given by authors working in the respective fields, they are highly specified. The same is true for the useful understanding of the other lectures mostly treating scattering problems in quantum mechanics and particle field theory. Finally some lectures deal with the mathematical aspects of inverse theory (e.g., the search for solutions of integral and differential equations arising mainly in these fields outside geophysics).

In conclusion, I can say that the book gives a good overview of applied (and theoretical) inverse problems for the scientist working in one of these fields and who is willing to overclimb the high mathematical cliffs which encompass this rather new research field. For the geophysicist who wants to know something about inversion there are a number of publications treating this problem at a lower and maybe more understandable level, lots of them are referenced in this book. Some of the geophysical lectures given here are already published elsewhere so that the geophysicist who wants to do inversion does not get much additional information from this book.

Manfred Koch

Cosmo- and Geochemistry

1981. 39 figures, 18 tables. Approx. 140 pages.

(Topics in Current Chemistry, Volume 99)

Cloth DM 68,-; approx. US\$ 30.00

ISBN 3-540-10920-X

Distribution rights for all socialist countries: Akademie-Verlag, Berlin

Contents: E. Anders, R. Hayatsu: Organic Compounds in Meteorites and Their Origins

C1 and C2 carbonaceous chondrites contain organic matter, mainly as a bridged aromatic polymer containing COOH, OH, and CO groups, as well as heterocyclic rings containing N, O, and S. The remaining 5.3% includes compound classes such as alkanes, alkenes, arenes, alicyclics, alcohols, aliphatic carboxylic acids, purines, pyrimidines, and other basic N compounds, amino acids, porphyrin-like pigments, and carbynes. This article discussed the formation of these compounds in the solar nebula by catalytic (Fischer-Tropsch) reactions of CO, H₂, and NH₃ at 360400 K, ~ 10⁻⁵atm. Such reactions are shown to be a possible source of prebiotic carbon compounds on the inner planets and interstellar molecules.

G. Winnewisser: The Chemistry of Interstellar Molecules

This article reviews the substantial new information molecular line astronomy has contributed over the last 10 years on a variety of long-standing astronomical problems such as (i) the distribution of matter within our galaxy and other galaxies, (ii) the mass loss associated with old and young stars, (iii) the formation of young stars occurring in dense molecular clouds, and (iv) the determination of isotope ratios as a useful indicator of the past chemical history of the galaxy. Molecular line astronomy has also opened several new areas of research, thus widening our understanding of the cosmos. The most fascinating of these new research activities certainly seems to be cosmochemistry, the chemistry prevailing in the highly diluted medium of space as compared to terrestrial conditions. Although some of the organic interstellar molecules are by far not as complex as biologically important molecules such as desoxyribonucleic acid or ribonucleic acid or which govern most basic properties of life reproduction, they are, however, fairly exotic even when compared with present day organic chemistry carried out in the most advanced Laboratories.

R. Lüst: Chemistry in Comets

About 100 years ago, photography and spectroscopy opened a new area in astronomy, making it possible to attempt the first crude chemical analysis of the comets. As early as 1911, K. Schwarzschild and E. Kron explained emissions by resonance fluorescence of the solar radiation from various molecules. Studies have since yielded the presence of the elements H, C, O, and N. New results on the chemistry of comets are summarized in this review.

V. Marchig: Marine Manganese Nodules

Marine manganese nodules are black to dark brown concretions lying on the sea floor. The chemical composition of the concentric growth zones around the nucleus shows that they are built of hydrated manganese and iron oxides (predominantly Mn⁴⁺ and Fe³⁺ in different ratios. Manganese and iron oxides incorporate some trace metals like Cu, Ni, Co, and Zn. The content of Cu and Ni can to make exceed 2% manganese nodules economically interesting.

Springer-Verlag
Berlin
Heidelberg
New York



6065/4/2h

Electrical Processes in Atmospheres

Edited by
H. Dolezalek and R. Reiter

Proceedings of the 5th International Conference on Atmospheric Electricity held at Garmisch-Partenkirchen.

1977, 865 pages, 478 figures, 42 tables.
Hardcover DM 350,-
ISBN 3-7985-0435-0

With 127 contributions on such fields as Basic Research in Ions - Atmospheric Electricity and Meteorology - Principles and Problems of Instrumentation - Application of Atmospheric Electricity - Cloud Electricity - Thunderstorms and Showers - Global Circuit - Solar-Terrestrial Effects - Atmospheric Electricity on Planets and Moon - Physics of Lightning and Sferics - Recent Research Problems - and a Subject Index with 3800 entries.

"This volume provides those interested in atmospheric electrical processes with a contemporary basis of knowledge and thinking on this topic. Its completeness and its excellent indexes, plus its comments throughout on future research challenges, should be useful to scientists and engineers in a variety of related fields." (Bulletin of the American Meteorological Society).

If your library has not yet purchased a copy please place your order to your bookshop or to



Dr. Dietrich Steinkopff Verlag
Saalbaustrasse 12
D-6100 Darmstadt 11
F. R. Germany

Palaeomagnetic Constraints on Allochthony and Age of the Krol Belt Sequence, Garhwal Himalaya, India*

Chris T. Klootwijk¹ **, Arvind K. Jain² ***, and Rakesh Khorana³

¹ Research School of Earth Sciences, Australian National University, P.O. Box 4 Canberra A.C.T. 2600, Australia

² Department of Geology and Geophysics, University of Roorkee, Roorkee 247672, India

³ Geological Survey of India, Kashmir Circle, Rajbagh Srinagar, India

Abstract. Thermal demagnetization of 223 carbonate and sandstone samples from the Late Palaeozoic – Mesozoic Krol Belt sequence (Ganges section, Garhwal Lesser Himalaya) showed five magnetic components:

- a) a magnetically soft component of recent origin;
- b) a universally present secondary component of mixed polarity and of post-folding origin. Its NNW declination indicates an approximately 25° clockwise rotation of the sampled region with respect to peninsular Indo-Pakistan since Late Tertiary.
- c) another universally present secondary component of mixed polarity and of pre-folding origin and with a NNW declination which indicates a clockwise rotation of the sampled region over approximately 45° since Early Tertiary.
- d) an ENE to WSW directed secondary component of very shallow inclination and of mixed polarity, observed throughout the entire sample collection. This component may represent a low-temperature oxidation phase, aligned preferentially towards the foliation plane.
- e) primary magnetic components could be determined only in the basal part of the sequence studied, i.e. in the Lower Blaini Diamictite, in the Blaini Limestone and in the Krol-A Limestone. Comparison of their rotation-corrected pole positions with the Indo-Pakistan apparent polar wander path (APWP) supports the disputed Permo-Carboniferous to Permian age of these rocks. The high palaeolatitudes indicated by these primary components favour a glacial origin for the Blaini Diamictite.

Clockwise rotation of tectonic units within the Lesser Himalaya similar to the rotation described herein have been documented recently for the Kashmir nappe, and can be interpreted also from palaeomagnetic results from other regions of the NW Himalaya. If of regional consistency, such rotations may put constraints on the magnitude of intracontinental underthrusting along the Main Central Thrust (MCT).

Key words: Himalaya – Krol Belt – Palaeomagnetism – Tectonics

Introduction

Early Tertiary or earlier collision of Greater India with an island arc occurred at equatorial latitudes (Klootwijk 1981). The Indian plate has subsequently moved northwards over more than 3,000

km, with Eurasia undergoing hardly any latitudinal movement and only a slight clockwise rotation. This large-scale convergence has been accommodated not only by crustal shortening within South Central Asia (Molnar and Tapponnier 1975; Tapponnier and Molnar 1976), but also resulted in intracontinental underthrusting within extra-peninsular Indo-Pakistan and in telescoping of the sedimentary cover. The magnitude of underthrusting is a matter of debate, particularly after Powell and Conaghan (1973) revived Argand's (1924) hypothesis that Tibet's twice average crustal thickness represents doubling of the crust through large-scale underthrusting by Indo-Pakistan's continental lithosphere. This hypothesis conflicts with the established notion that buoyancy constraints would preclude large-scale continental subduction (McKenzie 1969; Dewey and Bird, 1970). Recent surface wave (Knopoff and Chang 1981) and deep seismic sounding studies of the Tibetan Plateau (Teng-Ji wen 1981; Teng-Ji wen et al. 1981) have been interpreted, however, in support of doubling of the crust. Estimates of thrust magnitudes within the Lesser Himalaya (Andrieux personal communication 1981) and along the MCT (Gansser 1974; Le Fort 1975) indicate a minimal underthrusting of 300–500 km.

The India-Asia convergence occurred nearly completely in a north-south direction. Palaeomagnetism can therefore be most usefully applied to quantifying and dating such a large-scale underthrusting, and on a more regional scale to determining the magnitude of rotational movement of individual thrust units. This formed the main objective for our ongoing palaeomagnetic project in extra-peninsular Indo-Pakistan (see reviews by Klootwijk 1979, 1981, Klootwijk and Radhakrishnamurty 1981). A second and closely related objective concerned further refinement of the Indo-Pakistan APWP and determination of its poorly constrained Palaeozoic and Tertiary segments. Rocks of those eras are not well developed in peninsular Indo-Pakistan. The present reconnaissance study of the Krol Belt region in the north-western Himalaya has been carried out as part of this project, with as specific objectives:

- 1) To constrain the age range of the Krol Belt sequence through comparison with well-dated palaeomagnetic results from peninsular Indo-Pakistan.
- 2) To obtain more data on the controversial autochthonous or allochthonous position of the Krol Belt sequence.
- 3) To establish the palaeolatitude of deposition of the Blaini Diamictites, whose glacio-marine origin has been disputed.

Regional Geology and Sampling

Since the pioneering investigations of Auden (1934, 1937) in the Lesser Himalaya of Garhwal, who regarded their structure

* IGP Contribution no. 577

** *Present address.* Institute de Physique du Globe de Paris and U.E.R. des Sciences Physiques de la Terre, Universités Paris 6 and 7, 4 Place Jussieu, 75230 Paris Cedex 05, France

*** *Present address.* Geologisches Institut, Universität Karlsruhe, 7500 Karlsruhe 1, Federal Republic of Germany

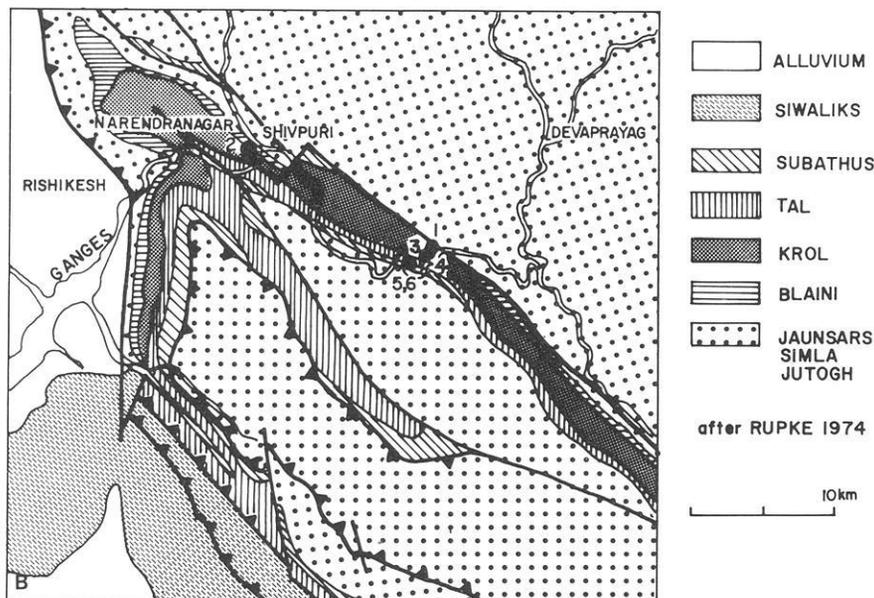
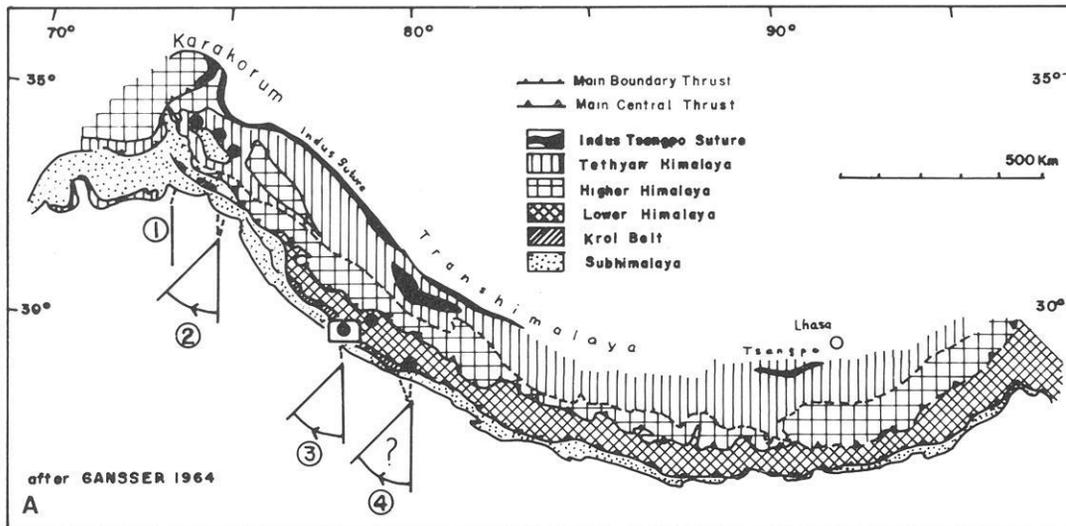


Fig. 1. A Main structural features of the Himalaya after Gansser (1964). The apical angle and arrow of the halfcones indicate magnitude and sense of rotation with respect to peninsular Indo-Pakistan (see text). 1 = NW Kashmir (Klootwijk 1979), 2 = Central and Eastern Kashmir (Klootwijk 1979), 3 = present study, 4 = Kumaon Lesser Himalaya (Athavale et al. 1980). The inset shows the location of B: a geological sketchmap of the sampled region after Rupke (1974)

as a telescoped pile of large-scale thrusts, controversies still persist regarding the chronostratigraphy and tectonics of this very complex region. A nearly complete absence of fossil records from these formations and a profusion of lithostratigraphical correlation schemes have hindered interpretation of the complicated tectonics. It is not surprising, therefore, that controversies have arisen on the regional tectonics and, in particular, on the allochthony of the Krol Belt sequence (Krol Nappe). Likewise, there is as yet no universal acceptance of a glacial origin for the Blaini diamictites of the Krol Belt sequence. This poses problems for dating the younger sequence of the Krol Belt, as the lithological correlation of the Blainis with the Permo-Carboniferous Talchir beds of peninsular Indo-Pakistan forms a keystone in dating the Krol Belt sequence. A probable Late Palaeozoic-Mesozoic age for this sequence is still contested.

It is beyond the scope of this paper to review the abundant literature and multitude of controversies on the Krol Belt, and we confine ourselves here to a description of the structural and stratigraphical outlines of the Krol Belt region as far as is relevant to our palaeomagnetic study.

The Garhwal Lesser Himalaya are made up of a pile of

southwards directed nappes which are of higher metamorphic grade in the tectonically higher units (Fig. 1a, b). The cryptic roots of these thrusts probably emanate from the Main Central Thrust (MCT) in Garhwal and elsewhere in Himachal Pradesh and Nepal (Ravi Shankar and Ganesan 1972; Ashgiri et al. 1977; Andrieux et al. 1980; Brunel and Andrieux 1980). The main nappe units northwards and upwards of the autochthonous Siwaliks of the sub-Himalaya are: the Krol Nappe, mainly a lower greenschist metamorphism, and the Almora Nappe. The allochthony of the Krol Nappe was disputed by Ranga Rao (1968, 1970), Rupke (1974) and Fuchs and Sinha (1978), who interpreted the folded Krol-Tons thrustplane dissecting a basement of Simla slates and the unconformably overlying Subathu Formation as a pre-Eocene unconformity. Consequently, they doubted the tectonic window structures beneath the eroded parts of the Krol Belt, near Bhidalna and Pharat (Auden 1934, 1937, 1970, Jain 1972). The majority of workers favour, however, an allochthonous origin for the Krol Belt sequence (Auden 1934, 1937, 1970; Gansser 1964; Bhattacharya and Niyogi 1971; Jain 1972; Bhargava 1972; Valdiya 1973a, 1977; Kumar and Agarwal 1975; Ashgiri et al. 1975; Acharya and Shah 1975; Singh

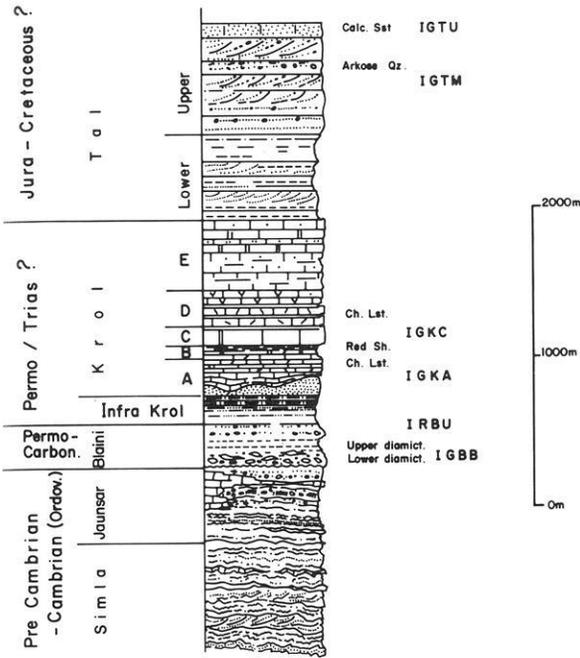


Fig. 2. Stratigraphical ideal column of the Krol Belt region after Gansser (1964)

1975a; Geological Survey of India, 1976; Ashgiri 1977; Sinha 1980; Prashra 1980)

Understandably, neither the magnitude nor the date of displacement of the Krol Nappe have been established accurately, though Auden (1937) and Prashra (1980) suggest minimal southward movements of 50 km and 60 km respectively. The youngest thrust movements of the Krol Nappe are of Plio-Pleistocene to Quaternary age as indicated by overriding of the Krol Belt sequence on the Siwaliks and Doon gravels of the sub-Himalaya. The general overthrusting of the Krol Nappe over an Upper Proterozoic to Lower Palaeozoic basement of low-grade metasediments of the Simla Group and unconformably overlying Tertiary sediments of the Subathu and Dagshai Formations (Figs. 1 b, 2) indicates, according to Auden (1937), that the major south to southwestwards directed thrust movements and the subse-

quent folding of the thrustplane occurred during post-Burdigalian time and most probably during the Early Miocene diastrophic phase of the Himalayan orogeny.

The base of the Krol Nappe is formed locally by the Jaunsars which are overlain unconformably by a Late Palaeozoic-Mesozoic succession comprising the Blaini Formation and the actual Krol Belt Formations, i.e. the Infra-Krol, the Krol limestone and shale and the Tal sandstone and limestone (Figs. 1 b, 2).

The Blaini Formation is characterized by two main rock types, i.e. two diamictite horizons (lower horizon sampled, Table 1) and an upper succession of siliceous dolomitic purple limestone and silicified purple sandy shale (sampled, Table 1). The tillitic aspect of the diamictite has been disputed by Bhattacharya and Niyogi (1971), Valdiya (1973b), and Rupke (1974) who stressed the turbiditic character of the Blaini beds. The majority of recent sedimentological and geochemical studies have brought forth evidence for a glacial and most probably a glaciomarine origin of the Blaini Formation (Gaur 1971; Gaur and Dave 1974; Bhargava and Bhattacharya 1975; Singh 1975b; Bhatia and Prasad 1975, 1981; Ahmed 1975; Jain and Varadaj 1978; Jain in press 1981). The Blaini Formation is overlain by slate and sandstone of the Infra-Krol Formation (not sampled), which is succeeded by an up to 1,500 m thick succession of limestone and calcareous shale of the Krol Formation which is generally assumed to be of Permo-Triassic age. From the five members of the Krol Formation (A-E), only thin-bedded limestone of the Krol-A member and more massive dolomitic limestone of the Krol-C member were selected for sampling. The Tal Formation of a presumed Jurassic-Cretaceous age forms the highest lithostratigraphic unit of the Krol Belt. Sampling was carried out in white to pinkish coloured arkosic sandstone and shelly sandy limestone from the Upper Tal member (Table 1).

Treatment and Results

A total of 223 samples from the Blaini-Tal succession were obtained with a portable drill along the Ganges profile in the northeastern limb of the Lansdowne Syncline (Fig. 1 b). The beds have a steeply southward to locally vertical dip, which facilitated sampling of stratigraphic profiles. Generally one sample was taken per bed (except for the Lower Blaini Diamictite). All samples were oriented with a magnetic and a solar compass.

Table 1. Sampling details Krol Belt region Ganges river section (30.2° N 78.4° E)

Formation	Locality (Fig. 1)	Age	Lithology	Section length	Samples (specimens)	Init. intensity 10^{-3} m. Am ⁻¹	Bedding Strike (o)/ Dip (o)
Upper Tal Lst.	6	Jurassic-Cretaceous	oolitic arenaceous limestone	1 m	16 (21)	545- 3,725	100/ 54 S
Middle Tal Sst.	5	Jurassic-Cretaceous	pinkish to white sst., occ. cross-bedding	6.5 m	24 (25)	139- 2,290	95/ 55 S
Krol-C Lst.	4	Permo-Triassic?	grey thin-bedded dolomitic limestone	4.5 m	37 (68)	42- 1,763	134-140/ 67.5-79.5 SW
Krol-A Lst.	3	Permo-Triassic?	limestone, calcareous shale	15.5 m	40 (62)	29- 389	120-125/ 85 S
Blaini Lst.	2	Late Carboniferous-Early Permian	purple dolomites silicified red and grey sandy shales	14 m	62 (79)	46-102,964	83-93/ 73-81 S
Lower Blaini Diamictite	1	Late Carboniferous-Early Permian	grey boulder beds	11.5 m	34 (51)	88- 3,628	140-145/ 92-97 SW

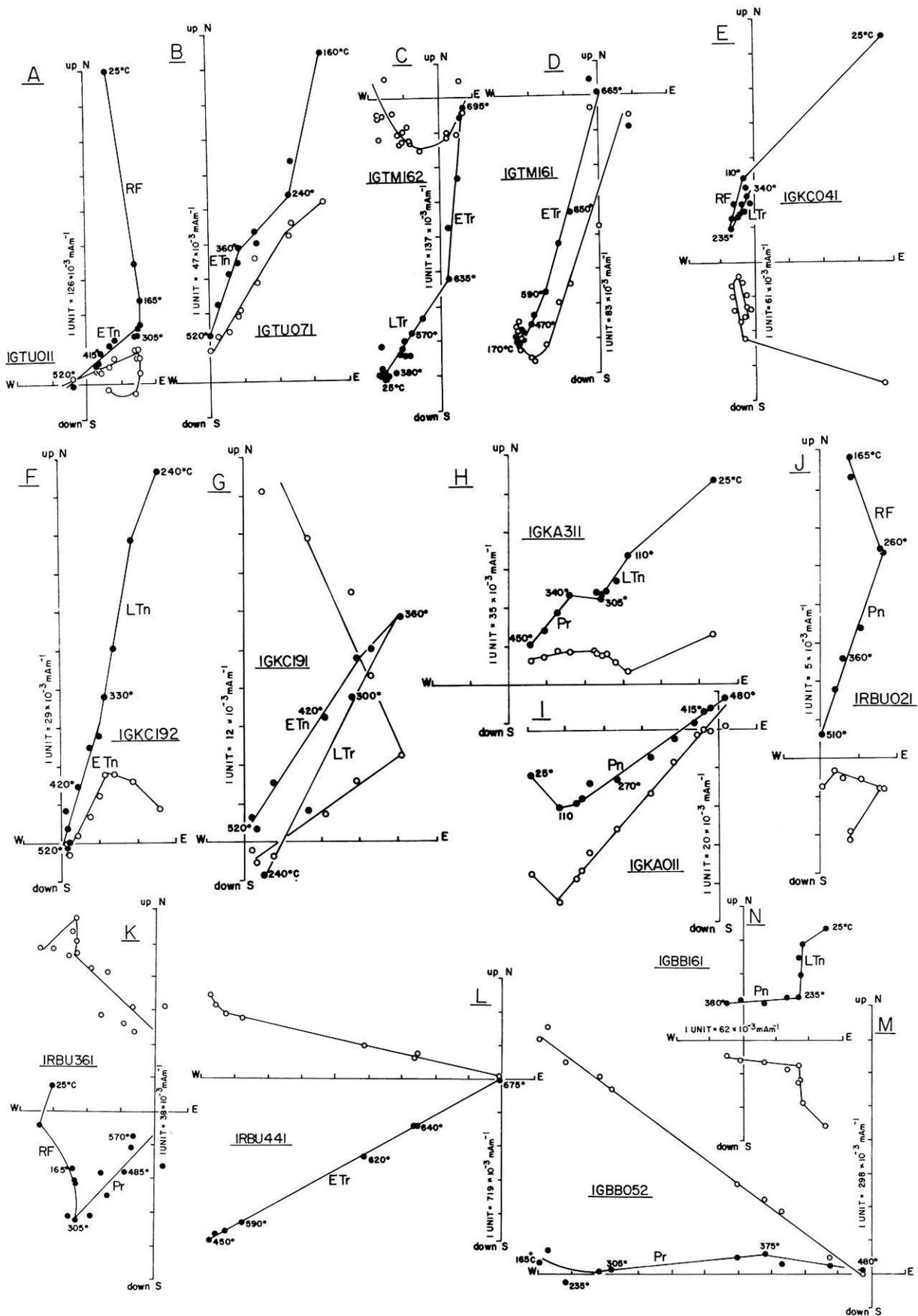


Fig. 3. Demagnetization diagrams of some representative specimens. The points denote successive positions – in orthogonal projection – of the end points of the resultant magnetization vector during progressive thermal demagnetization. *Open circles* denote projections on the vertical east-west plane, *dots* denote projections on the horizontal plane. *Numbers* denote successive peak temperature values. RF=recent field component, LT=Late Tertiary secondary component, ET=Early Tertiary secondary component, P=Primary component, n=normal polarity, r=reversed polarity. Results are not corrected for bedding. See Fig. 2 for sample codes

Table 2. Mean palaeomagnetic results, Krol Belt

Formation	Direction						Pole position (southpoles)					E ₉₅ Palaeolat ^a	No Fig. 5, 6
	Decl (o)	Incl (o)	K	α_{95} (o)	N	Polarity	Lat (o)	Long (o)	dp (o)	dm (o)			
<i>A) Recent field component</i>													
Upper Tal lst.	355	+41	31.5	10	8	N							1
Blaini Lst.	346.5	+42	21.5	9	13	N							2
<i>B) Late Tertiary secondary component, post-folding</i>													
Middle Tal sst.	203	-31.5	11	14.5	11	R	65 S	15.5 E	9	16.5			3
Krol-C Lst.	203.5	-40.5	34.5	3.5	50	R	68 S	1.5 E	2.5	4			4
Krol-C Lst.	18.5	+39	21.5	7.5	18	N	71.5 S	10 E	5.5	9			5
Krol-C Lst.	21.5	+40	29.5	3	68	N+R comb.	69.5 S	4.5 E	2	3.5			6
Krol-A Lst.	12.5	+43.5	8	12.5	20	N (+R:2)	78 S	9 E	9.5	15.5			7
Blaini Lst.	4	+30	9.5	10	25	N (+R:4)	75.5 S	63 E	6	11			8
Lower Blaini Diamictite	201.5	-40	22.5	14.5	6	R	69.5 S	4.5 E	10.5	17.5			9
Lower Blaini Diamictite	12	+44	26.5	8.5	12	N	78.5 S	8 E	6.5	10.5			10
Lower Blaini Diamictite	15	+42.5	25	7	18	N+R comb.	75.5 S	7.5 E	13.5	22			11
Mean pole	16	+38	97	7	5		∞73 S	14 E			6°		A
<i>C) Middle to Early Tertiary secondary component, pre-folding</i>													
Upper Tal lst.	29	+16	13	10	19	N	55 S	21.5 E	5.5	10.5			12
Middle Tal lst.	197	-28.5	11	12.5	14	R	68.5 S	28.5 E	7.5	13.5			13
Middle Tal lst.	18	+17	13.5	11	15	N	62.5 S	36.5 E	6	11.5			14
Middle Tal lst.	18.5	+22	11.5	8	29	N+R comb.	64.5 S	32 E	4.5	8.5			15
Krol-C Lst.	27.5	+34.5	18.5	5	46	N(+R:1)	62.5 S	6.5 E	3.5	5.5			16
Krol-A Lst.	27	+27	40	9	8	N	60.5 S	15 E	5.5	10			17
Blaini Lst.	233.5	-20.5	5	13.5	27	R	36.5 S	12 W	7.5	14			18
Blaini Lst.	25.5	+18	19.5	9.5	14	N	58.5 S	24.5 E	5	10			19
Blaini Lst.	224	-17	8	8.5	41	N+R comb.	43.5 S	6.5 E	4.5	9			20
Lower Blaini Diamictite	15	+15.5	13.5	26	4	N(+R:1)	63.5 S	43 E	13.5	26.5			21
Mean pole	27	+22.5	45	10	6		59.5 S	19.5 E			8.5		B
<i>D) Primary components</i>													
Krol-A Lst.	115	+71.5	10	18.5	8	R	12.5 N	109.5 E	28.5	32.5			22
Krol-A Lst.	317	-61	6	30	6	N	7.6 S	109 E	35	46			23
Krol-A Lst.	306.5	-67.5	8	15	14	N+R comb.	3.5 N	109.5 E	21	25			24
Blaini Lst.	89.5	+77	6.5	9	47	R	27.5 N	106.5 E	15.5	16.5			25
Blaini Lst.	270	-67	8.5	14	15	N	22.5 N	123 E	19	23			26
Blaini Lst.	269.5	-74.5	7	7.5	62	N+R comb.	26.5 N	116 E	12.5	13.5			27
Lower Blaini Diamictite	187.5	+62	8.5	16.5	11	R	16.2 S	72.5 E	20	25.5			28
Lower Blaini Diamictite	321	-79	7	18	12	N	13 N	92 E	32.5	34			29
Lower Blaini Diamictite	354	-72	7	12	23	N+R comb.	2.5 S	81.5 E	18.5	21			30
<i>E) Primary components, corrected for inferred rotation 45° clockwise</i>													
Krol-A Lst.	261.5	-67.5	8	15	14	N+R comb.	28 N	124 E	21	25	50.5S ^{33S} _{75S}		24'
Blaini Lst.	224.5	-74.5	7	7.5	62	N+R comb.	47.5 N	108.5 E	12.5	13.5	61 S ^{50S} _{74.5S}		27'
Lower Blaini Diamictite	319	-72	7	12	23	N+R comb.	4 N	99.5 E	12.5	19	57 S ^{41S} _{78S}		30'
<i>F) East-west directed secondary component, all formations combined</i>													
Before bedding correction	289.5	- 8.5	4.5	10.5	54	N+R comb.							31
After bedding correction	286.5	- 1.5	7.5	7.5	54	N+R comb.							32

^a Upper and lower limits corresponding to the apex of the 95% cone of confidence

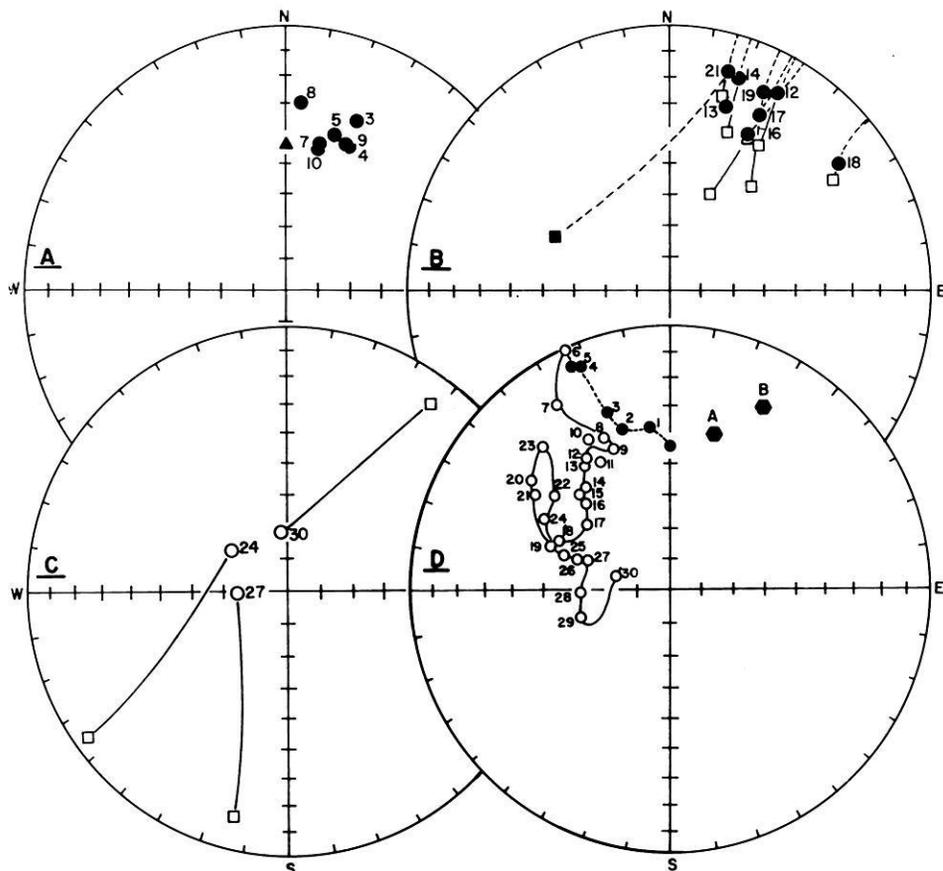


Fig. 5A-D. Formation mean directions not corrected for bedding. **A** For the secondary Late Tertiary component. The *triangle* represents the present field direction at the sampled region. **B** for the secondary Early Tertiary component. *Squares* indicate directions prior to bedding correction and *circles* thereafter; **C** for the primary component. Legend as for **B**. **D** expected directions in the sampled region according to palaeomagnetic data from the Indian plate: 1=DSDP cores (20 my), 2=DSDP cores (30 my), 3=DSDP cores (40 my), 4=DSDP cores (50 my), 5=Sanjawi Lst. (Pal.-Eoc.), 6=Brewery Lst. (Palaeoc.), 7=DSDP cores (60 my), 8=Deccan Traps, upper normal epoch (>60-65 my), 9=Deccan Traps, lower reversed epoch (>60-65 my), 10, 11=Tirupatti Sst. (K1), 12=DSDP cores (70 my), 13=Satyavedu beds (K1), 14=Sylhet Traps (K1??), 15, 16=Goru Fm. - Parh Lst. (K1-u), 17=Rajmahal Traps (>100-105 my), 18=Loralai Lst. (Jm), 19=Chiltan Lst. (Jm-u), 20, 21, 22=Transferred directions for the mean pole positions from Australia (Jm), Australia (Jl) and Antarctica (Jm) respectively. 23=Parsora beds (Tru), 24=Pachmarhi beds (Tru), 25=Kamthi beds (Pu-Trl), 26=Mangli beds (Pu-Trl), 27=Panchet beds (Pu-Trl), 28=Kamthi beds (Pu-Trl), 29=Talchir beds (P-C), 30=Alozai Fm. (P-C). Hexagons *A* and *B* indicate the mean formation directions for the Late- and Early Tertiary secondary components respectively. All figures are in equal area projection. Full symbols denote downwards pointing directions and open symbols denote upwards pointing directions

observations on Precambrian glacial successions in particular, which are accompanied by substantial dolomitic sequences (Bhatia and Prasad 1981).

It is open to further study to what extent this approximately 45° clockwise rotation may be characteristic also for other regions of the Lesser Himalaya. The few palaeomagnetic results so far available, indicate that similar rotations can indeed be interpreted for other parts of the Lesser Himalaya (Fig. 1a). A similar clockwise rotation by about 45° of the Panjal nappe has been concluded from secondary post-folding components observed in Kashmir (Klootwijk 1979). Athavale et al. (1980) recently reported results for the Permo-Carboniferous Durgapipal volcanics and for the Siluro-Devonian Rudraprayag volcanics from the Kumaon Lesser Himalaya, and interpreted them as indicative of autochthony of the sampled regions. These regions are situated, however, directly to the north of the Main Boundary Thrust (MBT) and in the region of the Almora Thrust respectively (Fig. 1b). The assumed primary origin of the observed magnetic directions is, moreover, open to some doubt in view of the degree of metamorphism of the sampled regions, which is at least of greenschist facies (Raina and Dungrakote 1975; Kumar

and Agarwal 1975). If these directions are indeed of primary origin, comparison of their pole positions with the Palaeozoic segment of the Indo-Pakistan APWP (Fig. 6b) can be interpreted in terms of a clockwise rotation of the sampled regions of a magnitude comparable to those observed in the present study and in Kashmir.

The importance of such clockwise rotations of the Lesser Himalayan nappes (Fig. 1a) for determination of the magnitude of intracontinental underthrusting of Greater India along the MCT (Powell and Conaghan 1973) remains to be studied. The similarity between the interpreted magnitude of rotations in this study and the azimuthal angle between the MBT and the reconstructed northern boundary of Greater India (Powell 1979; Fig. 7) may be more than fortuitous.

Conclusions

Comparison of Early and Late Tertiary secondary magnetic components from the Krol Belt region of the Garhwal Lesser Himalaya and the Indo-Pakistan APWP evidently indicates an approx-

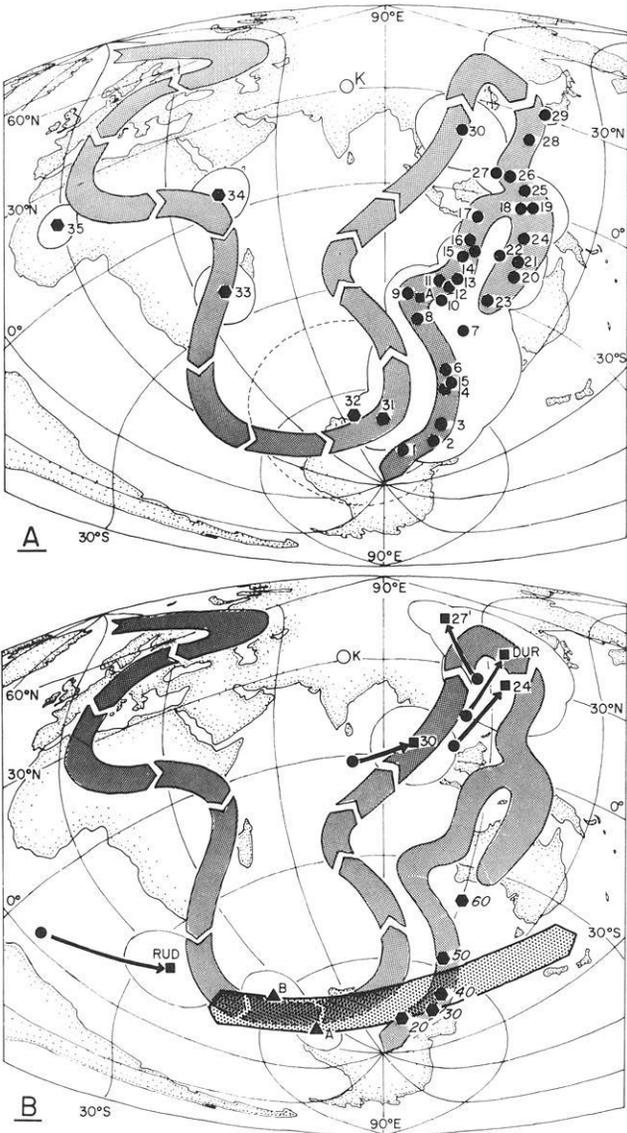


Fig. 6. A Phanerozoic APWP for Indo-Pakistan, based on (south) pole positions from the Indian plate (1–30, see legend to Fig. 5D) and transferred pole positions from south-east Australia (Goleby 1981; 31–35). 31, 32=Devonian, 33=Silurian (M–U), 34=Silurian (L–U), 35=Ordovician (M–U). A=Panjal Traps. The latter result may represent a secondary (McElhinny et al. 1978) or possibly a primary magnetization. **B** Comparison of the Indo-Pakistan APWP as shown in Fig. 6a with the primary (Table 2: 24, 27, 30) and secondary mean pole positions (Table 2: A, B) obtained here for the Krol Belt region. The loci of the secondary mean positions (A, B) intersect the Tertiary part of the Indo-Pakistan APWP (ages indicated in m.y.) in the Early and Late Tertiary segments respectively (see text). Their offset indicates a clockwise rotation of the sampled region with respect to peninsular Indo-Pakistan which has proceeded over approximately 45° since the Early Tertiary (compare Fig. 5D). The primary pole positions have been corrected for this rotation (24° , 27° , 30°). The effect of a similar correction on palaeopositions obtained by Athavale et al. (1980) for the Permo-Carboniferous Durgapal volcanics and the Siluro-Devonian Rudraprayag volcanics of the Kumaon Lesser Himalaya is shown for comparison. The hexagon (K) indicates the sampled region. Aitoff projection

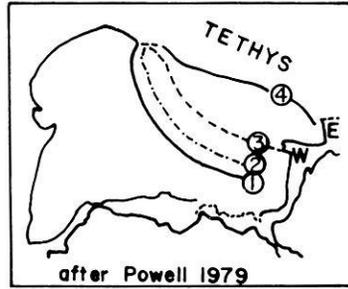


Fig. 7. A reconstruction of Greater India within Gondwana after Powell (1979). This figure displays the azimuthal orientation of the Main Boundary Thrust (I) with respect to the inferred outline of the maximum northern extent of Greater India. This extent is based on underthrusting of Greater India beneath Tibet up to the present Kun Lun – Astin Tagh – Nan Shan mountain front. Within a reconstructed East Gondwana this outline leads up to the Exmouth Plateau (E) off western Australia (4). 2=Indus-Tsangpo suture zone, 3=Inferred minimum northern extent of Greater India in an arc parallel to 2 and leading up to the Wallaby Plateau (W)

imately 45° clockwise rotation of the sampled region. This finding clearly supports the allochthonous nature of the Krol Nappe.

Comparison of the pole positions for the primary magnetic components of the Lower Blaini Diamictite, the Blaini Limestone and the Krol-A Limestone, with the Indo-Pakistan APWP supports a Permo-Carboniferous to Permian age of these formations. This strengthens the lithological correlation of the Blaini Formation with the Talchir boulder beds of peninsular India as has also been pointed out (Jain and Thakur 1975; Jain in press 1981) for other Late Palaeozoic diamictites. The observed high palaeolatitudes further support the glacial origin of the Blaini diamictites.

The presently emerging pattern of clockwise rotations in the northwestern Lesser Himalaya has to be tested further for consistency, as it may form an important constraint in determining the magnitude of intracontinental underthrusting along the MCT.

Acknowledgments. Thanks are due to Shri V.K.S. Varadan and Shri M.N. Sehgal for provision of logistic support. Professor R.S. Mithal and Dr. M.W. McElhinny are thanked for continuing support and encouragement throughout this study. Part of this study was carried out during a post-doctoral fellowship (A.K.J.) from the Alexander von Humboldt Foundation.

References

- Acharyya, S.K., Shah, S.C.: Biostratigraphy of the marine fauna associated with diamictites of the Himalayas. *Bull. Indian Geol. Assoc.* **8**, 9–23, 1975
- Ahmed, N.: The origin and age of the Blaini Boulder Beds. *Bull. Indian Geol. Assoc.* **8**, 204–208, 1975
- Andrieux, J., Arthaud, F., Brunel, M., Sauniac, S.: Le caractère post-métamorphique des grands chevauchements himalayens dans l'Himalaya du Nord-Ouest. Aperçu sur les relations possibles avec le métamorphisme inverse du Nepal. *C.R. Acad. Sci. Paris.* **291**, Sér. D, 525–528, 1980
- Argand, E.: La tectonique de l'Asie. *Comptes Rendus 13ème Congr. Geol. Int.* 171–372, 1924
- Ashgirei, G.D., Sinha, A.K., Raaben, M.E., Dimitrenko, O.B.: New findings on the geology of Lower Himalaya, Himachal Pradesh, India. *Chayanica Geologica* **1**, 143–151, 1975
- Ashgirei, G.D., Sinha, A.K., Pande, I.C., Mallik, B.C.: A contribution to the geology, geochronology and history of regional metamorphism of Himachal Himalaya. *Himal. Geol.* **7**, 102–117, 1977

- Athavale, R.N., Rao, G.V.S.P., Rao, M.S.: Palaeomagnetic results from two basic volcanic formations in the western Himalayas and a Phanerozoic polar wandering curve for India. *Geophys. J.R. Astron. Soc.* **60**, 419–433, 1980
- Auden, J.B.: The geology of the Krol Belt. *Rec. Geol. Surv. India.* **69**, 357–354, 1934
- Auden, J.B.: The structure of the Himalaya in Garhwal. *Rec. Geol. Surv. India* **71**, 407–433, 1937
- Auden, J.B.: Comment "On the Krol Nappe hypothesis" by Ranga Rao. *J. Geol. Soc. India* **11**, 288–294, 1970
- Bhargava, O.N.: A reinterpretation of the Krol Belt. *Himal. Geol.* **2**, 47–81, 1972
- Bhargava, O.N., Bhattacharya, B.K.: The Blaini Formation of Himachal Pradesh and Uttar Pradesh. *Bull. Indian Geol. Assoc.* **8**, 71–99, 1975
- Bhatia, M.P., Prasad, A.K.: Some sedimentological, lithostratigraphic and genetic aspects of Blaini Formation of parts of Simla Hills, Himachal Pradesh, India. *Bull. Indian Geol. Assoc.* **8**, 162–185, 1975
- Bhatia, M.R., Prasad, A.K.: Evolution of the Blaini sedimentation in Simla Hills, Lesser Himalaya, India. *Neues Jahrb. Geol. Palaeontol.* **1981(5)**, 267–288, 1981
- Bhattacharya, S.C., Niyoga, D.: Geologic evolution of the Krol Belt in Simla Hills, H.P. *Himal. Geol.* **1**, 178–212, 1971
- Brunel, M., Andrieux, J.: Sur un modèle explicatif du métamorphisme "inverse" Himalayen. *C.R. Acad. Sci. Paris.* **291**, Sér. D, 609–612, 1980
- Dewey, J.F., Bird, J.M.: Mountain belts and the New Global tectonics. *J. Geophys. Res.* **75**, 2625–2647, 1970
- Fuchs, G., Sinha, A.K.: The tectonics of Garhwal-Kumaun Lesser Himalaya. *Jahrb. Geol. Bundesanst. (Austria)* **121**, 219–241, 1978
- Gansser, A.: The Geology of the Himalayas. London, New York, Sydney: Interscience 1964
- Gansser, A.: Himalaya. Mesozoic-Cenozoic orogenic belts: Data for orogenic studies. *Geol. Soc. London Spec. Publ.* **4**, Spencer, ed., pp 267–278, 1974
- Gaur, G.C.S.: Sedimentary features of the interglacial sediments of the Blaini Series, Rishikesh, Garhwal Himalaya. *Himal. Geol.* **1**, 231–243, 1971
- Gaur, G.C.S., Dave, V.K.S.: Blaini tillites near Rishikesh and their origin. *J. Geol. Soc. India* **12**, 164–172, 1974
- Geological Survey of India: Geology of Dehra Dun – Mussoorie – Devprayag area, Uttar Pradesh Himalaya, pp 1–22, 1976
- Goleby, B.R.: Early Palaeozoic palaeomagnetism in south east Australia. Global reconstruction and the geomagnetic field during the Palaeozoic. In: *Advances in Earth Planet. Sci.*, **10**, McElhinny, Khramov, Ozima, Valencio, eds: pp 11–21, 1981
- Jain, A.K.: Structure of the Bidhalna-Pharar windows and Garhwal thrust unit, Garhwal U.P. *Himal. Geol.* **2**, 188–205, 1972
- Jain, A.K.: Stratigraphy, petrography and paleogeography of the Late Palaeozoic diamictites of the Lesser Himalaya. *Sediment. Geol.*, in press 1981
- Jain, A.K., Thakur, V.C.: Stratigraphy and tectonic significance of the Eastern Himalayan Gondwana Belt with special reference to the Ranjit Pebble Slate. *Bull. Indian Geol. Assoc.* **8**, 50–70, 1975
- Jain, A.K., Varadaraj, N.: Stratigraphy and provenance of Late Palaeozoic diamictites in parts of Garhwal Lesser Himalaya, India. *Geol. Rundsch.* **67**, 49–72, 1978
- Kirschvink, J.L.: The least-square line and plane and the analysis of palaeomagnetic data. *Geophys. J.R. Astron. Soc.* **62**, 699–718, 1980
- Klootwijk, C.T.: A summary of palaeomagnetic data from Extrapeninsular Indo-Pakistan and south central Asia: Implications for collision tectonics. In: *Structural Geology of the Himalaya*, Saklani, ed.: pp 307–360. Delhi: Today and Tomorrow Publ. 1979
- Klootwijk, C.T.: The India-Asia collision: A summary of palaeomagnetic constraints. In: *Geology, Geological History and origin of Qinghai-Xizang Plateau*, Liu Dong-sheng, ed.: pp 951–966. Beijing: Science Press 1981
- Klootwijk, C.T., Sharma, N.L., Gergan, J., Tirkey, B., Shah, S.K., Agarwal, V.: The extent of Greater India 2: Palaeomagnetic data from the Ladakh Intrusives at Kargil, northwestern Himalayas. *Earth Planet. Sci. Lett.* **44**, 47–64, 1979
- Klootwijk, C.T. Bingham, D.K.: The northern extent of Greater India 3: Palaeomagnetic data from the Tibetan Sedimentary Series, Thakkhola region, Nepal Himalaya. *Earth Planet. Sci. Lett.* **51**, 381–405, 1980
- Klootwijk, C.T., Radhakrishnamurty, C.: Phanerozoic palaeomagnetism of the Indian plate and the India-Asia collision. In: *Palaeoreconstruction of the continents*, McElhinny, Valencio eds. *Geodynamic series 2*, pp 93–105, 1981
- Klootwijk, C.T., Nazirullah, R., De Jong, K.A., Ahmed, H.: A palaeomagnetic reconnaissance of northeastern Baluchistan. *J. Geophys. Res.* **86**, 289–306, 1981
- Knopoff, L., Chang, F.S.: Upper Mantle Structure under the Tibetan Plateau. In: *Geology, Geological History and origin of Qinghai-Xizang Plateau*, Liu Dong-sheng, ed.: pp 627–632. Beijing: Science Press 1981
- Kumar, G., Agarwal, N.C.: Geology of the Srinagar – Nandprayag area (Alaknanda Valley), Chamoli, Garhwal and Tehri Garhwal districts, Kumaun Himalaya, Uttar Pradesh. *Himal. Geol.* **5**, 29–59, 1975
- Le Fort, P.: Himalayas: the collided Range. Present knowledge of the continental arc. *Am. J. Sci.* **275-A**, 1–44, 1975
- McElhinny, M.W., Luck, G.R., Edwards, D.: A large volume magnetic fieldfree space for thermal demagnetization and other experiments in palaeomagnetism. *Pure Appl. Geophys.* **90**, 126–130, 1971
- McElhinny, M.W., Cowley, J.A., Edwards, D.J.: Palaeomagnetism of some rocks from peninsular India and Kashmir. *Tectonophysics* **50**, 41–54, 1978
- McKenzie, D.: Speculations on the consequences and causes of plate motions. *Geophys. J.R. Astron. Soc.* **18**, 1–32, 1969
- Molnar, P., Tapponnier, P.: Cenozoic tectonics of Asia. Effects of a continental collision. *Science* **189**, 419–426, 1975
- Powell, C. McA.: A speculative tectonic history of Pakistan and surroundings: some constraints from the Indian Ocean. In: *Geodynamics of Pakistan*, Farah, De Jong, eds.: pp 5–24 Quetta: Geological Survey of Pakistan 1979
- Powell, C. McA., Conaghan, P.J.: Plate tectonics and the Himalaya. *Earth Planet. Sci. Lett.* **20**, 1–12, 1973
- Prashra, K.C.: Some recent observations on the tectonic framework of Western Himalaya. In: *Geologie des Chaînes Alpines issues de la Tethys*. Aubouin, Debelmas, Latreille, eds.: *Mém. Bureau des Recherches Géologiques et Minières.* 115, 153, 1980
- Raina, B.N., Dungrakote, B.D.: Geology of the area between Naini Tal and Champawat, Kumaun Himalaya, Uttar Pradesh, *Himal. Geol.* **5**, 1–27, 1975
- Ranga Rao, A.: On the Krol Nappe hypothesis. *J. Geol. Soc. India* **9**, 152–158, 1968
- Ranga Rao, A.: Reply to Auden's comments "On the Krol Nappe hypothesis" *J. Geol. Soc. India* **11**, 295–302, 1970
- Ravi Shankar, Ganesan, T.M.: A note on the Garhwal Nappe. *Himal. Geol.* **2**, 72–82, 1972
- Rupke, J.: Stratigraphic and structural evolution of the Kumaon Lesser Himalaya. *Sediment. Geol.* **11**, 81–265, 1974
- Singh, M.P.: A reappraisal of the geology of the Kalsi-Chakrata area, Uttar Pradesh and the problem of Blaini Mandhali correlation. *Bull. Indian Geol. Assoc.* **8**, 118–133, 1975a
- Singh, M.P.: Some observations on the Blaini Boulder Beds of Himachal Pradesh Himalaya. *Bull. Indian Geol. Assoc.* **8**, 134–142, 1975b
- Sinha, A.K.: Tectono-stratigraphic problems in the Lesser Himalayan zone of Simla region, Himachal Pradesh. In: *Stratigraphy and Correlations of Lesser Himalayan Formations*, pp 99–116. Delhi: Hindustan Publ. Co. 1980
- Tapponnier, P., Molnar, P.: Slip-line field theory and large-scale continental tectonics. *Nature* **264**, 319–324, 1976
- Teng Ji-wen.: Characteristics of geophysical field and continental plate tectonics of the Qinghai-Xizang Plateau and its neighbouring regions. In: *Geology, Geological History and origin of Qinghai-Xizang Plateau*, Liu Dong-sheng, ed.: pp 633–650. Beijing: Science Press 1981

- Teng Ji-wen, Xiong Shao-po, Sun Ke-zhong, Yin Zhou-xun, Yao Hung, Chen Li-fang, Mu tong, Lai Moing-hui, Wu Ming-chu, Su De-yuan, Wang Shao-zhou, Huang Wen-jian, Ou Ren-sheng, Hao We-cheng, Shao An-min.: Explosion seismic study for velocity distribution and structure of the crust and upper mantle from Damxung to Yadong of the Xizang Plateau. In: *Geology, Geological History and origin of Qinghai-Xizang Plateau*, Liu Dong-sheng, ed., pp 691–710. Beijing: Science Press 1981
- Valdiya, K.S.: Tectonic framework of India: A review and interpretation of recent structural and tectonic studies. *Geophys. Res. Bull.* **11**, 79–114, 1973 a
- Valdiya, K.S.: The Blaini Conglomerates of Himachal Pradesh and Garhwal Himalaya. In: *Recent researches in Geology*, Vol 1, V.K. Verma et al., eds: pp 61–81 Delhi: Hindustan Publ. Co. 1973 b
- Valdiya, K.S.: Structural set-up of the Kumaon Lesser Himalaya. *Ecologie et Géologie de l'Himalaya. Colloq. Int. CNRS.* **2**, 449–462, 1977
- Zijderveld, J.D.A.: AC-demagnetization of rocks: Analysis of results. In: *Methods in Palaeomagnetism*, Collinson, Creer, Runcorn, eds.: pp 254–286. Amsterdam: Elsevier 1967
- Zijderveld, J.D.A.: Palaeomagnetism of the Esterel rocks. Ph.D. Thesis, University of Utrecht 1975

Received October 1, 1981; Revised version December 21, 1981

Accepted December 22, 1981

Grain Size Effect on the Low-Temperature Oxidation of Titanomagnetite

Tadashi Nishitani* and Masaru Kono**

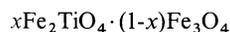
Geophysical Institute, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

Abstract. The low-temperature oxidation process in titanomagnetite has been investigated. Using samples with controlled grain sizes, it has been shown that there is a critical grain size in the oxidation behaviour of titanomagnetites; smaller grains undergo low-temperature oxidation, while larger ones separate to Fe-rich titanomagnetite and Ti-rich hemo-ilmenite (high-temperature oxidation). The difference between the results of similar experiments by Ozima and Sakamoto (1971) and by Readman and O'Reilly (1972) can be explained in terms of differences in particle sizes of titanomagnetites. Comparable results were obtained from natural subaerial and submarine basalts; when they are heated to temperatures between 150° C and 375° C, oxidation proceeds in submarine basalts, but it does not in subaerially erupted basalts and a high-temperature oxidation process occurs. This fact indicates that the grain size of the magnetic minerals is one of the most important controlling factors in low-temperature oxidation of titanomagnetites.

Key words: Rock magnetism – Low-temperature oxidation – Titanomagnetite – Grain size effects

Introduction

In igneous rocks, titanomagnetites

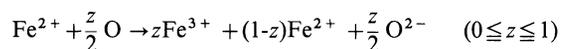


are often found to be the carriers of remanent magnetization and their magnetic properties have been extensively studied (e.g., Akimoto and Katsura 1959; Nagata 1961; Ozima and Sakamoto 1971; Readman and O'Reilly 1972). The magnetic properties of titanomagnetite, especially hysteresis properties, vary considerably due to grain size (Day et al. 1976, 1977). Levi and Merrill (1976) reported the grain size dependence of the intensity of anhysteretic remanent magnetization (ARM) and thermoremanent magnetization (TRM) in magnetite. The grain size of titanomagnetite may also affect the process of low-temperature oxidation. However, such effects have not been studied in detail in previous experiments.

The phenomenon of low-temperature oxidation was first investigated by Akimoto et al. (1957) and by Akimoto and Katsura (1959). They heated titanomagnetites in air at temperatures of

400°–500° C and studied magnetic properties of titanomagnetites and their changes with oxidation. However, it was later pointed out that they did not produce pure titanomaghemite as their samples contained high-temperature oxidation products (Ozima and Larson 1970). Sakamoto et al. (1968) demonstrated that titanomaghemites might be obtained if titanomagnetites were first ground in water and then heated in air at temperatures below 300° C. Readman and O'Reilly (1970) prepared samples by ball-milling sintered titanomagnetites in a water slurry for 100 h and obtained cation deficient titanomagnetites (titanomaghemites) by heating in air at temperatures between 200° and 300° C. As a result of the above experiments, it is generally thought that the experimental results can be trusted only if the samples are prepared by ball-milling or wet-grinding titanomagnetites.

Ozima and Sakamoto (1971) and Readman and O'Reilly (1972) carried out systematic studies of low-temperature oxidation. However, their results were not consistent with each other. Ozima and Sakamoto (1971) performed oxidation experiments on wet-ground titanomagnetites of high ulvospinel ratios ($x=0.7, 0.9$ and 1.0) and reported parameters such as lattice constants and Curie temperatures. Readman and O'Reilly (1972) used ball-milled samples with four compositions ($x=0.0, 0.4, 0.7$ and 1.0). The above two authors defined an oxidation parameter z , and in this study this parameter is also used. The definition of z is



$z=0.0$ stands for stoichiometric titanomagnetite and $z=1.0$ for fully oxidized state. As for the oxidation state reached under laboratory conditions, Ozima and Sakamoto (1971) concluded that $z \sim 0.3$ was the maximum after which the separation of titanomagnetites proceeded, while Readman and O'Reilly (1972) reported that titanomagnetites could be completely oxidized. These differences are clearly shown for $x=0.7$ titanomagnetite in Figure 1. In the results of Nishitani (1979), there is no hemo-ilmenite phase visible on the X-ray measurements. More recently, Keefer and Shive (1981) made Curie temperature and lattice parameter measurements on synthetic titanomaghemite samples. Their results are not very different from those of Ozima and Sakamoto (1971).

It appears necessary to re-examine the low-temperature oxidation scheme with special reference to the effect of grain size of the titanomagnetites. In the present work, the oxidation states were determined using the results of the detailed study of titanomaghemites by Nishitani (1979). Titanomagnetite samples with

* Present address: Institute of Mining Geology, Mining College, Akita University, Akita 010, Japan

** Present address: Department of Applied Physics, Tokyo Institute of Technology, Meguro-ku, Tokyo 152, Japan

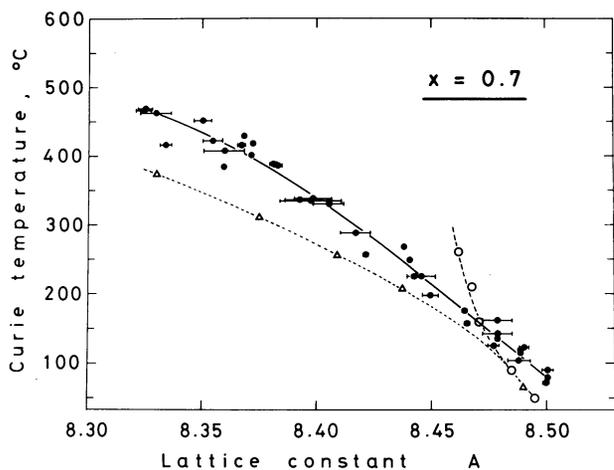


Fig. 1. Relation between lattice constant and Curie temperature for $x=0.7$ titanomagnetite. *Solid circles* represent experimental points for $x=0.7$ (ball-milled) titanomagnetite by Nishitani (1979). Error bars indicate standard deviations of lattice constants. Results of Ozima and Sakamoto (1971) are shown by *open circles* and those of Readman and O'Reilly (1972) by *open triangles*

particle sizes ranging from $0.57 \pm 0.33 \mu\text{m}$ – $300 \mu\text{m}$ were prepared and were oxidized at low temperatures.

Sample Preparation

To produce titanomagnetite of the desired composition, pure TiO_2 and Fe_2O_3 were mixed in the stoichiometric proportion and well ground in an agate mortar for several hours. The mixtures were then sintered in air at $1,200^\circ\text{C}$ for an hour. After that they were well ground and sintered again. Sintered materials were then crushed to about 2 mm and kept at $1,200^\circ\text{C}$ in a furnace with controlled oxygen fugacity (P_{O_2}) for several hours. They were quenched to about 0°C and the lattice constant was determined using an X-ray diffractometer to ascertain that a single phase titanomagnetite was formed. These heatings were repeated until the specimen homogeneity was sufficient that the two peaks of $\text{FeK}\alpha_1$ and $\text{FeK}\alpha_2$ in (511) and (400) reflections could be seen clearly. Details of the method of synthesis are described in Nishitani (1979).

Titanomagnetite samples ($x=0.7$) which cover the particle size from $0.57 \pm 0.33 \mu\text{m}$ – $300 \mu\text{m}$ were prepared. The following size fraction of titanomagnetite were obtained using sieves; 200–249 μm (sample No. 1), 105–149 (No. 2), 44–62 (No. 3). Samples with particle sizes less than 44 μm were obtained using an agate mortar or a ball-mill. Ball-milling was done in an ethyl alcohol. Titanomagnetites were ground in an agate mortar for about two hours, mixed with acetone and stirred well. After about 40 seconds the top layer of acetone was skimmed and this procedure was repeated several times. Acetone was added to the mixture and again stirred well. After about 15 seconds

Table 1. Grain size distribution of titanomagnetite ($x=0.7$)

Sample	Number of grains	Grain size
0.7 TM a precipitated part (2)	114	$1.14 \pm 0.71 \mu\text{m}$
0.7 TM ground in an agate mortar	459	0.84 ± 0.92
0.7 TM a precipitated part (1)	283	0.72 ± 0.57
0.7 TM ball-milled for 36.17 h	432	0.57 ± 0.33

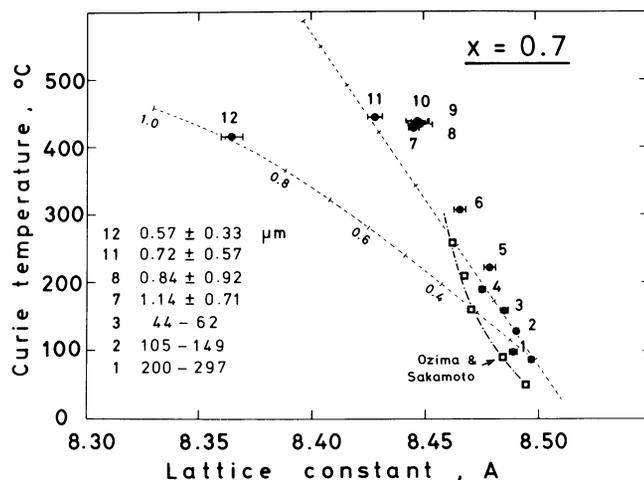


Fig. 2. Relationship between lattice constant and Curie temperature when different sizes of titanomagnetites ($x=0.7$) are heated for 300 minutes at 300°C . The results of Ozima and Sakamoto (1971) for $x=0.7$ are represented by *open squares*. Grain sizes are listed on the Figure

the top of the acetone was again skimmed. In this way, different time intervals classified titanomagnetite samples into different size ranges.

Precipitated parts were also available in this experiment. Samples classified into the size ranges were numbered No. 4 to No. 10. The same procedure was used on a ball-milled titanomagnetite and samples No. 11 and 12 were obtained. As the size of the titanomagnetite becomes small the sample number increases (No. 1–No. 12). Particle sizes less than 44 μm were determined from photographs taken by an electron microscope at magnifications of 5000–20000. The particle sizes of such samples are summarized in Table 1.

Other titanomagnetite samples ($x=0.5$ and 0.3) were synthesized and divided into two fractions: one was ball-milled and the other was ground in an agate mortar. Though their grain sizes were not measured directly, they are probably similar in size to $x=0.7$ titanomagnetites.

Grain Size Effect on Oxidation Behaviour

Samples of $x=0.7$ titanomagnetite were heated in air and kept at a temperature of 300°C for 300 minutes. The relation between the lattice constants and the Curie temperatures of the heated samples are shown in Figure 2. The broken lines show the trends for composition changes in stoichiometric titanomagnetite (right) and for oxidation of $x=0.7$ titanomagnetite obtained by Nishitani (1979). In this figure the sample number increases as the grain size of titanomagnetite becomes smaller. The observed particle sizes are shown in this figure. The results of Ozima and Sakamoto (1971) for $x=0.7$ are also shown in the same figure. Their trend, with a sharp rise in Curie temperature, coincides with the trend of the samples No. 1 to No. 10. They are almost on the stoichiometric titanomagnetite solid-solution line, so that they can be interpreted as the phase of separated Fe-rich titanomagnetite in high-temperature oxidation: $\text{Ti-rich titanomagnetite} + \text{O}_2 \rightarrow \text{Magnetite}_{\text{ss}} + \text{Ilmenite}_{\text{ss}}$ where $\text{Magnetite}_{\text{ss}}$ and $\text{Ilmenite}_{\text{ss}}$ are Fe-rich spinel and Ti-rich rhombohedral phases, respectively (Buddington and Lindsley 1964). For large grains the oxidation stage, that is the diffusion time of cations on oxidation may differ. However, in the present study Figure 2

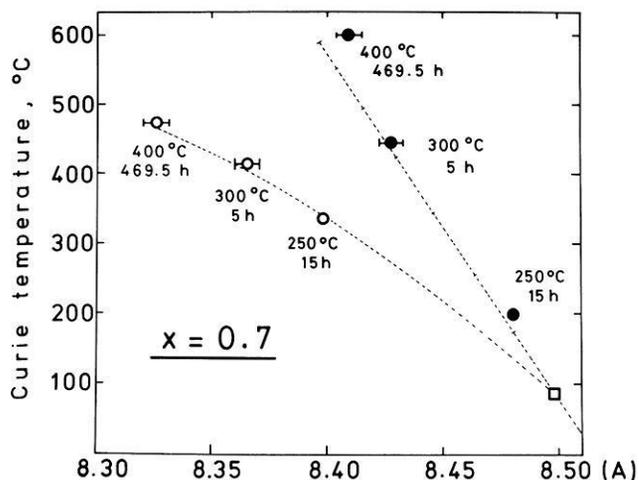


Fig. 3. Results of heat treatment in air for $x=0.7$ titanomagnetite. Temperatures and length of heating are medicated on the figure. Two samples, one is ball-milled (*open circles*) and the other is ground in an agate mortar (*solid circles*), were used

is interpreted supposing that the diffusion time of cations on oxidation is the same.

Samples with grain sizes between $0.72 \pm 0.57 \mu\text{m}$ (No. 11) and $1.14 \pm 0.71 \mu\text{m}$ (No. 7) reach almost the same position in Figure 2 when they are heated under the same conditions. However, the sample with grain size of $0.57 \pm 0.33 \mu\text{m}$ (No. 12) was oxidized and lay on the low-temperature oxidation line. Therefore, it can be concluded that a titanomagnetite sample ($x=0.7$) decomposes as a result of heating if its grain size is greater than 0.7 to $0.8 \mu\text{m}$. This indicates that there is a critical grain size in low-temperature oxidation. The critical size is about 0.7 to $0.8 \mu\text{m}$ in the case of $x=0.7$ titanomagnetite in a laboratory condition. The ball-milled titanomagnetite, and that ground in an agate mortar, behave in quite a different way when they are heated in air.

To show this distinction more clearly, six samples were prepared: they had three titanomagnetite compositions ($x=0.3, 0.5$ and 0.7) and for each composition one was ball-milled and the other was ground in an agate mortar. They were heated in air successively at 250°C for 15 h, at 300°C for 5 h (only for $x=0.7$) and at 400°C for 469.5 h. The results for $x=0.7$ titanomagnetite and $x=0.3, 0.5$ titanomagnetites are shown in Figures 3 and 4, respectively.

The ball-milled samples were oxidized along the oxidation line of Nishitani (1979) (Fig. 1). On the other hand, samples which were ground in an agate mortar were decomposed or separated to an iron-rich stoichiometric titanomagnetite (spinel phase) and a hemo-ilmenite (rhombohedral phase) by heating. Rhombohedral phases are seen clearly in a chart from an X-ray diffractometer. These experiments indicate that, by observing lattice constants and Curie temperatures before and after heating, it can be determined whether either the low-temperature oxidation process or the decomposition process dominates in oxidation of titanomagnetites.

Ozima and Sakamoto (1971) prepared titanomagnetite samples using an agate mortar and therefore their grain size was probably greater than the critical size of 0.7 to $0.8 \mu\text{m}$. As a result, decomposition processes occurred and a sharp rise in Curie temperature was observed as shown in Figure 2. On the other hand, the grain size of samples used by Readman and O'Reilly (1972) was estimated to be about $0.1 \mu\text{m}$ by the broaden-

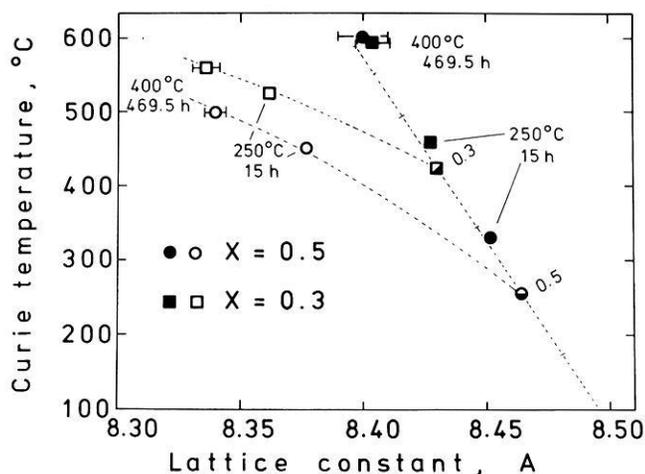


Fig. 4. Results of heat treatment in air. Temperatures and length of heating are marked. Four samples were prepared, two titanomagnetite compositions ($x=0.3$ and 0.5) and for each composition one is ball-milled (*open circles and squares*) and the other is ground in an agate mortar (*solid circles and squares*)

ing of X-ray peaks. Therefore, generally the same trend of oxidation process was obtained as in the present experiment. The slight but significant difference between the present and their results probably indicates the uncertainty in the determination of x and/or z values (Nishitani 1979).

Low-Temperature Oxidation in Submarine and Subaerial Basalts

Grains of titanomagnetites in oceanic basalts are usually smaller than $10 \mu\text{m}$, and a large number of grains finer than $1 \mu\text{m}$ exist (Lowrie 1974; Irving 1970). Continental basalts contain much larger magnetic minerals and the particle size commonly ranges up to $100 \mu\text{m}$ or more (Storetvedt et al. 1978).

In the previous section it was shown that the grain size controlled the state of low-temperature oxidation. This effect may appear in the difference between the oxidation states of oceanic basalts and those of continental basalts. Low-temperature oxida-

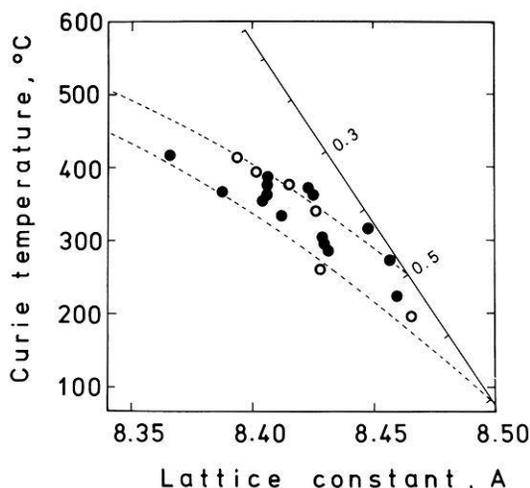


Fig. 5. Relationship between lattice constants and Curie temperatures of titanomagnetites extracted from submarine basalts (*solid circles*) and subaerially erupted basalts (*open circles*). Subaerial basalts were confined to only low-temperature oxidation products

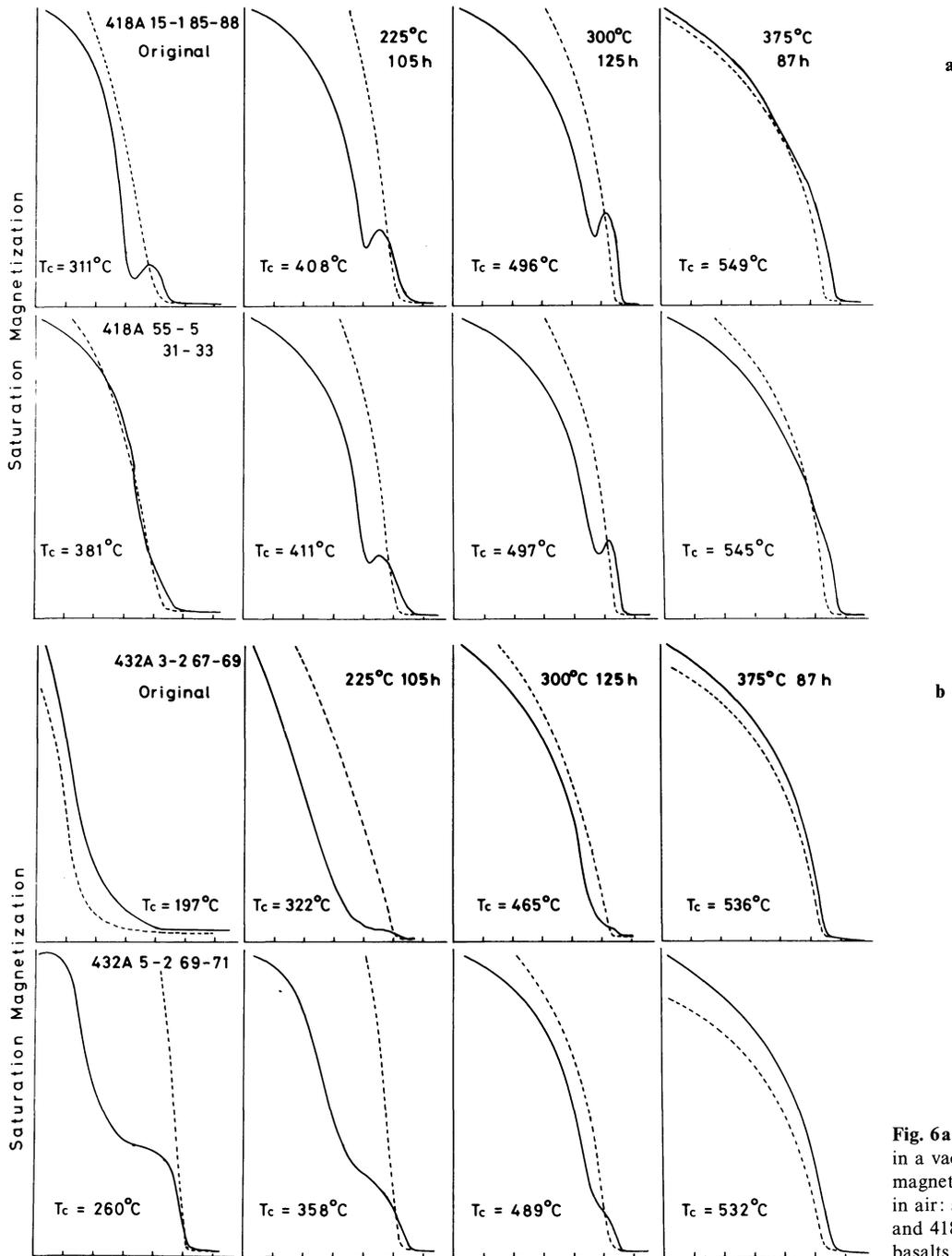


Fig. 6a and b. Thermomagnetic curves in a vacuum (below 1×10^{-5} Torr) for magnetic minerals which were heated in air: **a** submarine basalts (418A 15-1 and 418A 55-5), **b** subaerially erupted basalts (432A 3-2 and 432A 5-2)

tion is affected by other factors such as temperature and f_{O_2} . In this study we pay attention only to the effect of grain size.

Oxidation experiments were carried out on magnetic minerals, obtained by the Deep Sea Drilling Project (DSDP) legs 51, 52, 53 and 55. The samples of Legs 51, 52 and 53 are submarine basalts which originated at a Mid-Atlantic ridge-type spreading center about 100 m.y. ago (Donnelly et al. 1980), while those of Leg 55 are, although now completely submerged, subaerially erupted basalts from the Hawaiian hotspot with ages between 55 and 65 m.y. (Jackson et al. 1980).

Sixteen samples from Holes 417D and 418A (Legs 51, 52 and 53), and six samples from Holes 430A and 433C (Leg 55) were selected and magnetic minerals were separated using a hand magnet. The samples of Legs 51, 52 and 53 basalts were ran-

domly selected, while those of Leg 55 were preferentially selected from those that have undergone only slight high-temperature oxidation and later subjected to varying degrees of low-temperature oxidation. Most of the Leg 55 samples show only the effects of high-temperature oxidation, and their Curie temperatures and lattice parameters were close to those of magnetite (Kono 1980). The relationships between Curie temperatures (T_c) and lattice constants (A) are shown in Figure 5.

It is clear from Figure 5 that most of the data of submarine basalts (solid circles) in a T_c vs. A diagram are confined between two oxidation lines of $x=0.5$ and 0.7 and various oxidation states exist. This indicates that the magnetic minerals are titanomaghemites with a narrow range of x (0.5–0.7) provided that all magnetic minerals contain no impurities. Analysis with the

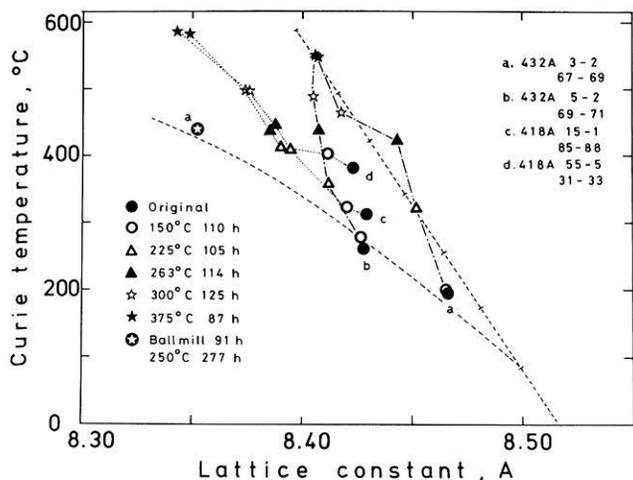


Fig. 7. The effect of heat treatment of magnetic minerals of submarine basalts (418A 15-1 and 418A 55-5) and subaerially erupted basalts (432A 3-2 and 432A 5-2) on the Curie temperature – lattice constant curves. Sample codes, and preparation: conditions are keyed on the figure

microprobe is necessary for exact investigation. The effect of Mg or Al on the magnetic properties of titanomagnetites has been investigated (Nishitani 1981; O'Donovan and O'Reilly 1977a, b). However, lattice constants and Curie temperatures of titanomaghemites containing Mg or Al over the whole range of x have not been reported at the present stage. Therefore the effect of impurities is not considered in this study.

Four magnetic minerals were selected and heated in laboratory conditions. As submarine basalts, 418A 15-1 85-88 and 418A 55-5 31-33 were chosen, which have relatively low z values (Hamano et al. 1980). 432A 3-2 67-69 and 432A 5-2 69-71 were chosen as subaerially erupted basalts because they have low oxidation state and high ulvospinel ratio (Kono 1980). The separated magnetic minerals were heated in air for about 100 h at various temperatures. Thermomagnetic curves are shown in Figure 6. These curves were obtained using a magnetic balance under a magnetic field of 5.5 kOe. All the samples were heated in a vacuum (below 1×10^{-5} Torr) for thermomagnetic analysis up to about 650°C. The change of lattice constants and Curie temperatures of the extracted titanomagnetites as a result of heat treatment are summarized in Figure 7.

The samples 418A 15-1 and 418A 55-5 show similar trends in Figure 7. However, experimental points for 418A and 432A show a different trend as the treatment temperature increases. 432A samples show a sharp rise in Curie temperature, indicating that they are decomposed into Fe-rich titanomagnetite and hemilmenite (high-temperature oxidation). 418A samples show a trend similar to the ball-milled samples, that is, low-temperature oxidation processes seems dominant. However, 418A samples do not follow the low-temperature oxidation line completely, because they contain a slight high-temperature oxidation product. It appears that low-temperature oxidation occurs at the grain size of the magnetic minerals of 418A, but that it does not occur at the grain size of the magnetic minerals of 432A under laboratory conditions.

If small grain size is essential for low-temperature oxidation to occur, 432A samples may also be low-temperature oxidized when their grain size becomes small enough. The magnetic minerals of 432A 3-2 were ball-milled for 91 h and heated for 277 h at 250°C. The result of this heating is also shown in Figure 7. It is evident from this figure that low-temperature oxidation

occurred under the same conditions because the particle size is now smaller than before. Though the sizes of these magnetic minerals were not measured, they are probably less than 0.5 μm , by analogy to other ball-milled samples. It can be concluded from above experiments that the size of the magnetic minerals is one of the controlling factors in low-temperature oxidation.

The sizes of the magnetic minerals were measured only in the samples 418A 55-5 and 432A 5-2 using an electron microscope. The grain sizes which were determined from the photograph of the electron microscope are $1.47 \pm 0.83 \mu\text{m}$ ($N=31$) for 418A 55-5 and $2.11 \pm 1.25 \mu\text{m}$ ($N=30$) for 432A 5-2. As the numbers of grains counted are extremely small, the mean size is not very reliable and likely to contain some error. The magnetic minerals were separated using a hand magnet after they were ground in an agate mortar in a slurry of ethyl alcohol, so that magnetic minerals were probably smaller than the real size in a rock. Magnetic minerals in subaerial basalts in particular, are much larger than the observed value. Nevertheless, it may be concluded that the grain size effect is one of the reasons why oxidation states are relatively low in subaerially erupted basalts and why very severe oxidation states ($z \sim 1.0$) are observed in submarine basalts.

The difference in grain sizes of samples 418A and 432A is not so pronounced. However, 418A samples showed a tendency to low-temperature oxidation, while 432A samples were separated by heating. Particle sizes of 432A samples are greater than the critical size of low-temperature oxidation, therefore only high-temperature oxidation occurs. The effect of ball-milling reduces their grain sizes to less than the critical size. On the other hand, in 418A samples, high-temperature oxidation as well as low-temperature oxidation occurs, because their particle sizes are widely spread.

Ozima and Sakamoto (1971) showed that self-reversal of J_s (saturation magnetization) occurred during progressive low-temperature oxidation if the samples did not suffer initial high-temperature oxidation. They showed that initial conditions such as high-temperature oxidation might influence low-temperature oxidation. Samples which suffered high-temperature oxidation might resist the process of low-temperature oxidation. Therefore, in a re-heating process, subaerially erupted basalts may reveal this difficulty of low-temperature oxidation as a result of high-temperature oxidation.

Conclusions

As a result of experiments on samples of controlled grain size, it has been shown that there is a critical size for low-temperature oxidation in laboratory conditions. The critical size may vary under other conditions, such as those realized in nature. If titanomagnetites smaller than the critical size are heated at moderate temperatures (100–400°C), they are low-temperature oxidized. If the size of a titanomagnetite is greater than the critical size, it decomposes to Fe-rich titanomagnetite and hemo-ilmenite as a result of heating. This critical size is $0.7 \sim 0.8 \mu\text{m}$ for $x=0.7$ titanomagnetite. Through this experiment, it has been revealed that the main difference between the results of Ozima and Sakamoto (1971) and Readman and O'Reilly (1972) can be explained in terms of grain size. However, the results of Readman and O'Reilly (1972) do not completely coincide with those of a detailed study (Nishitani 1979).

Magnetic minerals were separated from submarine basalts and subaerially erupted basalts, and lattice constants and Curie temperatures have been measured. In Curie temperature vs. lat-

tice constant diagrams, magnetic minerals of submarine basalts are in the range $0.5 < x < 0.7$ and $0 < z < 1.0$. The oxidation state (z) of the magnetic minerals of subaerial basalts, restricted to low-temperature oxidation products, is less than 0.6. The low-temperature oxidation proceeds in submarine basalts, but unmixing occurs in subaerial basalts as a result of heating. The magnetic minerals of subaerial basalts are also low-temperature oxidized when their grain size is made small enough. It can be concluded from above experiment that the grain size of the magnetic minerals is one of the controlling factors of low-temperature oxidation.

Acknowledgement. I wish to thank Drs. M. Ozima, and Y. Hamano for their frequent discussions and helpful suggestions.

References

- Akimoto, S., Katsura, T., Yoshida, M.: Magnetic properties of TiFe_2O_4 - Fe_3O_4 system and their change with oxidation. *J. Geomagn. Geoelectr.* **9**, 165–178, 1957
- Akimoto, S., Katsura, T.: Magneto-chemical study of the generalized titanomagnetite in volcanic rocks. *J. Geomagn. Geoelectr.* **10**, 69–90, 1959
- Buddington, A.F., Lindsley, D.H.: Iron-titanium oxide minerals and synthetic equivalents. *J. Petrology* **5**, 310–357, 1964
- Day, R., Fuller, M.D., Schmidt, V.A.: Magnetic hysteresis properties of synthetic titanomagnetite. *J. Geophys. Res.* **81**, 873–880, 1976
- Day, R., Fuller, M., Schmidt, V.A.: Hysteresis properties of titanomagnetites: Grain-size and compositional dependence. *Phys. Earth Planet. Inter.* **13**, 260–267, 1977
- Donnelly, T.W. et al.: Introduction and explanatory notes. Initial Reports of the Deep Sea Drilling Project, Vol. 51, 52 and 53, pp. 5–22: Washington, D.C.: U.S. Government Printing Office, 1980
- Hamano, Y., Nishitani, T., Kono, M.: Magnetic properties of basalt samples from Deep Sea Drilling Project holes 417D and 418A. Initial Reports of the Deep Sea Drilling Project, Vol. 51, 52, and 53, pp. 1391–1405: Washington, D.C.: U.S. Government Printing Office, 1980
- Irving, E.: The Mid-Atlantic ridge at 45° N, XIV. Oxidation and magnetic properties of basalt; review and discussion. *Can. J. Earth Sci.* **7**, 1528–1538, 1970
- Jackson, E.D., Koizumi, I., Dalrymple, G.B., Clague, D.A., Kirkpatrick, R.J., Greene, H.G.: Introduction and summary of results from DSDP Leg 55, the Hawaiian-Emperor Hot-spot experiment. Initial Reports of the Deep Sea Drilling Project Vol. 55 pp. 5–31. Washington, D.C.: U.S. Government Printing Office 1980
- Keefer, C.M., Shive, P.N.: Curie temperature and lattice constant reference contours for synthetic titanomagnhemites. *J. Geophys. Res.* **86**, 987–998, 1981
- Kono, M.: Magnetic properties of DSDP Leg 55 basalts. Initial Reports of the Deep Sea Drilling Project Vol. 55 pp. 723–736: Washington, D.C.: U.S. Government Printing Office, 1980
- Levi, S., Merrill, R.T.: A comparison of ARM and TRM in magnetite. *Earth Planet. Sci. Lett.* **32**, 171–184, 1976
- Lowrie, W.: Oceanic basalt magnetic properties and the Vine and Matthews Hypothesis. *J. Geophys.* **40**, 513–536, 1974
- Nagata, T.: *Rock Magnetism*. Tokyo: Maruzen 1961
- Nishitani, T.: Low-temperature oxidation of titanomagnetites and its effect on the remanent magnetization in basaltic rocks. PhD Thesis, University of Tokyo, 1979
- Nishitani, T.: Magnetic properties of titanomagnetites containing spinel (MgAl_2O_4). *J. Geomagn. Geoelectr.* **33**, 171–179, 1981
- O'Donovan, J.B., O'Reilly, W.: Range of non-stoichiometry and characteristic properties of the products of laboratory magnetization. *Earth Planet. Sci. Lett.* **34**, 291–299, 1977a
- O'Donovan, J.B., O'Reilly, W.: The preparation, characterization and magnetic properties of synthetic analogues of some carriers of the palaeomagnetic record. *J. Geomagn. Geoelectr.* **29**, 331–344, 1977b
- Ozima, M., Larson, E.E.: Low- and high-temperature oxidation of titanomagnetite in relation to irreversible changes in the magnetic properties of submarine basalts. *J. Geophys. Res.* **75**, 1003–1017, 1970
- Ozima, M., Sakamoto, N.: Magnetic properties of synthesized titanomagnhemite. *J. Geophys. Res.* **76**, 7035–7046, 1971
- Readman, P.W., O'Reilly, W.: The synthesis and inversion of non-stoichiometric titanomagnetites. *Phys. Earth Planet. Inter.* **4**, 121–128, 1970
- Readman, P.W., O'Reilly, W.: Magnetic properties of oxidized (cation deficient) titanomagnetite ($\text{Fe, Ti, } \square)_3\text{O}_4$. *J. Geomagn. Geoelectr.* **24**, 69–90, 1972
- Sakamoto, N., Ince, P.I., O'Reilly, W.: The effect of wet-grinding on the oxidation of titanomagnetites. *Geophys. J.* **15**, 509–515, 1968
- Storetvedt, K.M., Carmichael, C.M., Hayatsu, A., Palmer, H.C.: Palaeomagnetism and K/Ar results from the Duncansby volcanic neck, NE Scotland: superimposed magnetizations, age of igneous activity, and tectonic implications. *Phys. Earth Planet. Inter.* **16**, 379–392, 1978

Received February 2, 1981; Revised version December 8, 1981

Accepted January 21, 1982

The Romanian Earthquake of March 4, 1977

III. Improved Focal Model and Moment Determination *

E. Räckers^{1**} and G. Müller²

¹ Geophysical Institute, University of Karlsruhe, Hertzstr. 16, 7500 Karlsruhe, Federal Republic of Germany

² Institute of Meteorology and Geophysics, University of Frankfurt, Feldbergstr. 47, 6000 Frankfurt, Federal Republic of Germany

Abstract. The focal process of the Romanian earthquake of March 4, 1977, which had already been studied in an earlier paper (Müller et al. 1978), is reinvestigated on the basis of additional data from the Russian seismic network. The general features of this multiple event, as given in the first investigation, were confirmed, but several details changed significantly. It is shown that the weak foreshock of this earthquake (depth according to CSEM and ISC 93 and 86 km, respectively) took place at about the same location as the beginning of the main phase of rupture, shock 1. Additional accelerations of rupture which are summarized as shock 2 are difficult to localize with respect to shock 1. Rupture propagated essentially horizontally and towards the SW and terminated abruptly 13 s later at the place of shock 3, about 50 km away from shock 1; this produced exceptionally strong stopping phases in far-field seismograms. The average rupture velocity was 3.7 km/s. A new fault-plane solution is given for shock 1 and an approximate one for shock 3 which indicates significant deviations of the rupture surface from a plane. *P*-wave seismograms of several long-period WWNSS stations are modelled with the aid of theoretical seismograms, using the concept of a station-dependent equivalent moment function of a single double-couple source. The moments determined in this way fall into the range $2\text{--}3 \times 10^{27}$ dyne-cm.

Key words: Earthquakes – Rupture process – Fault-plane solution – Seismic moment

Introduction

In a first study of the Romanian earthquake of March 4, 1977 (Müller et al. 1978, hereafter called paper I) evidence was presented for the multiple nature of this event. The earthquake consisted of a weak foreshock, followed by at least three main shocks (called shocks 1, 2, 3 in the following), whose relative locations were determined by a master-event analysis. A fault-plane solution was derived for shock 1 from long-period seismograms from the Worldwide Network of Standard Seismographs (WWNSS). According to the proposed rupture model, rupture propagated uphill over a distance of about 20 km from the location of the foreshock at a depth of 93 km (source depth according

to Centre Sismologique Européo-Méditerranéen (CSEM), to the place of shock 1, where the main phase of the earthquake was initiated about 5 s later. Rupture propagation after this moment was in general directed towards the SW and downhill. A location with acceleration of rupture, shock 2, was located somewhat off the steep, NW-dipping nodal plane of the fault-plane solution. Rupture stopped abruptly and close to this nodal plane at a slant distance of 72 km from, and 14.5 s after, shock 1 at a depth of 109 km (shock 3), producing strong stopping phases in the seismograms.

The purpose of the present paper is to summarize some new aspects of the focal process of the Romanian earthquake which emerged from a detailed analysis of the data of paper I and of additional data from the Russian network of seismic stations (Räckers 1981). We present an improved fault-plane solution for shock 1, put forth evidence for a closer collocation of the foreshock and shock 1 than derived in paper I, redetermine the location of shock 3 relative to shock 1, and discuss briefly the (uncertain) nature of shock 2. Our results on these points represent significant, although not basic, changes of the rupture model suggested in paper I. Moreover, we infer from an approximate fault-plane solution of shock 3 that a rotation of the rupture surface took place during the earthquake, and determine the seismic moment by time-domain modelling of the wave-group consisting of the phases *P*, *pP* and *sP* at several stations. We

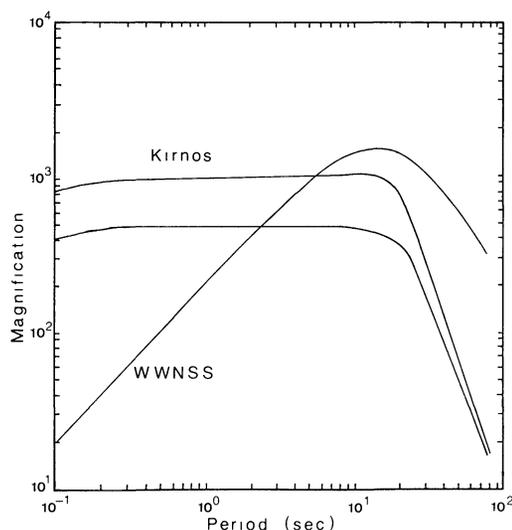


Fig. 1. Typical magnification curves of long-period electro-dynamical seismographs in the Russian network (Kirnos instruments) and the Worldwide Network of Standard Seismographs (WWNSS)

* Contribution No. 242, Geophysical Institute, University of Karlsruhe

** Present address: Institut für Geophysik, Westfälische Berggewerkschaftskasse, Herner Str. 45, D-4630 Bochum

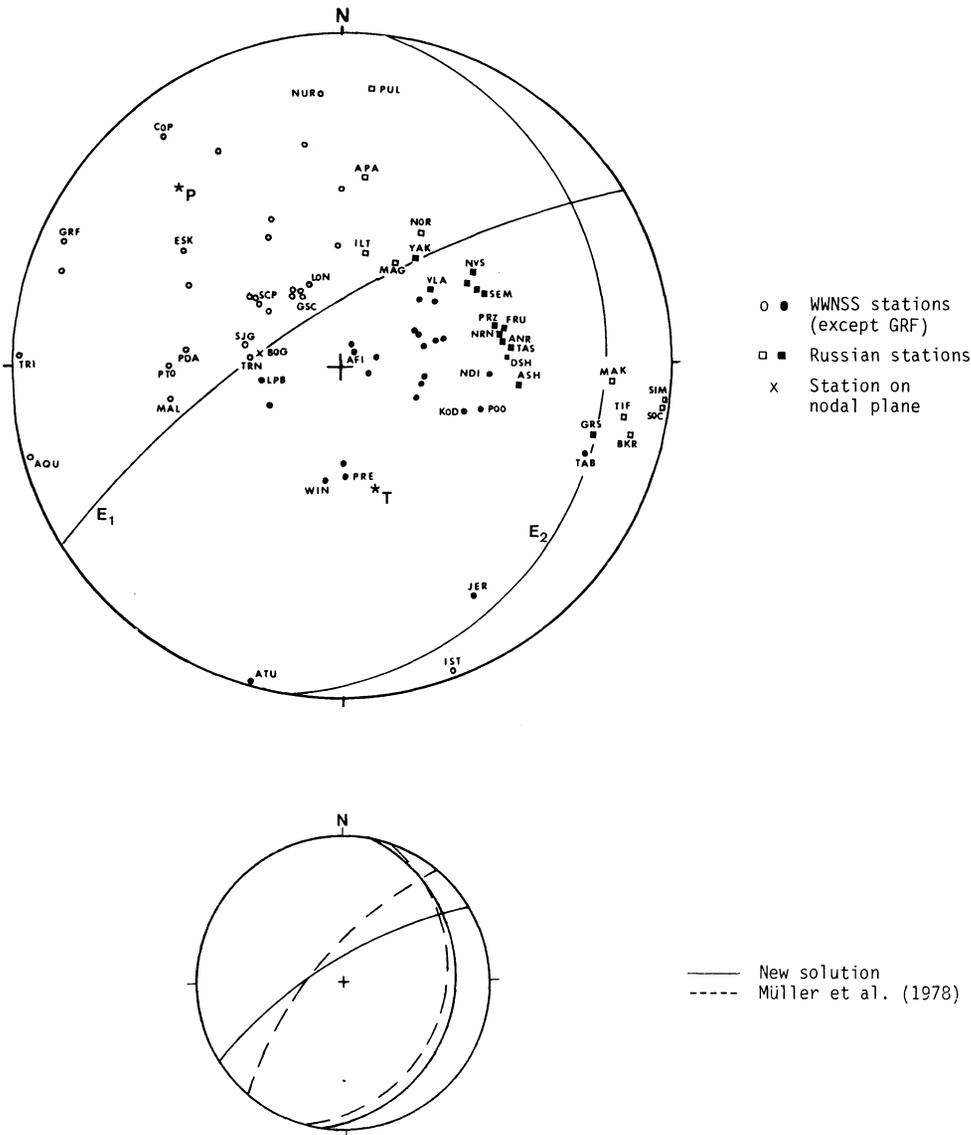


Fig. 2. New fault-plane solution of shock 1 (*top*) and comparison with the old solution (*bottom*). Equal-area projections of the lower focal hemisphere, *open symbols* are dilatations, *closed symbols* compressions

also discuss the focal model which Hartzell (1979) investigated in order to explain the accelerograms of the earthquake recorded at Bucharest.

The following records of the Romanian earthquake were at our disposal:

a) about 60 long-period and a few short-period records from WWNSS stations,

b) long- and short-period records from 28 Russian stations (see Fig. 1 for a comparison of the magnification curves of the most commonly used long-period Kirnos electrodynamic seismographs and a long-period WWNSS instrument),

c) broadband seismograms of the station GRF (Gräfenberg, Federal Republic of Germany),

d) copies of Mainka pendulum records of the Romanian stations BAC, BUC, CMP and IAS.

New Fault-Plane Solution of Shock 1

Figure 2 shows the new fault-plane solution of shock 1, based on long-period WWNSS and Russian data. The essential changes with respect to the fault-plane solution of paper I are a rotation

of the strike direction of nodal plane E1 by 18° and an increase of its dip angle by 6°. These changes are required by the downward motions at the Russian stations ILT, MAG and NOR. A reinspection of the WWNSS stations close to nodal plane E1 showed, (1) that the polarities of stations TRN and GSC had to be changed from upward to downward, (2) that BKS (located close to GSC and LON) had to be omitted because of strong background noise at the expected arrival time of energy from shock 1, and (3) that shock 1 is invisible at BOG, i.e., that BOG is very close to nodal plane E1. (Expected arrival times for shock 1 at BKS and BOG followed from safe readings of this shock at nearby stations, corrected for the travel-time difference due to the difference in epicentral distance.)

The parameters of the new fault-plane solution are:

	Azimuth (deg)	Dip (deg)
Pole of nodal plane E1	148	14
Pole of nodal plane E2	278	69
<i>P</i> -axis	317	24
<i>T</i> -axis	165	57

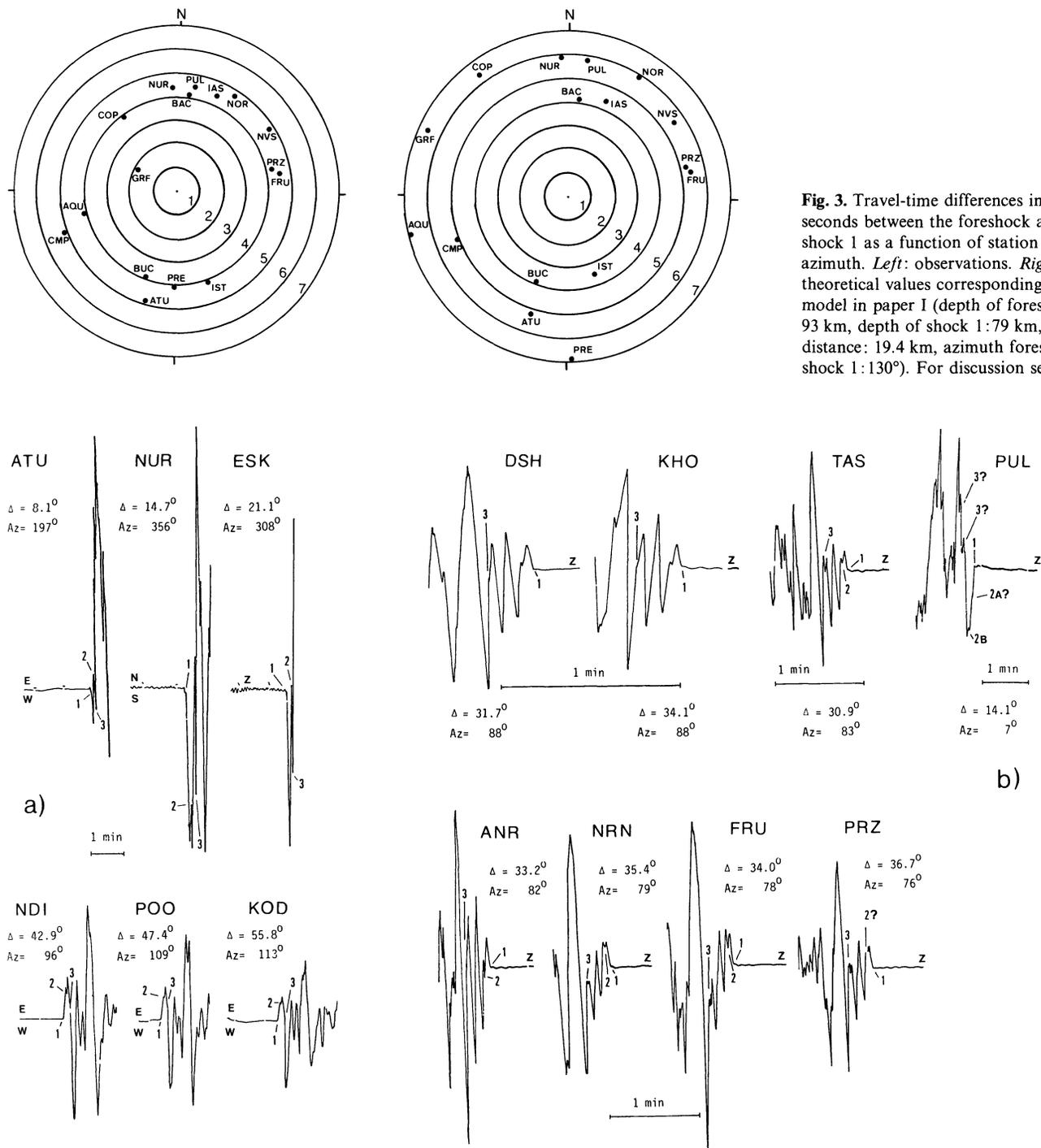


Fig. 4a and b. Long-period seismograms of **a** the WWNSS and **b** the Russian stations that were used in the master-event analysis of shock 3. Note that in the Russian seismograms time increases from right to left and time scales are different. Numbers indicate *P* arrivals due to the different shocks

Location of the Foreshock

Figure 3 shows the azimuthal distribution of the observed travel-time differences between shock 1 and the foreshock, together with the distribution of travel-time differences which would be expected if the foreshock location was as suggested in paper I. This model obviously does not satisfy the observations. We do not want to discuss the reasons given in paper I for the choice of the foreshock location, but rather say that the azimuthal distribution of the travel-time differences with values

almost exclusively between 4 and 5 s points to the locations of the foreshock and shock 1 being quite close together. The exceptionally low value of the travel-time difference for GRF may be related to the weaker amplitudes of shock 1 at this station, compared with other stations, such that at GRF the noise level masked the foreshock for longer. We disregard the GRF data point because of the rather large number of other observations which are all compatible. Our model, then, is that the foreshock (CSEM coordinates: 26.78 E, 45.78 N, depth 93 km, origin time 19:21:56.2; International Seismological

Fig. 3. Travel-time differences in seconds between the foreshock and shock 1 as a function of station azimuth. *Left*: observations. *Right*: theoretical values corresponding to the model in paper I (depth of foreshock: 93 km, depth of shock 1: 79 km, distance: 19.4 km, azimuth foreshock-shock 1: 130°). For discussion see text

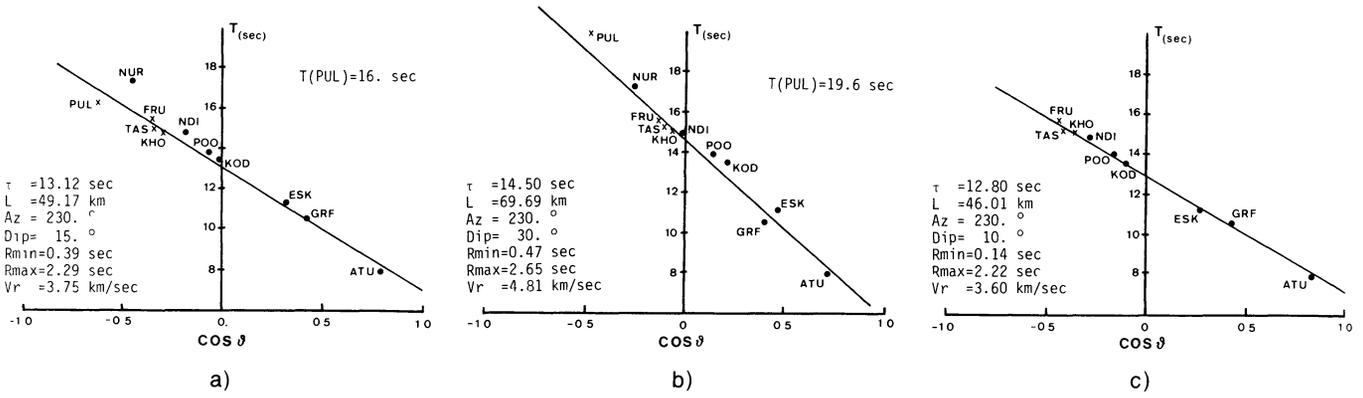


Fig. 5. Master-event analysis for shock 3 relative to shock 1: regression lines for T versus $\cos \vartheta$ in three different cases. ● = WWNSS stations, x = Russian stations. The Russian stations ANR, DSH, NRN and PRZ are not shown since their data are close to those of FRU, KHO and TAS. Az = azimuth of the direction from shock 1 to shock 3, Dip = dip angle of this direction, R_{min} and R_{max} = minimum and maximum residue, $V_r = L/\tau$ = apparent rupture velocity

Centre (ISC) coordinates: 26.72 E, 45.83 N, depth 86 km, origin time 19:21:54.1) and shock 1 occurred at about the same place with a time separation of about 4.5 s.

New Master-Event Analysis for Shock 3

The key role that the broadband record of station GRF played in the identification of shock 3 was described in paper I. In Figure 4 those long-period WWNSS and Russian seismograms are reproduced in which the P arrival due to shock 3 can also be seen. The signatures of the two types of instruments involved are different, but as in the WWNSS seismograms, shock 3 in the Russian seismograms is connected with a strong deflection *opposite* to that of shock 1. The travel-time differences T between shock 3 and shock 1, corrected for the screw-line effect of drum recording if necessary, were used in a master-event analysis, as in paper I. Three regression lines $T = \tau - L \cos \vartheta / \alpha$ are shown in Figure 5 (τ = origin-time difference, L = distance between the two shocks, ϑ = angle between the direction from shock 1 to

shock 3 and the direction of the P -wave ray to the station, $\alpha = P$ velocity in the focal volume, assumed here to be 8.3 km/s):
 – one regression line for all stations mentioned in Figure 4, plus GRF, assuming $T(PUL) = 16.0$ s (Fig. 5a),
 – one for all stations and $T(PUL) = 19.6$ s (Fig. 5b),
 – and one with PUL and NUR omitted (Fig. 5c).
 The result of Figure 5b comes closest to the results of paper I, but we consider the other two solutions to be closer to reality, since the larger value of $T(PUL)$ is less probable (see Fig. 4). In all three cases the location of shock 3 is very near the nodal plane E1 of the fault-plane solution of shock 1 (Fig. 2), such that we retain the identification of this plane as the rupture plane and the interpretation of shock 3 as an abrupt termination of rupture.

Expressing the new coordinates of shock 3 with respect to those of shock 1 by some numbers, we have the time difference 13 s, the slant distance 50 km, the depth difference 13 km (shock 3 is deeper), the epicentral distance 48 km, the azimuth 230° (from shock 1 to shock 3) and the dip angle 15° . The average

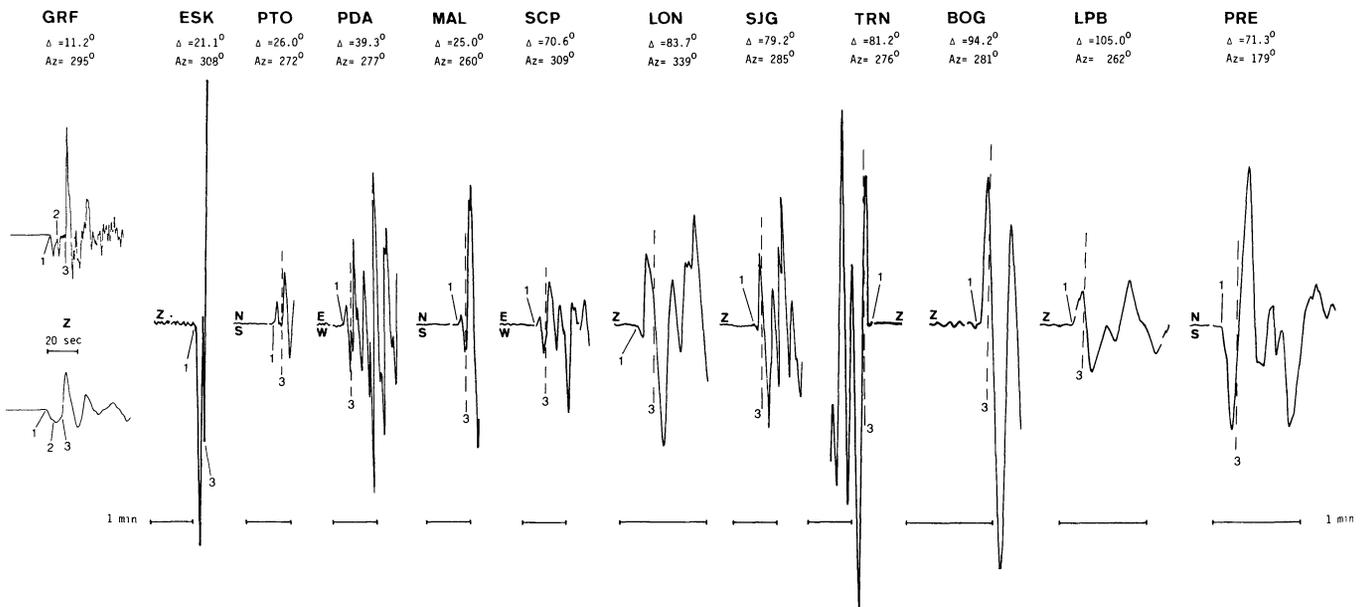


Fig. 6. Seismogram examples showing the polarities of shock 3. The records of GRF (broadband and WWNSS simulation) and ESK, in conjunction with other records (Fig. 4), have been used to predict the shock 3 onset times at stations PTO to PRE, indicated by the dashed lines. The shock 3 polarities at these stations correspond to downward motion

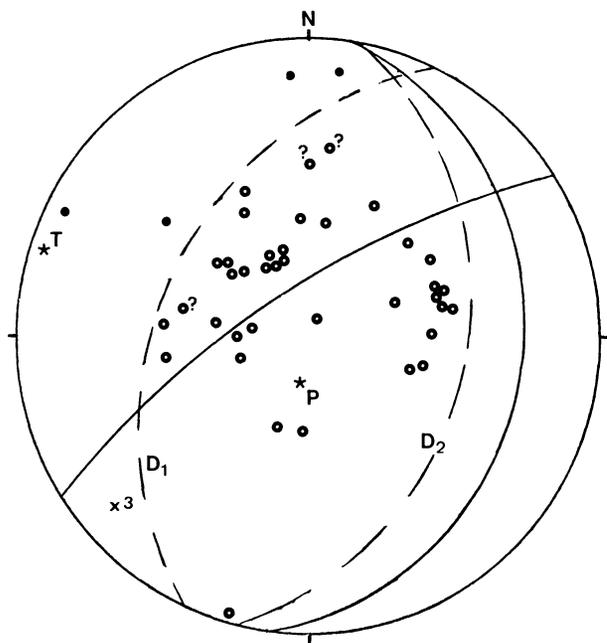


Fig. 7. Approximate fault-plane solution of shock 3 (dashed nodal lines), compared with the fault-plane solution of shock 1 (solid nodal lines). The cross labelled 3 is the location of shock 3 in the fault-plane solution of shock 1

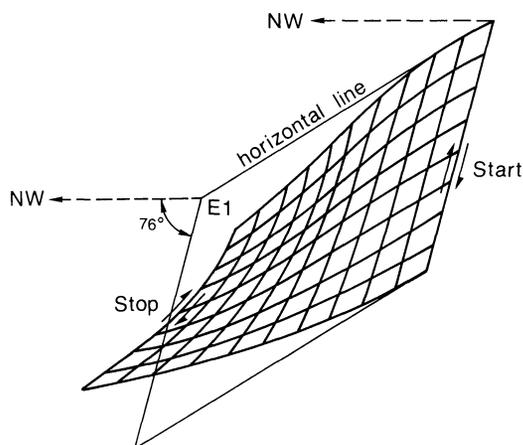


Fig. 8. Schematic diagram, not to scale, of the contorted rupture surface, based on the fault-plane solutions of shock 1 and shock 3. E1 is the initial rupture plane of shock 1

rupture velocity 3.7 km/s is considerably lower than the value 4.98 km/s given in paper I; this is mainly due to the decrease in slant distance from 72 to 50 km. The rupture velocity is about 0.8 times the S velocity in the focal region which is assumed to be about 4.6 km/s.

Polarity Distribution for Shock 3

For the identification and localization of shock 3 only those seismograms could be used in which the P -wave onset times of this shock could be determined. The polarities of these arrivals could be read in many more records, because shock 3 was so strong. Figure 6 shows examples of seismograms with the expected onset times for shock 3, according to the results of the foregoing section, indicated by dashed lines. The polarities of

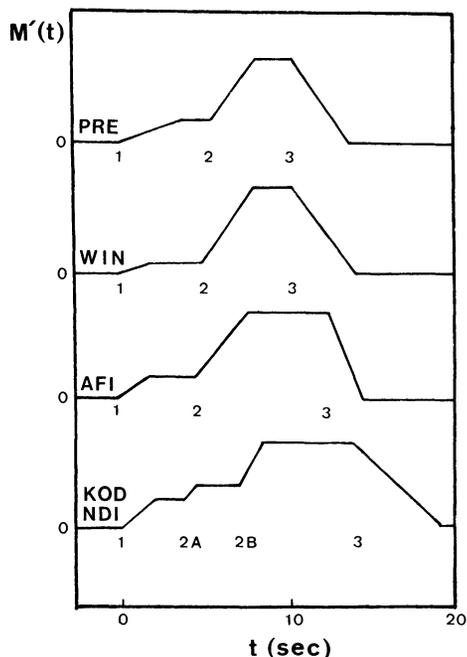


Fig. 9. Moment function time derivatives (normalized) which were used for modelling the long-period records of the WWNSS stations PRE, WIN, AFI, NDI and KOD. The numbers give the times of shocks 1, 2, and 3, as seen by these stations. The form of the curves represents true ground motion

the deflections following these lines all correspond to dilatational motion. Figure 7 shows a polarity plot in which all useful stations are included. The nodal plane D1 of this approximate fault-plane solution is relatively well determined by the many dilatations and the four safe compressions (at GRF, ESK, PUL and NUR); for nodal plane D2 there is more freedom. The azimuths of the pole directions of D1 and D2 are 115° and 281° , the dip angles 43° and 44° , respectively.

Our interpretation of these polarities is that D1 is the rupture plane of shock 3. Combination with the fault-plane solution of shock 1 indicates that the rupture surface of the whole earthquake deviated considerably from a plane: towards the SW end the rupture surface became less steep and changed its strike direction more towards the N (Fig. 8). These contortions of the rupture surface make it difficult to give a fault-plane solution which is representative for the whole earthquake; this is a certain drawback for the moment determination in a later section.

Re-Investigation of Shock 2

There are a few clear indications that between the shocks 1 and 3 there was an additional shock with rather strong P -wave radiation (see the GRF records in Figure 6 and the WWNSS records in Fig. 4a). The corresponding master-event analysis in paper I, however, did not give a very convincing result: the location of shock 2 was quite far off the rupture surface and had great uncertainty. The Russian data (see the examples in Fig. 4b) complicate the matter even more, since they indicate more than one shock between the shocks 1 and 3. We determined the time differences with respect to shock 1 of all onsets between shock 1 and shock 3 in all available seismograms and tried to group them, in order to find out whether shock 2 could possibly consist of two shocks, 2A and 2B (see the record for PUL in Fig. 4b). As a result, we actually consider this a possibility, but of questionable significance, since we probably would

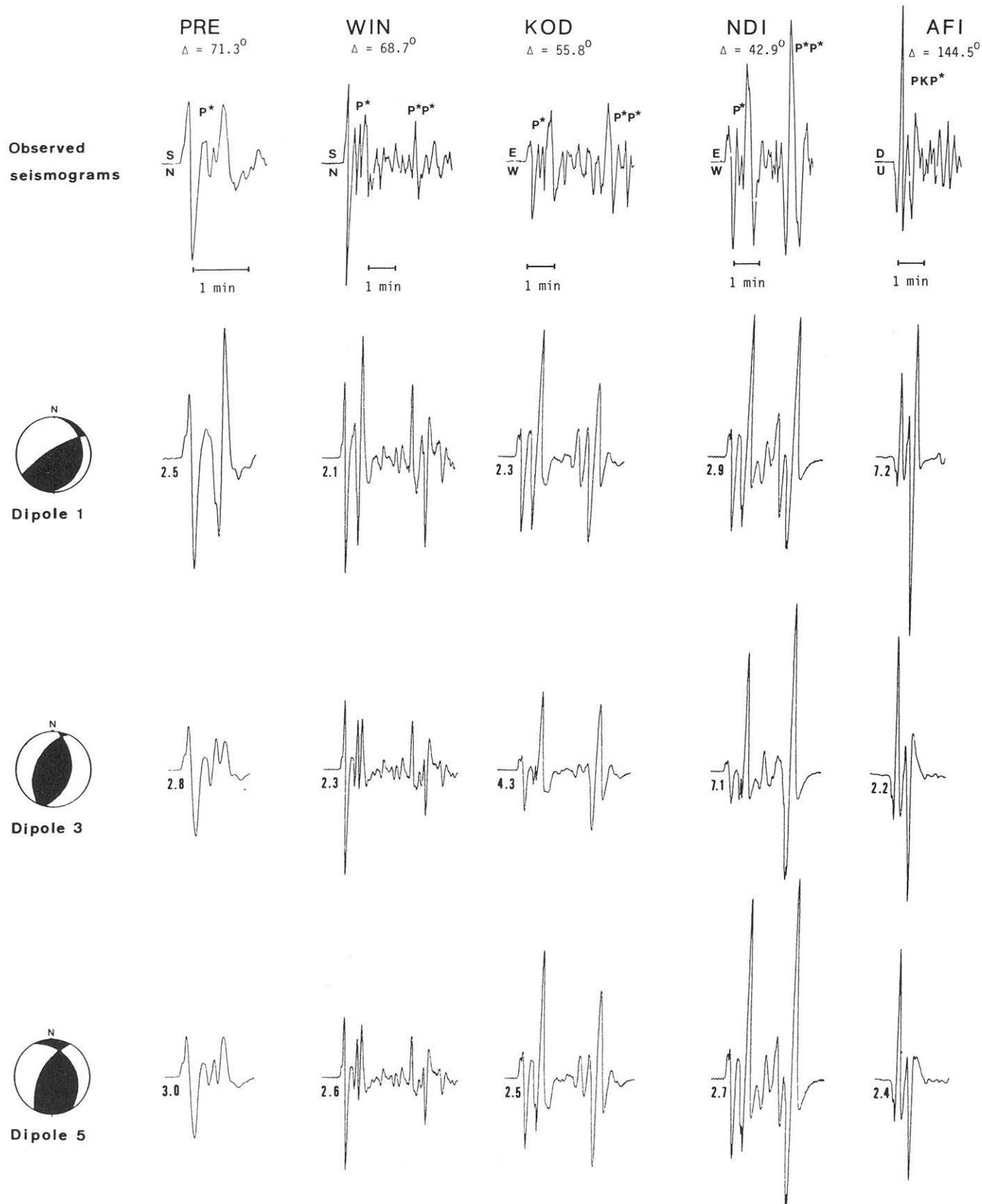


Fig. 10. Comparison of observed and theoretical seismograms for three different double couples. Dipole 1 corresponds to shock 1 (Fig. 2), dipole 3 to shock 3 (Fig. 7) after a change in sign, and dipole 5 is one of a few more orientations investigated. The moment derivatives assumed for each station are shown in Fig. 9. The numbers below the beginning of the theoretical seismograms are the moment values in units of 10^{27} dyne-cm

also have been able to infer more than two shocks. We therefore conclude only that apparently the rupture process between the shocks 1 and 3 was more complex than implied by the model of one additional shock, but that these complexities are not resolvable, at least not with far-field observations.

Moment Determination

Hartzell (1979) has estimated the moment of the Romanian earthquake from surface-wave amplitudes, as recorded by WWNSS, Seismic Research Observatory (SRO) and High Gain Long Period (HGLP) stations, and found a value in the range $1\text{--}2 \times 10^{27}$ dyne-cm. The rupture process was modelled by a SW-moving point shear dislocation which had a step moment function and the fault-plane solution of shock 1, as given in paper I. Here, we attempt to model with theoretical seismograms the compressional body-wave phases P , pP and sP which on long-period records form one complicated arrival, called P^* in the following; for one distant station PKP^* is used. Additional to the complexity due to surface reflections is the complexity due to the rupture process itself. We condense this complexity into the moment function of a stationary point-source at a depth of 95 km, i.e., each station which is used in the analysis is assigned an equivalent moment function whose qualitative features – mainly the times of slope discontinuities of the moment derivative – follow from the multiple-event analysis. Another simplifying assumption is that we work with a time-independent orientation of the double couple; it is varied in a few discrete steps between the fault-plane solution for shock 1 (Fig. 2) and the reversed fault-plane solution of shock 3 (Fig. 7). The whole procedure is a rather time-consuming trial and error procedure which, however, gives direct insight into the effects of parameter changes.

The theoretical seismograms were calculated with the reflectivity method for a typical earth model (Model SP-1 of Faber and Müller (1980)). Absorption was approximately taken into account by applying an acausal frequency-domain operator whose value of t^* was assumed to be 0.8 s for all stations. Instrument characteristics were also included. Seismogram modelling was attempted for the WWNSS stations PRE, WIN, NDI, POO, HKC and AFI which were at distances greater than 30° and not too close to the nodal planes of shock 1. Figure 9 shows the normalized moment function derivatives assumed for some of these stations, with the times of shocks 1, 2 and 3 indicated. Integrating these functions gives a gradual build-up of the moment function whose final value, after 15–20 s, is the moment of the earthquake. Theoretical seismograms, calculated for these source functions and three different double-couple orientations, are presented in Figure 10 in comparison with observed seismograms. Matching observed and theoretical seismograms at the first peak-to-peak amplitude gives the moment values indicated below the traces.

The fit of observed and theoretical waveforms in Figure 10 is always good during the first 15–20 s, which confirms the general character of the moment functions and the rupture duration of about 13 s implied by our model. For later times the agreement is of variable quality mainly because of the variable success in modelling the amplitudes of sP relative to those of P . The amplitude ratio of these two phases depends strongly on the source orientation. In the case of dipole 1 the radiation of sP relative to P (or of $sPKP$ relative to PKP for station AFI) is always too strong. For AFI this is due to the fact that this station is relatively close to the P -wave nodal plane E1 (see Fig. 2); this leads to an over-estimated moment of 7.2×10^{27}

dyne-cm. Disregarding this station, we infer a moment of $2\text{--}3 \times 10^{27}$ dyne-cm. In the case of dipole 3 there is, for the stations PRE, WIN and AFI, a considerable improvement of the fit for the *complete* pulses P^* and PKP^* , respectively. The stations KOD and NDI are now too close to the nodal plane D2 (see Fig. 7), such that for these stations there is less agreement of theory and observation than for the other stations. PRE, WIN and AFI point to a moment of $2\text{--}3 \times 10^{27}$ dyne-cm and thus support the value obtained for dipole 1. Finally, in the case of dipole 5, which has a roughly intermediate orientation between the dipoles 1 and 3, the success of pulse-form modelling is fair to excellent for all stations, and all moments fall into the range $2.4\text{--}3.0 \times 10^{27}$ dyne-cm. A cautious conclusion from these (and several other) modelling experiments, which takes into account the simplifying assumptions that we have made, is that the Romanian earthquake had a moment of $2\text{--}3 \times 10^{27}$ dyne-cm. This is somewhat higher than the value $1\text{--}2 \times 10^{27}$ dyne-cm, inferred by Hartzell (1979) from surface waves. In view of the uncertainties which enter the moment determination from both body and surface waves, it is probably not justified to consider this as an essential disagreement.

A dipole such as dipole 5 is probably a good average focal mechanism of the Romanian earthquake. For completeness we give the orientation of this dipole: the pole directions of the nodal planes have the azimuths 118° and 240° , and the dip angles 30° and 56° , respectively.

Discussion and Conclusions

The successful modelling of far-field seismograms supports our interpretation of shock 3 as being due to rupture termination at the SW end of the fault zone. Thus, the distance of about 50 km between shock 1 and shock 3, found in the master-event analysis, is a lower limit for the horizontal extent of the rupture surface. The dimension in the dip direction probably is neither larger nor much less than the horizontal dimension. We thus assume fault dimensions of 60 and 40 km, respectively, giving a rupture area of about 2,400 km². A rupture surface of this size falls well within the aftershock volume of the earthquake, as determined by Fuchs et al. (1979). For a moment value of 2.5×10^{27} dyne-cm and a rigidity of 7.0×10^{11} dyne/cm² in the focal zone we then obtain an average dislocation of about 2.2 m. If we apply the well-known formula for the stress drop on a circular dislocation we find a stress-drop estimate of 52 bar.

Hartzell (1979) studied the accelerograms of the Romanian earthquake, recorded at Bucharest, and compared the displacement records, obtained by integration, with theoretical seismograms, calculated for three different types of source models. His *line-source model*, consisting of a linear array of 11 point sources which are triggered successively over 13 s, comes closest to our rupture model, since the time during which the moment of the earthquake is built up is about the same as in our model. It appears possible that long-period, far-field seismograms calculated for this model (after a change in the orientation of the line source) resemble observed seismograms. Hartzell's *one-point source models*, however, correspond to a rupture duration of only about 1 s. These models are able to explain the periods and pulse duration of the Bucharest displacement records, but the long-period far-field seismograms of the earthquake definitely cannot be modelled by them. The same is true for a *two-point source model* by which Hartzell tries to represent in a simple way the multiple nature of the earthquake. The second source, which has the opposite polarity of the first and radiates about 15 s later, is thought to model the termination of rupture (which,

in actuality, it does not do: it models a second, separate and complete earthquake). Since its strength is only 1/4 of the strength of the first source, the additional radiation does not essentially change the seismograms of the one-point source model.

In conclusion, if we try to combine Hartzell's results and ours, it seems possible that a line-source model made up by a pointshear dislocation, moving horizontally over a distance of 50 km from NE to SW with a velocity of about 3.7 km/s, can be found which explains the main features of both near-field and far-field seismograms. The orientation of the shear dislocation and the moment release as functions of time are constrained by our results, but there appears to be enough freedom left in these quantities for a successful modelling of the near-field seismograms.

Acknowledgments. We are grateful to Professor N. Kondorskaya, Institute of Physics of the Earth, Moscow, for making available the records of the Russian seismic network. The computations for this paper were performed at the computing center, University of Karlsruhe. We thank H. Berckhemer, W. Brüstle, S. Faber, and R. Kind for helpful discussions and reading the manuscript, and I. Hörnchen for typing it.

References

- Faber, S., Müller, G.: Sp phases from the transition zone between the upper and lower mantle. *Bull. Seismol. Soc. Am.* **70**, 487–508, 1980
- Fuchs, K., Bonjer, K., Bock, G., Cornea, I., Radu, C., Enescu, D., Jianu, D., Nourescu, A., Merkle, G., Moldoveanu, T., Tudorache, G.: The Romanian earthquake of March 4, 1977. II. Aftershocks and migration of seismic activity. *Tectonophysics* **53**, 225–247, 1979
- Hartzell, S.: Analysis of the Bucharest strong ground motion record for the March 4, 1977 Romanian earthquake. *Bull. Seismol. Soc. Am.* **69**, 513–530, 1979
- Müller, G., Bonjer, K., Stöckl, H., Enescu, D.: The Romanian earthquake of March 4, 1977. I. Rupture process inferred from fault-plane solution and multiple-event analysis. *J. Geophys.* **44**, 203–218, 1978
- Räkers, E.: Untersuchungen zum Bruchvorgang des rumänischen Erdbebens am 4.3. 1977 und Bestimmung des seismischen Moments mit Hilfe von synthetischen Seismogrammen für P-Wellen. Diploma thesis, University of Karlsruhe, 131 p. 1981

Received October 27, 1981; Revised version December 21, 1981
Accepted January 20, 1982

An Improved Algorithm for Magnetotelluric and Direct Current Data Interpretation

S. Goldberg¹, D. Loewenthal², and Y. Rotstein¹

¹ Institute for Petroleum Research and Geophysics, Holon, Israel

² Dept. of Planetary Sciences and Geophysics, Tel-Aviv University, Tel-Aviv, Israel

Abstract. A simple iterative inversion method for MT and DC data, which is a significant improvement of the usual automatic trial and error process, is presented. The basic idea has long been used, i.e. inversion using approximate solution and exact forward computation for check of fit. In the new version the iterations are performed in the resistivity vs. depth domain instead of the apparent resistivity vs. frequency in MT or apparent resistivity vs. electrode separation in DC soundings. This modification totally avoids the appearance of the physically meaningless solutions which are a major obstacle in the common iterative procedure. Thus a sufficiently good fit is always reached after only a few iterations. As in all similar inversion methods, a solution consists of as many layers as the number of sampled points. The number of layers may be reduced by various methods. A simple interactive method for the reduction of number of layers, which was found very useful, is presented. It is based on an approximate solution which enables the simultaneous creation of a wide range of equivalent sections from the original multilayered section derived in the inversion process. The new inversion and reduction of layers procedures can be easily programmed. They have been applied to synthetic as well as on numerous field soundings and were found to be powerful in the interpretation of field surveys.

Key words: Inversion – Magnetotellurics – Direct current resistivity soundings

Introduction

A quantitative interpretation of magnetotelluric and direct current field data is still the most problematic and critical step in magnetotelluric (MT) and DC resistivity surveys. Since two and three dimensional interpretation is not routinely possible for practical reasons, a one-dimensional interpretation is the most which may be expected from the exploration geophysicists who handle a large number of soundings.

Anyone who has tried to carry out an interpretation based on a purely automatic process will probably agree that such interpretation is often very far from being conclusive and realistic. The reason for the discrepancy between true resistivities and the derived one, apart from inhomogeneity, is in the existence of a wide range of equivalent solutions which

are indistinguishable for geoelectrical methods. Thus it seems reasonable to derive some basic multilayered section which fits the field data sufficiently well and then to create a wide range of equivalent sections from which the most suitable section may be chosen (with the aid of additional geological data).

Two main automatic inversion methods for MT and DC data are widely used. One of them is based on least mean square fitting technique (Wu 1968; Patrick et al. 1969; Inman et al. 1973). The other one is an iterative method based on approximate (asymptotic) inversion (Maillet 1947; Berdichevsky 1965; Bostick 1977) and an exact forward computation for comparison of the solution with the data (Zohdy 1973, 1975; Chaipayungpun and Landisman 1977).

The least mean square technique is based on the generalized linear inverse theory. This method requires the exhaustive computation of all the eigenvalues and eigenvectors of the system matrix. By introducing logarithmic variables, controlling the eigenvalue cut-off and looking at the eigenvectors, equivalences are detected.

The second method seems to be very attractive since it fits exactly the given set of data without any need for an initial guess. The disadvantage of this method, when used in the usual way, is in the appearance of physically meaningless solutions such those that include negative resistivities or negative thicknesses. The modification of the original scheme, made in order to avoid this problem (Zohdy 1973; Chaipayungpun and Landisman 1977), usually results in loss of accuracy.

An improved algorithm, suitable for the inversion of MT apparent resistivity and DC Kernel function curves, is presented. This algorithm is absolutely free from physically meaningless resistivities and thicknesses and always leads to a high degree of accuracy. The algorithm is based on fitting of resistivity-depth curves, which are a product of asymptotic inversion, instead of apparent resistivity curves.

The number of layers derived in the inversion process is equal to the number of sampled points. In order to get a useful solution, the number of layers must be reduced. This may be done by an automatic method (Zohdy 1975), but a simple and more powerful technique is presented here. This technique, also based on the asymptotic equations, allows not only a reduction in the number of layers, but also the simultaneous creation of a wide range of equivalent sections without significant loss of accuracy. The use of this technique permits the inclusion of all available geological and geophysical data, thus leading to the best possible subsurface section.

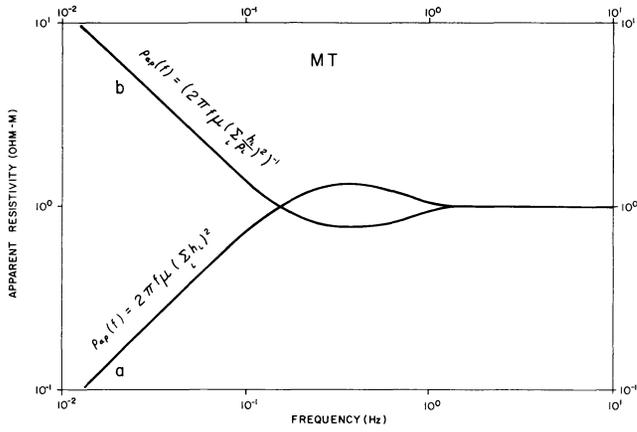


Fig. 1. MT apparent resistivity curves computed for a single layer (1 ohm-m) 100 m thick, underlayed by an ideal conducting (curve a) and an ideal insulating (curve b) half space

The algorithms for inversion and for reduction of the number of layers were applied to a large number of synthetic and field examples. Our experience shows that the interpretation procedure which is described in this paper is very useful due to simplicity, the consistent, meaningful results and its almost negligible cost.

Approximate (Asymptotic) Expressions for Inversion of MT and DC Curves

The idea of deriving an approximate solution by computing apparent resistivity as a point of intersection of two low frequency asymptotes, was originally introduced by Bostick (1977) for an inversion of MT data, assuming a continuous change of resistivity with depth. Using the same method, similar asymptotic solutions for discrete changes of resistivity may be derived. The same approach may be extended to the case of DC and thus the nature of the expressions used by Zohdy (1973, 1975) for the inversion of Schlumberger apparent resistivity curves can be clarified.

Figures 1 and 2 illustrate MT apparent resistivity and DC Kernel function curves for a single layer ($\rho = 1.0$ ohm-m, $h = 100$ m) overlaying a perfect conductor (curve a) or a perfect insulator (curve b). One can see that the apparent resistivity (Kernel function for DC) values at frequencies higher than the frequency of intersection for both low frequency asymptotes are only slightly affected by an infinite change in resistivity of the half-space. Furthermore, at the frequency of the asymptotes intersection, the corresponding apparent resistivity (Kernel function) values appear to be a reasonable single-value, central estimate of the small range of values, over which the exact apparent resistivity varies in response to all possible changes in the underlying half-space conductivity. Thus it may be assumed that it is almost independent of underlying structure as well. Therefore, at any frequency the apparent resistivity value can be approximately presented as a common value of both low frequency asymptotes.

Starting with a well known recursive formula for direct computation of electromagnetic impedance over a one dimensional layered earth (Word et al. 1970), asymptotic expressions for apparent resistivity (ρ_{ap}) are derived:

$$\lim_{\omega \rightarrow 0} \rho_{ap}/\omega = \mu \left(\sum_{j=1}^{N-1} h_j \right)^2, \text{ for a case of conducting half space } (\rho_N \rightarrow 0)$$

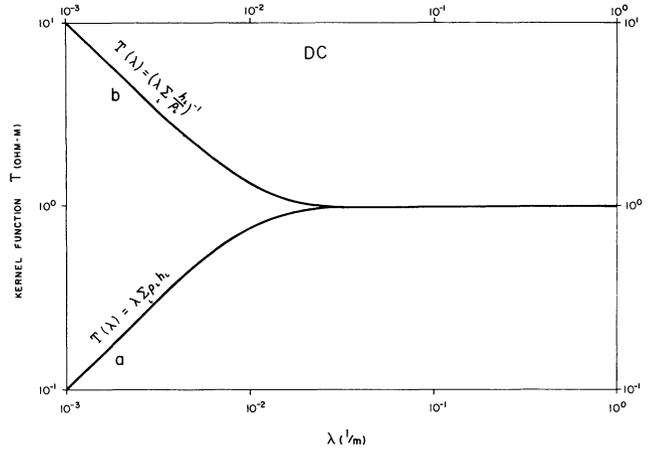


Fig. 2. DC kernel function computed for a single layer (1 ohm-m), 100 m thick, underlayed by an ideal conducting (curve a) and an ideal insulating (curve b) half space

and $\lim_{\omega \rightarrow 0} \omega \rho_{ap} = 1 / \left[\mu \left(\sum_{j=1}^{N-1} h_j/\rho_j \right)^2 \right]$, for a layered medium overlaying a perfect insulator ($\rho_N \rightarrow \infty$).

As usual, ω denotes the frequency μ is the permeability, j is the layer numbered from the top and N is the number of layers.

Combining those asymptotic expressions, we get a system of equations which make possible an approximate computation of MT apparent resistivity in terms of sub-surface parameters:

$$\begin{aligned} \rho_{ap}(\omega_n) &= \mu\omega \left(\sum_{j=1}^n h_j \right)^2 \\ \rho_{ap}(\omega_n) &= 1 / \left[\mu\omega \left(\sum_{j=1}^n h_j/\rho_j \right)^2 \right] \end{aligned} \quad (1)$$

where n is the deepest layer (it may be a fraction of a layer as well) which still affects the apparent resistivity at frequency ω_n .

Slightly modifying and introducing $D_n = \sum_{j=1}^n h_j$ - the depth to the $(n+1)$ -th layer:

$$\begin{aligned} D_n &= (\rho_{ap}(\omega_n)/\mu\omega_n)^{1/2} \\ \rho_{ap}(\omega_n) &= D_n^2 / \left(\sum_{j=1}^n h_j/\rho_j \right)^2 \end{aligned} \quad (2)$$

Equation (2) was originally derived by Jain (1966) using simple physical concepts. The expression found here for D_n is just $1/\sqrt{2}$ of the homogeneous earth skin depth and is close to two-thirds of the skin depth experimentally found and used by Jain.

Since $\sum_{j=1}^{n-1} h_j/\rho_j = D_{n-1}/\rho_{ap}(\omega_{n-1})$ and $h_n = D_n - D_{n-1}$, convenient expressions for approximate inversion of MT apparent resistivity curves can be derived from Eq. (2) (a decreasing order of frequencies is assumed):

$$\begin{aligned} \rho_n &= h_n / \left(\frac{1}{\sqrt{\rho_{ap}(\omega_n)\omega_n\mu}} - \frac{1}{\sqrt{\rho_{ap}(\omega_{n-1})\omega_{n-1}\mu}} \right) \\ h_n &= \frac{\rho_{ap}(\omega_n)}{\omega_n\mu} - \frac{\rho_{ap}(\omega_{n-1})}{\omega_{n-1}\mu}, \quad n = 2, \dots, N. \end{aligned} \quad (3)$$

$\rho_1 = \rho_{ap}(\omega_1)$ and $h_1 = \sqrt{\rho_{ap}(\omega_1)/\mu\omega_1}$ are obtained from (1). h_N , being the thickness of a half space, is meaningless, and is required only for the computation of the last resistivity ρ_N .

The number of layers derived by this method is equal to the number of sampled points. This does not mean that it is the actual number of layers since such a layer may be only a fraction of a real geological layer.

The depth to the n -th layer depends on $\rho_{ap}(\omega_n)$ and ω_n only. Therefore, it is possible to invert independently separate parts of the apparent resistivity curves. This property is most significant when, due to a poor signal to noise ratio, the apparent resistivity data is not continuous over a wide frequency band, but rather consists of a few discrete data segments.

Assuming a continuous change of resistivity with depth and using the same asymptotic expressions, it is possible to derive a similar set of equations for MT apparent resistivity data inversion (Bostick 1977):

$$D = \sqrt{\frac{\rho_{ap}(\omega)}{\omega\mu}},$$

$$\rho = \rho_{ap}(\omega) \frac{1 - d(\log \rho_{ap}(\omega))/d(\log \omega)}{1 + d(\log \rho_{ap}(\omega))/d(\log \omega)}.$$

The asymptotic expressions for DC Kernel function are derived from direct computation recursive formula:

$$\lim_{\lambda \rightarrow 0} T/\lambda = \sum_{j=1}^{N-1} \rho_j h_j \text{ for a layered medium overlaying a conducting half space } (\rho_N \rightarrow 0) \text{ and}$$

$$\lim_{\lambda \rightarrow 0} \lambda T = 1 \left/ \left(\sum_{j=1}^{N-1} h_j / \rho_j \right) \right. \text{ for a layered medium overlaying a perfect insulator } (\rho_N \rightarrow \infty).$$

Combining the asymptotic expressions we get a system of equations which makes possible an approximate computation of DC Kernel function values in terms of sub-surface parameters:

$$T(\lambda_n) = \lambda_n \sum_{j=1}^n \rho_j h_j \quad (4)$$

$$T(\lambda_n) = 1 \left/ \left(\lambda_n \sum_{j=1}^n h_j / \rho_j \right) \right.$$

From Eq. (4), using $\sum_{j=1}^{n-1} \rho_j h_j = T(\lambda_{n-1})/\lambda_{n-1}$ and $\sum_{j=1}^{n-1} h_j / \rho_j = 1/(T(\lambda_{n-1}) \cdot \lambda_{n-1})$, convenient expressions for the inversion of DC Kernel function curves are derived:

$$h_n = \sqrt{\left(\frac{T(\lambda_n)}{\lambda_n} - \frac{T(\lambda_{n-1})}{\lambda_{n-1}} \right) \left(\frac{1}{\lambda_n T(\lambda_n)} - \frac{1}{\lambda_{n-1} T(\lambda_{n-1})} \right)}$$

$$\rho_n = \sqrt{\left(\frac{T(\lambda_n)}{\lambda_n} - \frac{T(\lambda_{n-1})}{\lambda_{n-1}} \right) \left(\frac{1}{\lambda_n T(\lambda_n)} - \frac{1}{\lambda_{n-1} T(\lambda_{n-1})} \right)}$$

$\rho_1 = T_1$ and $h_1 = 1/\lambda_1$ are obtained from (4).

Similar expressions were used by Zohdy (1975) in the inversion of Schlumberger DC apparent resistivity curves (but with apparent resistivity instead of Kernel function). The slopes of the Kernel function are limited by ± 1 , whereas the slopes of DC apparent resistivity curves may exceed -1 . Thus the use of expression (5), which is well suited to kernel curve inversion, in the inversion of apparent resistivity curves, requires certain modifications (Zohdy 1973, 1975).

Inversion of MT Apparent Resistivity and DC Kernel Curves by an Iterative Process

The expressions for MT and DC curve inversion derived in the previous section represent a low frequency approximation. There will be an inherent discrepancy between the original curves and the curves computed by an exact forward computation for the approximately derived resistivity section. Furthermore, the approximate multi-layered section includes almost gradual changes of resistivity instead of well defined interfaces.

It was shown by Zohdy (1973, 1975) for DC, and by Chaipayungpun and Landisman (1977) for MT, that improving the accuracy of the asymptotic method is possible by an iterative process based on the same asymptotic expressions. The most serious limitation of these proposed methods is that they produce slopes which exceed the range of -1 to $+1$ and thus lead to meaningless solutions. The modifications of the original iterative process scheme, introduced by Zohdy (1973, 1975) and Chaipayungpun and Landisman (1977) in order to avoid the above mentioned problem, may result in a significant loss of accuracy or an inability to converge.

In an effort to avoid unreasonable slopes and the problems which they create, a new algorithm has been developed. The proposed algorithm is based on the existence of a one-to-one correspondence between the apparent resistivity (MT) and Kernel function (DC) curves or resistivity sections and the corresponding approximate resistivity sections derived using the asymptotic solutions. This fact enables us to compare and fit these resistivity sections instead of apparent resistivity curves.

Since approximate resistivity sections for DC soundings are computed from Kernel function values, the low resolving power of Kernel function curves for a C-type ($\rho_1 > \rho_2 > \rho_3$) and certain HK-type ($\rho_1 > \rho_2 < \rho_3 > \rho_4$) sections is inherent in corresponding approximate resistivity sections and may result in a final solution which fits the Kernel function curve well but does not necessarily fit the apparent resistivity curve. Therefore, as in interpretation in the Kernel domain, the calculation of apparent resistivity curves for comparison with the field data is advisable, especially for curves with steeply descending branches (Zohdy 1975).

Starting with any section, usually with an asymptotic solution for the given apparent resistivity curve, we shall compare the approximate sections corresponding to this particular section and to the given apparent resistivity curve. We assume that the disagreement existing between the approximate sections is close to the disagreement existing between the exact solutions.

Thus, we are able to change, step by step, the previous guess toward the exact solution. Resistivity and thickness values are compared and modified, using a logarithmic scale, so that negative values are avoided. The different steps for the proposed method are as follows:

1. From the given apparent resistivity curve $\rho_{ap}^m(\omega_i)$, $i = 1, 2, \dots, n$, where i is the index number of each sample, using asymptotic expressions, compute the corresponding resistivity section ρ_i^m, h_i^m , $i = 1, 2, \dots, n$. In order to preserve the one-to-one correspondence existing between the samples of apparent resistivity and the segments of the resistivity section, the samples must be arranged from high frequencies towards the lower ones, if the layers are counted downwards.

2. Use ρ_i^m, h_i^m , $i = 1, 2, \dots, n$ as ρ_i, h_i , $i = 1, 2, \dots, n$ - a first approximation of the true layering.

3. From $\rho_i, h_i, i=1, 2, \dots, n$ compute $\rho_{ap}^c(\omega_i), i=1, 2, \dots, n$ by exact forward computation.

4. Compare $\rho_{ap}^c(\omega)$ with the given curve $\rho_{ap}^m(\omega)$ at all ω_i ; if the requested fit is achieved then $\rho_i, h_i, i=1, 2, \dots, n$ is the requested solution.

5. If significant disagreement between ρ_{ap}^c and ρ_{ap}^m still exists, from $\rho_{ap}^c(\omega)$, using asymptotic inversion, compute the corresponding resistivity section $\rho_i^c, h_i^c, i=1, 2, \dots, n$.

6. Change the values of ρ_i and h_i according to the disagreement between ρ_i^m and ρ_i^c, h_i^m and h_i^c , i.e. replace $\log \rho_i$ by $(\log \rho_i + \log \rho_i^m - \log \rho_i^c)$ and $\log h_i$ by $(\log h_i + \log h_i^m - \log h_i^c)$ for all and use the new values in the next iteration. Return to step 3.

All the stages are valid for DC Kernel curve inversion if $\rho_{ap}(\omega)$ is changed to $T(\lambda)$.

Inversion of MT Data

The proposed method of iterative inversion was applied to a large number of synthetic and field data. The synthetic apparent resistivity curves were computed using the usual forward computation technique (Word et al. 1970). The iterative inversion, when used in all practical cases, was found to be very fast (3 to 5 iterations to get an accuracy better than 3–5%), successful and sufficiently accurate. However, in order to demonstrate the features of this method in the present discussion we shall look for the usually unreasonable relative accuracy of 10^{-6} and we shall limit the maximum number of iterations to 300. As will be illustrated, the significant improvement occurs after just 3 to 5 iterations. Higher number of iterations, even if improving the fit, usually lead to geologically meaningless solutions.

A typical example is presented in Figure 3. When two samples per decade are used, the convergence is very fast and the requested accuracy (10^{-6}) is usually achieved after only 20–40 iterations. A problem arises, however, that except for the sampled points, the accuracy was only slightly improved when compared to the results of the direct asymptotic inversion (Fig. 3b). When the apparent resistivity curve is sampled at four points per decade, the accuracy is improved over all of the curve (Fig. 3c), but the convergence becomes very slow and the requested fit (10^{-6}) was not achieved even after 300 iterations.

In Figure 4 the corresponding resistivity – depth cross-sections are presented (wavy line at the bottom of the resistivity profile represents the transition to the homogeneous half space): derived without iterations (a); after three iterations (b); after 300 iterations (c). The apparent resistivity curve in this example was sampled at four points over a decade.

The solution reached after three iterations (Fig. 4b) seems to be the one most similar to the given subsurface cross-section. The solution derived after 300 iterations (Fig. 4c), which may be expected to be the best, since the corresponding apparent resistivity curve is the closest to the one given, is actually the worst. Such a correlation indicates that a good fit, with an accuracy much better than that usually available in field data, does not guarantee that the solution reached is the real one rather than one of a wide range of equivalent solutions. This fact illustrates the limited resolution of the magnetotelluric method: very simplified solution, like the sections used in this example, may be found in case of a very complicated geology, similar to the solution found after 300 iterations.

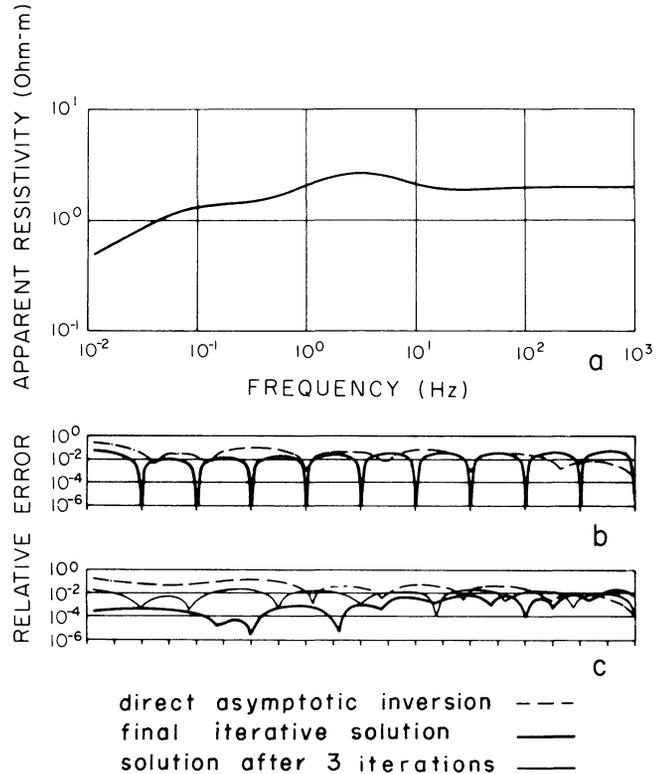


Fig. 3a-c. Inversion of synthetic model – error analysis (MT): **a** apparent resistivity curve (5 layers); **b** relative error for rough sampling (2 points per decade); the iterative process was terminated when the requested accuracy was reached at in the sampled points; **c** relative error for increased sampling (4 points per decade); the iterative process was terminated after 300 iterations

For practical purposes, it is customary to derive a simple solution as a geological model. With this in mind, the solution derived after only a few iterations seems to be sufficient in the sense that both the apparent resistivity and the resistivity sections are similar to the synthetic ones.

The solution derived using this inversion method consists of as many layers as there are sampled points. In the next section, a method of interpretation based on the reduction of the number of layers is presented.

Reduction of a Number of Layers – a Practical Approach to the Interpretation of MT Data

A product of 3–5 iterations, such as those presented in Figure 4b with all of the disadvantages mentioned, usually cannot be used as a final solution. In its present form the solution consists of a large number of layers while a geological interpretation based on several major lithological layers is usually needed. Thus, the next step must be a reduction of the number of layers with a minimal loss of accuracy achieved in fitting the apparent resistivity curves.

The technique presented in this section is a very simple one and is based on the same asymptotic expressions used in the inversion process. This method was found to be a very effective tool in interpretation, since it makes possible the inclusion of all the known subsurface information from the surveyed area.

Let us examine the section derived using the proposed iterative inversion method. Instead of only a few thick homogeneous and geologically justifiable layers, this section con-

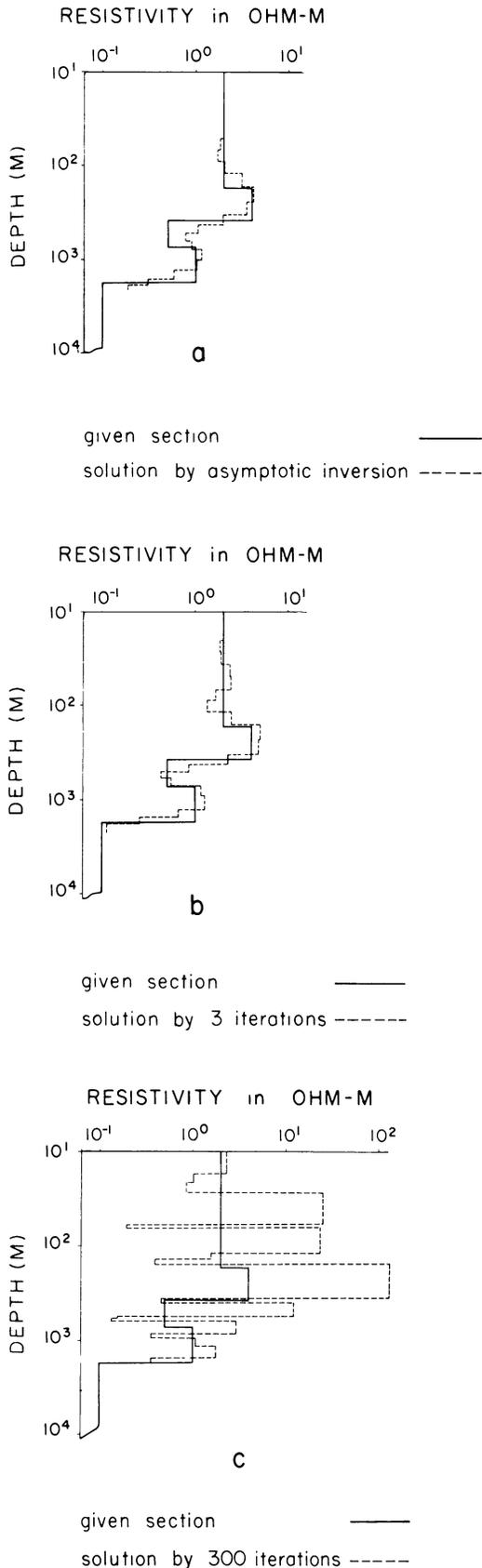


Fig. 4a-c. Subsurface resistivity structure derived for synthetic model (shallow layers are omitted): a by direct asymptotic inversion; b by 3 iterations; c by 300 iterations

sists of a large number of layers with an almost gradual change of resistivity from one layer to another (Fig. 4b). If it is certain that two layers (if no geological data is available, layers with extreme resistivities may be used) k and l belong to two neighbouring geological layers, it is necessary to divide and to add (in a certain way) the interlying pseudolayers $k+1, k+2, \dots, l-1$ to the layer k and l . Thus, an equivalent two-layers solution with resistivities ρ_k and ρ_l and corresponding thicknesses $h_k + \Delta h_k$ and $h_l + \Delta h_l$ will be found for this part of the multilayered section.

If the resistivities of the two-layer section are defined a priori, the choice of Δh_k and Δh_l must be made in such a way that the corresponding apparent resistivity curve will remain similar to the one corresponding to the original multilayered section.

It is clear that if

$$\Delta h_k + \Delta h_l = \sum_{j=k+1}^{l-1} h_j \quad (6)$$

$$\frac{\Delta h_k}{\rho_k} + \frac{\Delta h_l}{\rho_l} = \sum_{j=1}^{l-1} \frac{h_j}{\rho_j},$$

the apparent resistivity values, computed using approximate forward computation expressions (1), will not change at frequencies $\omega_1, \dots, \omega_k, \dots, \omega_n$ and the curve will be smoothed only between ω_k and ω_l . It was found that even the exact computed apparent resistivity curves remain almost unchanged if equalities (6) hold. Consequently, dividing the multilayered section, using this method, a geologically meaningful equivalent section can be obtained. If the existence of any layer with a given resistivity is known, and it does not appear in the resistivity section produced by the automatic inversion process, it can be added using a zero thickness at a suitable depth; through Eq. (6), its interfaces with the neighbouring layers can then be defined.

Sometimes, negative values for $h_k + \Delta h_k$ or $h_l + \Delta h_l$ may be reached. Usually, this happens only when an unsuitable choice of k and l is made.

Finally, when the requested number of layers is reached, it is necessary to compute the corresponding apparent resistivity curve and compare it with the given one in order to guarantee that the final solution remains satisfactory.

In Table 1 the possibility for layer number reduction based on Eq. (6) in the case of a synthetic model is illustrated. At the first step, a 20-layered section presented in Fig. 4b was derived by the iterative method. The iterative process was terminated after three iterations. The relative error reached in corresponding apparent resistivity curve fitting is presented in Figure 3c; the errors are less than 2.8% (without iterations -16.6%). Using the method described, the number of layers was reduced to five. Extremely high and low resistivity values were chosen to represent the actual structure. As seen from Table 4 the derived section is close to the synthetic one. The fit of the apparent resistivity curve for the final solution is better than 4.4%.

A field example is presented in Fig. 5. A good fit of both apparent resistivity and phase data is evidently reached by the proposed method.

Inversion and Interpretation of MT Apparent Resistivity Curves - Summary

As shown above relatively good accuracy may be reached by the proposed iterative method after only a few iterations. This

Table 1. Original resistivity model and a resistivity depth section derived from the computed MT apparent resistivity values using the proposed inversion routine and layer number reduction scheme

Layer No.		1	2	3	4	5
Synthetic model	ρ (ohm-m)	2.00	4.00	0.50	1.00	0.10
	h (m)	170.00	300.00	250.00	1,000.00	—
Derived section	ρ (ohm-m)	2.01	5.04	0.43	1.28	0.12
	h (m)	176.78	294.96	211.26	993.28	—

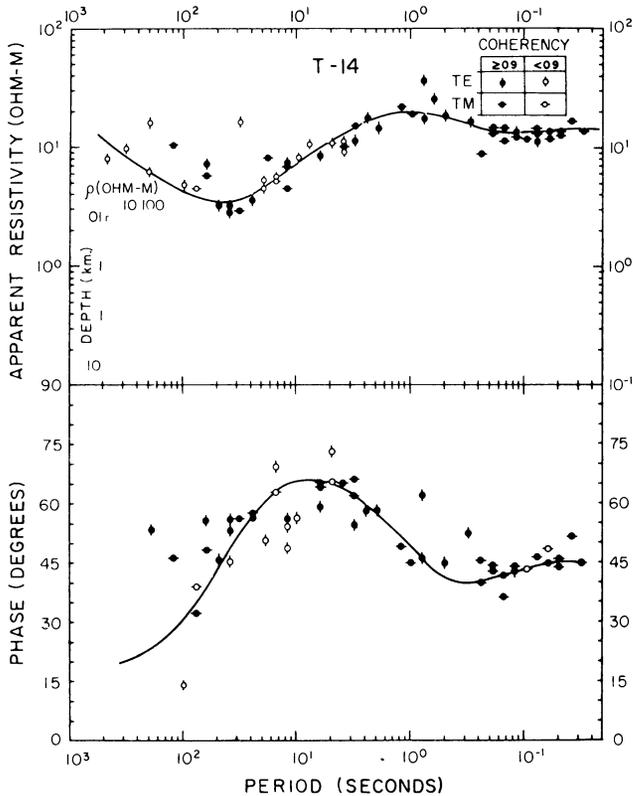


Fig. 5. Field example for MT apparent resistivity curve fitting

accuracy is only slightly affected by the layer number reduction. Thus, the different stages of the proposed inversion and interpretation process may be summarized as follows:

(a) Sampling of the apparent resistivity curve, 3–4 samples per decade (not necessarily uniformly spaced); if the data is noisy, the phase values may be taken into account in order to define the general shape of the noisy segments more clearly¹.

Since the inversion process does not use phases, the resulting phase curve in the presence of noise may be inconsistent with the field measured data even when sufficient accuracy of apparent resistivity is reached.

(b) Inversion of the sampled apparent resistivity curve using the iterative process; from experience, reasonable values for accuracy requested are 3 to 6%. While trying to improve accuracy, the number of iterations will essentially increase, thus leading to very complicated sections which do not allow any reasonable geological interpretation. For this same reason, the number of iterations must be limited to 7–10 (it was

1 The relation between phases and the slope is given approximately

$$\text{by } \frac{d \log \rho_{ap}(\omega)}{d \log \omega} = \frac{4}{\pi} \phi - 1 \text{ where } \phi \text{ is phase (Bostick 1977)}$$

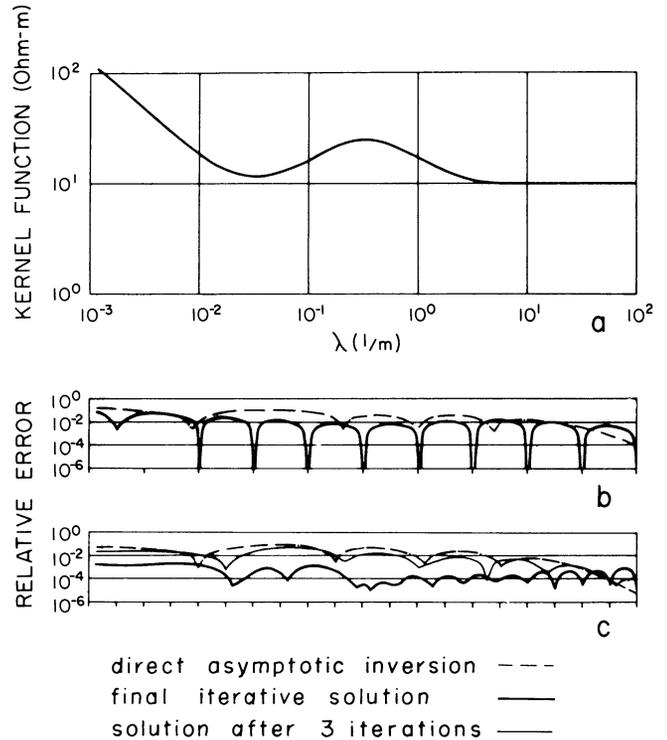


Fig. 6a-c. Inversion of synthetic model – error analysis (DC): **a** apparent resistivity curve (5 layers); **b** relative error for rough sampling (2 points per decade); the iterative process was terminated when the requested accuracy was reached at the sampling points; **c** relative error for increased sampling (4 points per decade); the iterative process was terminated after 300 iterations

found that the most significant improvement in the fitting of apparent resistivity curves occur on just the first few iterations). The requested accuracy may, possibly, not be reached at several samples. Usually, this occurs when dealing with very noisy data (it does not happen at all with synthetic data).

In such a case it would be advantageous to compare the accuracy reached with the data quality.

(c) Interpretation of the solution using the layer number reduction technique. All available geological information must be used at this step in order to obtain a reasonable solution. The interpretation must be made carefully and the limited resolution of the MT method must always be considered.

(d) Finally, it is necessary to compute the apparent resistivity curve corresponding to the final solution in order to make sure that the derived geological section is consistent with the measured data.

Inversion of DC Data

The proposed method of iterative inversion was applied to a large number of synthetic and field measured DC data.

In the following example (Fig. 6), in order to demonstrate the inversion techniques, the iterative process was terminated when accuracy better than 10^{-6} was reached or after 300 iterations. However, in all practical cases the field data quality is not better than 1 to 5%. For this level of accuracy to be reached in the inversion 10 to 20 iterations are usually sufficient.

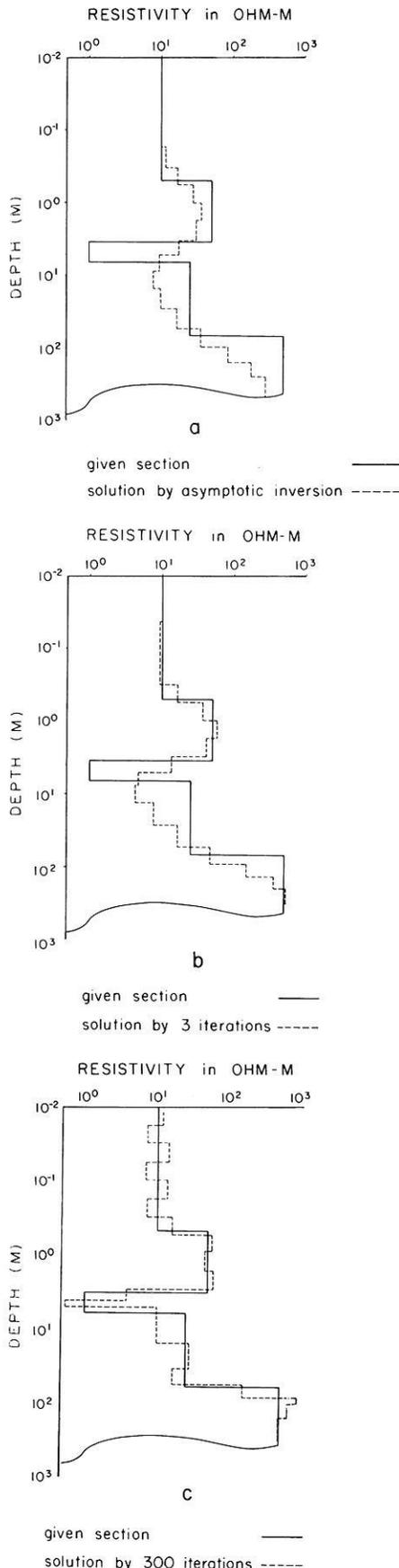


Fig. 7a-c. Subsurface resistivity structure derived for synthetic model (shallow layers are omitted); a by direct asymptotic inversion; b by 3 iterations; c by 300 iterations

Table 2. Original resistivity model and two possible solutions: a resistivity depth section A derived from the computed DC kernel function values and section B improved solution based on "known" resistivity value for the third layer

Layer No.		1	2	3	4	5
Synthetic model	ρ (ohm-m)	10.00	50.00	1.00	20.00	500.00
	h (m)	0.50	3.00	3.00	60.00	—
Derived section A	ρ (ohm-m)	9.87	57.49	0.46	27.95	493.75
	h (m)	0.51	2.62	1.62	69.96	—
Derived section B	ρ (ohm-m)	9.87	57.49	1.00	27.95	493.75
	h (m)	6.51	2.62	3.50	69.88	—

The corresponding resistivity-depth section reached by an asymptotic inversion is presented in Fig. 7a. The resulting sections reached after 3 and 300 iterations are presented in Fig. 7b and c, respectively. Unlike magnetotellurics, except for shallow depth, there is a gradual improvement in the resulting solution as the number of iterations increases.

Reduction of Number of Layers – a Practical Approach to the Interpretation of DC Data

The final product of the proposed iterative process usually cannot be accepted as a final geological section because of the larger number of layers involved. Therefore, a geological interpretation based on several major geological units is still needed.

The technique of layer number reduction, presented for magnetotellurics, may also be successfully used in DC interpretation. The corresponding expressions resulting from Eqs. (4) are as follows:

$$\rho_k \Delta h_k + \rho_l \Delta h_l = \sum_{i=k+1}^{l-1} \rho_i h_i \quad (9)$$

$$\frac{\Delta h_k}{\rho_k} + \frac{\Delta h_l}{\rho_l} = \sum_{i=k+1}^{l-1} \frac{h_i}{\rho_i}$$

The layer number reduction is illustrated in Table 2. First, a 20-layered section was derived by the iterative inversion process (300 iterations, see Fig. 7c). The accuracy reached in the fitting of kernel function curves was better than 0.5% (compared with 8% without iterations). The layer number reduction made by the proposed method only slightly effected the kernel function values, and for a 5-layered Sect. (A), the accuracy was better than 1.3%.

Section B presents a possible choice for improving the final solution using available data. It can be seen that even when a good accuracy was reached, a serious, almost 50% disagreement exists between the third layer of the given section and the derived one (Sect. B). In order to improve the final solution, a 1 ohm-m, zero thickness layer was added at an appropriate depth to the original 20-layered section. Then the number of layers was reduced using the usual method. The final solution (B) was consequently improved and is closer to the given one than section A. The accuracy in the fitting of the kernel function curves is still better than 1.3%, the same as for Sect. A.

The possibility of achieving a good fit to a given set of data with a number of distinctly different resistivity sections, illustrates a well known limitation of the DC sounding method, i.e. the existence of a wide range of equivalent solutions.

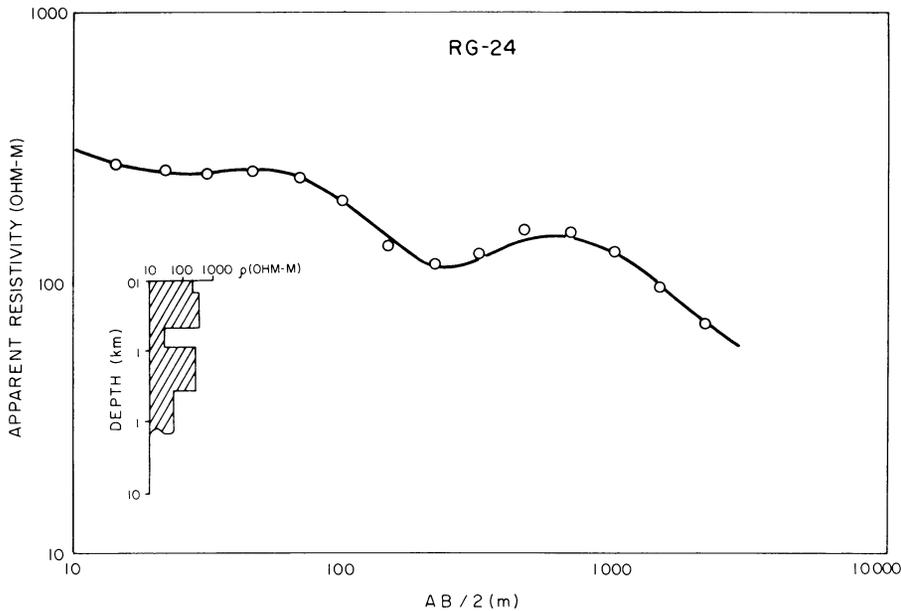


Fig. 8. Field example for DC apparent resistivity curve fitting

An advantage of the proposed method is the possibility of easily creating a wide range of equivalent solutions and choosing the most suitable one. The range of acceptable solutions may be narrowed down by using the available geological information, as illustrated in the discussion on synthetic sounding.

The scheme of inversion and interpretation of DC data is similar to the one presented for MT. Since the inversion is made on kernel function values, an appropriate transformation of the field apparent resistivity data must be done first (Koefoed 1968; Gosh 1971).

A field example is presented in Fig. 8. The apparent resistivity curve drawn corresponds to the derived resistivity section and it illustrates the good fit reached by the proposed method.

Summary

A simple and yet very effective technique for inversion of MT and DC data, using approximate equations together with exact forward computation for checking the solutions was described. This technique was extensively used in interpretation of field MT and DC data. It was found to be very effective and inexpensive.

The algorithm developed may be programmed easily using a small amount of computer storage; the computer time needed for inversion on a large computer is negligible in all practical cases. When applied on a CDC 6600 computer, inversion of usual MT and DC curves, i.e. 10 iterations of up to 15 samples lasts about half a second. An additional few milliseconds are required for layer number reduction and final interpretation.

References

- Berdichevsky, M.N.: Electrical prospecting with the telluric current method (English translation). Quarterly of the Colorado School of Mines, **60**, No. 1, 1965
- Bostick, F.X., Jr.: A simple and almost exact method of MT analysis (abstract): Workshop on Electrical Methods in Geothermal Exploration, Snowbird, Utah, 1977
- Chaipayungpun, W., Landisman, M.: Crust and upper mantle near the western edge of the Great Plains: Geophysical Monograph 20, American Geophysical Union, Washington, D.C. 1977
- Gosh, D.P.: The application of linear filter theory to the direct interpretation of geoelectric resistivity sounding measurements. *Geophys. Prosp.* **19**, 192-217, 1971
- Inman, J.R., Ryu, J., Ward, S.H.: Resistivity inversion. *Geophysics* **38**, 1088-1108, 1973
- Jain, S.: A simple method of magnetotelluric interpretation. *Geophys. Prosp.* **14**, 143-148, 1966
- Koefeld, O.: The application of the kernel function in interpreting geoelectrical resistivity measurements. Geophysical Monograph Series 1 No. 2, Gebruder, Borntraeger, 1968
- Maillet, R.: The fundamental equations of electrical prospecting. *Geophysics* **12**, 529-556, 1947
- Patrick, F.M., Bostick, F.X., Jr.: Magnetotelluric modeling techniques. E.G.R.L. Technical Report No. 59, University of Texas, Austin, Texas, 1969
- Word, D.R., Smith, H.W., Bostick, F.X., Jr.: An investigation of the magnetotelluric impedance method. E.G.R.L. Technical Report No. 82, University of Texas, Austin, Texas, 1970
- Wu, F.T.: The inverse problem of magnetotelluric sounding. *Geophysics* **33**, 972-979, 1968
- Zohdy, A.: Automatic interpretation of resistivity sounding curves using modified Dar Zarrouk functions (abstr.). *Geophysics* **38**, 196-197, 1973
- Zohdy, A.: Automatic interpretation of resistivity sounding curves using modified Dar Zarrouk functions. US Geol. Surv. Bull. **1313-E**, 1975

Received November 27, 1980; Revised version January 15, 1982
Accepted January 21, 1982

Inverse Problems = Quest for Information

A. Tarantola and B. Valette

Institut de Physique du Globe, L.E.G.S.P. (C.N.R.S.), Université Pierre et Marie Curie, 4 place Jussieu, F-75230 Paris Cedex 05, France

Abstract. We examine the general non-linear inverse problem with a finite number of parameters. In order to permit the incorporation of any a priori information about parameters and any distribution of data (not only of gaussian type) we propose to formulate the problem not using single quantities (such as bounds, means, etc.) but using probability density functions for data and parameters. We also want our formulation to allow for the incorporation of theoretical errors, i.e. non-exact theoretical relationships between data and parameters (due to discretization, or incomplete theoretical knowledge); to do that in a natural way we propose to define general theoretical relationships also as probability density functions. We show then that the inverse problem may be formulated as a problem of *combination of information*: the experimental information about data, the a priori information about parameters, and the theoretical information. With this approach, the general solution of the non-linear inverse problem is unique and consistent (solving the same problem, with the same data, but with a different system of parameters does not change the solution).

Key words: Information — Inverse problems — Pattern recognition — Probability

1. Introduction

Inverse problem theory was essentially developed in geophysics, to deal with largely underdetermined problems. The most important approaches to the solution of this kind of problem are well known to today's geophysicists (Backus and Gilbert 1967, 1968, 1970; Keilis-Borok and Yanovskaya 1967; Franklin 1970; Backus 1971, Jackson 1972; Wiggins 1972; Parker 1975; Rietsch 1977; Sabatier 1977).

The minimal constraints necessary for the formulation of an inverse problem are:

1. The formulation must be valid for linear as well as for strongly non linear problems.
2. The formulation must be valid for overdetermined as well as for underdetermined problems.
3. The formulation of the problem must be consistent with respect to a change of variables. (This is not the case with ordinary approaches: solving an inverse problem with a given parameter, e.g. a velocity v , leads to a solution v_0 ; solving the same problem with the same data but with another parameter, e.g. the slowness $n=1/v$, leads to a solution n_0 .

There is no natural relation between v_0 and n_0 in ordinary approaches).

4. The formulation must be general enough to allow for general error distributions in the data (which may be not gaussian, asymmetric, multimodal, etc.).

5. The formulation must be general enough to allow for the formal incorporation of any a priori assumption (positivity constraints, smoothness, etc.).

6. The formulation must be general enough to incorporate theoretical errors in a natural way. As an example, in seismology, the theoretical error made by solving the forward travel time problem is often one order of magnitude *larger* than the experimental error of reading the arrival time on a seismogram. A coherent hypocenter computation must take into account experimental as well as theoretical errors. Theoretical errors may be due, for example, to the existence of some random parameters in the theory, or to theoretical simplifications, or to a wrong parameterization of the problem.

None of the approaches by previous authors satisfies this set of constraints. The main task of this paper is to demonstrate that all these constraints may be fulfilled when formulating the inverse problem using a simple extension of probability theory and information theory.

To do this we will limit ourselves to the study of systems which can be described with a *finite set of parameters*. This limitation is twofold: first, we will only be able to handle *quantitative* characteristics of systems. All qualitative aspects are beyond the scope of this paper. The second limitation is that to describe some of the characteristics of the systems we should employ functions rather than discrete parameters, as for example for the output of a continuously recording seismograph, or the velocity of seismic waves as a function of depth. In such cases we decide to sample the corresponding function.

The problem of adequate sampling is not a trivial one. For example, if the sampling interval of a seismogram is greater than the correlation length of the noise (seismic noise, finite pass-band filter, etc.), errors in data may be assumed to be independent; this will not be true when densifying the sampling. We explicitly assume in this paper that the discretization has been made carefully enough, so that densifying the sampling will only negligibly alter the results.

In the next section we will define precisely concepts such as parameter, probability density function, information, and combination of information; in Sect. 3 we discuss the concept of null information; in Sect. 4 we define the a priori infor-

mation on a system; in Sect. 5 we define general theoretical relationships between data and parameters; finally in Sect. 6 we give the solution of the inverse problem. In Sects. 7-10 we discuss this solution, we examine particular cases, and give a seismological illustration with actual data.

2. Parameters and Information

Let \mathcal{S} be a *physical system* in a wide sense. By wide sense we mean that \mathcal{S} consists of a physical system *strictu sensu*, plus a family of measuring instruments and their outputs. We assume that \mathcal{S} is a discrete system or that it has been discretized (for convenience of description or because the mathematical model which describes the physical system is discrete). In that case \mathcal{S} can be described using a finite (perhaps large) set of parameters $\mathbf{X} = \{X_1, \dots, X_m\}$; any set of specific values of this set of parameters will be denoted $\mathbf{x} = \{x_1, \dots, x_m\}$. Each point \mathbf{x} may be named a *model* of \mathcal{S} . The m -dimensional space \mathcal{E}^m where the parameters \mathbf{X} take their values may be named the *model space*, or the *parameter space*.

When a physical system \mathcal{S} can be described by a set \mathbf{X} of parameters, we say that \mathcal{S} is *parametrizable*.

It should be noted that the parameterization of a system is not unique. We say that two parameterizations are *equivalent* if they are related by a bijection. Let

$$\mathbf{X} = \mathbf{X}(\mathbf{X}') \quad \mathbf{X}' = \mathbf{X}'(\mathbf{X}) \quad (2-1)$$

be two equivalent parameterizations of \mathcal{S} . We emphasize that Eqs. (2-1) represent a transformation between mathematically equivalent parameters and that they *do not* represent any relationship between physically correlated parameters. An example of equivalent parameters is a velocity v and the corresponding slowness *defined* by $n=1/v$. Let us remark that two equivalent parameterizations of \mathcal{S} can also be seen as two different choices of *system of coordinates* in \mathcal{E}^m .

The degree of knowledge that we have about the values of the parameters of our system may range from total knowledge to total ignorance. A first postulate of this paper is that *any* state of knowledge on the values of \mathbf{X} can be described using a *measure density function* $f(\mathbf{x})$; i.e. a real, positive, locally Lebesgue integrable function such that the positive measure defined by:

$$m(A) = \int_A f(\mathbf{x}) d\mathbf{x} \quad (A \subset \mathcal{E}^m) \quad (2-2)$$

is absolutely continuous with respect to the Lebesgue measure defined over \mathcal{E}^m . The quantity $m(A)$ is named the *measure* of A . If $m(\mathcal{E}^m)$ is finite then $f(\mathbf{x})$ can be normalized in such a way that $m(\mathcal{E}^m) = 1$; in that case $f(\mathbf{x})$ is named a *probability density function*, $m(A)$ is then noted $p(A)$ and is named the *probability* of A .

All through this paper a probability density function will be noted *p.d.f.* A measure density function, non normalized or non normalizable, will simply be named a density function and noted *d.f.*

Of course, the form of $f(\mathbf{x})$ depends on the chosen parameterization. Let \mathbf{X} and \mathbf{X}' be two equivalent parameterizations. As we want the measure $m(A)$ to be invariant, it is easy to see that there exists a d.f. $f'(\mathbf{x}')$, which is related to $f(\mathbf{x})$ by the usual formula:

$$f'(\mathbf{x}') = f(\mathbf{x}) \cdot \left| \frac{\partial \mathbf{x}}{\partial \mathbf{x}'} \right| \quad (2-3)$$

where the symbol $\left| \frac{\partial \mathbf{x}}{\partial \mathbf{x}'} \right|$ stands for the Jacobian of the transformation. (It never vanishes for equivalent parameterizations).

Let us define a particular d.f. $\mu(\mathbf{x})$ representing the state of total ignorance (Jaynes 1968; Rietsch 1977). Often the state of total ignorance will correspond to a uniform function $\mu(\mathbf{x}) = \text{const.}$, sometimes it will not, as discussed in Sect. 3. We will assume:

$$\mu(\mathbf{x}) \neq 0 \quad (2-4)$$

everywhere in \mathcal{E}^m (In fact this means that we restrict the space of parameters to the region not excluded by the state of total ignorance).

We should need a d.f., rather than a p.d.f. when we are not able to define the *absolute* probability of a subset A , but we can define the *relative* probabilities of two subsets A and B . The most trivial example is when $f(\mathbf{x}) = \text{const.}$ and the space is not bounded.

Two d.f. which differ only by a multiplicative constant will give the same *relative* probabilities, and all through this paper they will be considered identical:

$$f(\mathbf{x}) \equiv \text{const.} f(\mathbf{x}) \quad (2-5)$$

If the state of total ignorance corresponds to a p.d.f. $\mu(\mathbf{x})$, then the content of *information* of any p.d.f. $f(\mathbf{x})$ is defined by (Shannon 1948)

$$I(f; \mu) = \int f(\mathbf{x}) \text{Log} \frac{f(\mathbf{x})}{\mu(\mathbf{x})} d\mathbf{x}. \quad (2-6)$$

This definition has the following properties, which are easily verified:

a) I is invariant with respect to a change of variables:

$$I(f; \mu) = I(f'; \mu'). \quad (2-7)$$

(In Shannon's original definition of information for continuous variables the term $\mu(\mathbf{x})$ in Eq. (6) is missing, so that Shannon's definition is not invariant.)

b) Information cannot be negative:

$$I(f; \mu) \geq 0 \quad (2-8)$$

c) the information of the state of total ignorance is null:

$$I(\mu; \mu) = 0 \quad (2-9)$$

the reciprocal being also true:

$$I(f; \mu) = 0 \Rightarrow f = \mu. \quad (2-10)$$

We will say that each p.d.f. (or, by extension, each d.f.) $f_i(\mathbf{x})$ represents a *state of information*, which will be noted s_i .

Let us now set up a problem which appears very often under different aspects. Its general formulation may be: Let \mathbf{X} be a set of parameters describing some system \mathcal{S} . Let $\mu(\mathbf{x})$ be a d.f. representing the state of null information on the system. If we receive two pieces of information on our system, represented by the density functions $f_i(\mathbf{x})$ and $f_j(\mathbf{x})$, how do we combine f_i and f_j to obtain a d.f. $f(\mathbf{x})$ representing the final state of information?

We must first state *which* kind of combination we wish. To do this, let us first recall the way used in classical logic to define the combination of *logical propositions*. If p_i is a logical

proposition, one defines its *value of truth* $v(p_i)$ by taking the values 1 or 0 when p_i is respectively certain or impossible (true or false). Let p_i and p_j be two logical propositions. It is usual to combine them in order to obtain new propositions, as for example by defining the *conjunction* of two propositions, $p_i \wedge p_j$ (p_i and p_j) or by defining the *disjunction* of two propositions $p_i \vee p_j$ (p_i or p_j), and so on. The usual way for defining the result of these combinations is by establishing their values of truth. For example, the conjunction $p_i \wedge p_j$ is defined by:

$$\left. \begin{array}{l} v(p_i)=0 \\ \text{or} \\ v(p_j)=0 \end{array} \right\} \Leftrightarrow v(p_i \wedge p_j)=0. \quad (2-11)$$

For our purposes, we need the *definition of the conjunction of two states of information*, $s_i \wedge s_j$. This definition must be the generalization to the concept of states of information of the properties of the conjunction of logical propositions.

We will see later that this definition will allow the solution of many seemingly different problems, in particular it contains the solution of the general inverse problem as it has been stated in the preceding section.

Let us note:

$$f = f_i \wedge f_j \quad (2-12)$$

the operation which combines f_i and f_j (representing two states of information s_i and s_j) to obtain f (representing the conjunction $s = s_i \wedge s_j$). The definition must satisfy the following conditions:

a) $f_i \wedge f_j$ must be a d.f. In particular, the content of information of $f_i \wedge f_j$ must be invariant with respect to a change of parameters, i.e. Eq. (2-3) must be verified.

b) The operation must be commutative, i.e., for any f_i and f_j :

$$f_i \wedge f_j = f_j \wedge f_i. \quad (2-13)$$

c) The operation must be associative; i.e. for any f_i, f_j and f_k :

$$f_i \wedge (f_j \wedge f_k) = (f_i \wedge f_j) \wedge f_k \quad (2-14)$$

d) the conjunction of any state of information f_i with the null information μ must give f_i , i.e. must not result in any loss of information:

$$f_i \wedge \mu = f_i. \quad (2-15)$$

This equation means that μ is the neutral element for the operation.

e) The final condition corresponds to an extension of the defining property of the conjunction of logical propositions (Eq. (2-11)). For any measurable $A \subset \mathcal{E}^m$:

$$\left. \begin{array}{l} \int_A f_i d\mathbf{x} = 0 \\ \text{or} \\ \int_A f_j d\mathbf{x} = 0 \end{array} \right\} \Leftrightarrow \int_A (f_i \wedge f_j) d\mathbf{x} = 0 \quad (2-16)$$

which means that a necessary and sufficient condition for $f_i \wedge f_j$ to give a null probability to a subset A is that either f_i or f_j give a null probability for A .

This last condition implies that the measure engendered

by $f_i \wedge f_j$ is absolutely continuous with respect to the measures engendered by f_i and f_j respectively. Using Nikodym's theorem it can be shown (Descombes 1972) that $f_i \wedge f_j$ may then necessarily be written in the form:

$$f_i \wedge f_j = f_i \cdot f_j \cdot \Phi(f_i, f_j) \quad (2-17)$$

where $\Phi(f_i, f_j)$ is a locally Lebesgue integrable, positive function. This last condition is strong, because it is valid everywhere in \mathcal{E}^m , in particular where f_i or f_j are null (otherwise this equation would be trivial).

The simplest choice for $\Phi(f_i, f_j)$ in order to satisfy conditions b) and c) is to take it as independent of f_i or f_j . Condition d) then imposes:

$$\Phi(f_i, f_j) = \frac{1}{\mu}. \quad (2-18)$$

Condition a) is then automatically verified.

The above discussion suggests then the following definition:

Let s_i and s_j be two states of information represented respectively by the d.f.'s f_i and f_j , let μ be a d.f. representing the state of null information. By definition, the *conjunction* of s_i and s_j , denoted $s = s_i \wedge s_j$ is a state of information represented by the d.f. $f(\mathbf{x})$ given by:

$$f(\mathbf{x}) = \frac{f_i(\mathbf{x}) \cdot f_j(\mathbf{x})}{\mu(\mathbf{x})}. \quad (2-19)$$

The d.f. $f(\mathbf{x})$ is not necessarily normalizable, but except in some ad hoc examples, in most actual problems when one (or both) of the d.f. $f_i(\mathbf{x})$ or $f_j(\mathbf{x})$ is normalizable, the d.f. $f(\mathbf{x})$ is also normalizable, i.e. it is, in fact, a p.d.f.

In the following sections we will show that the Definition (2-19) allows for a simple solution of general inverse problems. In the appendix we recall the definition of *marginal* p.d.f.'s, we show that the *conditional* p.d.f. can be defined as a particular case of conjunction of states of information, and demonstrate the Bayes theorem (which is not used in this work because it is too restrictive for our purposes).

Let us emphasize that, given two states of information s_i and s_j on a system \mathcal{S} , the resulting state of information does not necessarily correspond to the conjunction $s_i \wedge s_j$. The conjunction, as defined above, must be used to combine two states of information only if these states of information have been obtained *independently*, as for example, for two independent physical measurements on a given set of parameters, or for combining experimental and theoretical information (see Sect. 6).

Let us conclude this section by the remark that in our use of probability calculus we do not use concepts such as *random variable*, *realization* of a random variable, *true value* of a parameter, and so on. Our d.f.'s are interpreted in terms of *human knowledge*, rather than in terms of *statistical properties* of the Earth. Of course, we accept statistics, and we use the language of statistics when statistical computations are possible, but this is generally not the case in geophysical experiments.

3. The Null Information on a System

As the concept of null information is not straightforward, let us discuss it in some detail and start with some examples. Assume that our problem consists in the location of an

earthquake focus from some set of data. Assume also that we are using cartesian coordinates (X, Y, Z) . The question is: which will be the d.f. $\mu(x, y, z)$ which is *least* informative on the location of the focus? The intuitive answer is that the least informative d.f. will be the one that assigns the same probability dP to all regions of equal volume dV :

$$dP = \text{const.} \cdot dV. \quad (3-1)$$

Since in cartesian coordinates $dV = dx \cdot dy \cdot dz$, Eq. (2-2) gives the solution:

$$\mu(x, y, z) = \text{const.} \quad (3-2)$$

If instead of cartesian coordinates we use spherical coordinates (R, θ, ϕ) , the null information d.f. $\mu'(r, \theta, \phi)$ may be obtained from Eq. (3-1) writing the elementary volume dV in spherical coordinates, $dV = r^2 \cdot \sin \theta \cdot dr \cdot d\theta \cdot d\phi$ or from Eq. (3-2) by means of a change of variables. We arrive at:

$$\mu(r, \theta, \phi) = \text{const.} \cdot r^2 \cdot \sin \theta \quad (3-3)$$

which is far from a constant function. We see in this example that the d.f. representing the null information need not be constant.

We will now try to solve a less trivial question. Let $V = |V|$ be some velocity. Could the null information d.f. $\mu(v) = \text{const.}$? To those who are tempted to answer *yes*, we ask another question. Let N be some slowness ($N = 1/V$). Could the null information d.f. $\mu'(n) = \text{const.}$? Obviously, if $\mu(v)$ is constant, $\mu'(n)$ cannot be, and vice-versa.

To properly define the null information density function μ , we will follow Jaynes (1968), who suggested that suitable density functions are obtained under the condition of *invariance of form* of the function μ under the transformgroups which leave invariant the equations of physics (see also Rietsch, 1977). Clearly, the form of μ must be invariant under a change of space-time origin and under a change of space-time scale. To see its consequences let \mathbf{O} and $\hat{\mathbf{O}}$ be two observers and let (X, Y, Z, T) and $(\hat{X}, \hat{Y}, \hat{Z}, \hat{T})$ be their coordinate system. The fact that observer $\hat{\mathbf{O}}$ has chosen a different space-time origin and scale is easily written in cartesian coordinates:

$$\begin{aligned} \hat{X} &= X_0 + a \cdot X \\ \hat{Y} &= Y_0 + a \cdot Y & \hat{T} &= T_0 + b \cdot T \\ \hat{Z} &= Z_0 + a \cdot Z \end{aligned} \quad (3-4)$$

where a and b are constants. Thus, by the definition of velocity:

$$\hat{v} = \frac{|d\hat{r}|}{d\hat{t}} = \frac{a}{b} \frac{|dr|}{dt} = c \cdot v \quad (3-5)$$

where $c = a/b$ is a new constant. Let $\mu(v)$ be the null information d.f. for \mathbf{O} and $\hat{\mu}(\hat{v})$ be the one of $\hat{\mathbf{O}}$. From Eq. (3-5) and Eq. (2-4) we must have:

$$\mu(v) = \hat{\mu}(\hat{v}) \cdot \left| \frac{d\hat{v}}{dv} \right| = c \cdot \hat{\mu}(c \cdot v). \quad (3-6)$$

The invariance under transformations (3-4) will be realized if μ and $\hat{\mu}$ are *the same function*, that is:

$$\mu(w) = \hat{\mu}(w) \quad (3-7)$$

for all w .

From Eqs. (3-6) and (3-7) it follows:

$$\mu(v) = c \cdot \mu(c \cdot v) \quad (3-8)$$

i.e.

$$\mu(v) = \frac{\text{const.}}{v}. \quad (3-9)$$

This result may appear puzzling to some. Let us ask which is the form of the null information d.f. for the slowness $N = 1/V$. We readily find:

$$\mu'(n) = \mu(v) \cdot \left| \frac{dv}{dn} \right| = \frac{\text{const.}}{n}. \quad (3-10)$$

We see that the equivalent parameters V and N have null information d.f. of exactly the same form. In fact, it was in order to warrant this type of symmetry between all the powers of a parameter that Jeffreys (1939, 1957) suggested assigning to all continuous parameters X *known to be positive* a d.f., representing the null information, of the form $\text{const.}/x$.

Some formalisms of inverse problems attempt a definition of some probabilistic properties in parameters space (computation of standard deviations, etc.). We claim that these kind of problems cannot be consistently posed without explicitly stating the null information d.f. μ .

In most ordinary cases the choice

$$\mu = \text{const.} \quad (3-11)$$

will give reasonable results. Nevertheless we must emphasize that the solution of the same problem using a different set of parameters will be inconsistent with the choice of Eq. (3-11) for representing the state of null information in the new set of parameters, unless the change of parameters is linear.

4. Data and A Priori Information

Among the set of parameters \mathbf{X} describing a system \mathcal{S} , the parameters describing the outputs of the measuring instrument are named *data* and written $\mathbf{D} = (D_1, \dots, D_n)$. The rest of the parameters are then named parameters *strictu sensu*, or, briefly, parameters, and are written $\mathbf{P} = (P_1, \dots, P_s)$. If a partition of \mathbf{X} into $\mathbf{X} = (\mathbf{D}, \mathbf{P})$ is made, then any d.f. on \mathbf{X} may be equivalently written:

$$f(\mathbf{x}) = f(\mathbf{d}, \mathbf{p}). \quad (4-1)$$

Let us consider a particular geophysical measurement, for example, the act of reading the arrival time of a particular phase on a seismogram. In the simplest case the seismologist puts all the information he has obtained from his measurement in the form of a given value, say t , and an "uncertainty", say σ_t . In more difficult cases, he may hesitate between two or more values. What he may do, more generally, is to define, for each time interval Δt on the seismogram, the probability ΔP which he assigns to the arrival time t to be in the interval Δt . Doing this, he is putting the information which he obtains from his measurement into the form of a probability density function $\rho(t) = \Delta P / \Delta t$. This p.d.f. can be asymmetric, multimodal, etc. Extracting from this p.d.f. a few estimators, such as mean or variance, would certainly lead to a loss of information, thus we have to take as an elementary datum the p.d.f. $\rho(t)$ itself.

Let us now consider a non-directly measurable parameter P_α . Some examples of a priori information are: a) We know only that P_α is bounded by two values $a \leq p_\alpha \leq b$. We will obviously represent this a priori information by a d.f. which is null outside the interval and which coincides with the null information d.f. inside the interval. b) Inequality constraint $P_\alpha \leq P_\beta$: we take a d.f. null for $p_\alpha > p_\beta$ and equal to the null information d.f. for $p_\alpha \leq p_\beta$. c) Some parameters $P_{\alpha+1}, P_{\alpha+2}, \dots, P_\beta$ are spatially (or temporally) distributed, and we know that their variation is smooth. Accordingly, we will represent this a priori information by using a joint d.f. $\rho(p_{\alpha+1}, p_{\alpha+2}, \dots, p_\beta)$ with the corresponding non-null assumed correlations (covariances). d) We have some diffuse a priori information about some parameters. In that case we will define a priori d.f. with weak limits and large variances. e) We have no a priori information at all. This a priori information is then represented by the null information function μ .

We see then that we may assume the existence of a density function:

$$\rho(\mathbf{x}) = \rho(\mathbf{d}, \mathbf{p}) \quad (4-2)$$

named the a priori d.f., representing both, the results of measurements and all a priori information on parameters.

5. Theoretical Relationships

A theoretical relationship is usually viewed as a functional relation between the values of the parameters:

$$\mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{d}, \mathbf{p}) = \mathbf{0}. \quad (5-1)$$

Often the form (5-1) of a functional relationship may be simplified and may be written (Fig. 1):

$$\mathbf{d} = \mathbf{G}(\mathbf{p}). \quad (5-2)$$

This view is too restrictive. In most cases, even if the value \mathbf{p} is given we are not able to *exactly* compute the corresponding value of \mathbf{d} , because our theory is incomplete, or because the theory contains some random parameters, or because we have roughly parametrized the system under study. In such cases, to be rigorous, we may exhibit not the value $\mathbf{d} = \mathbf{G}(\mathbf{p})$ but the probability density function for \mathbf{d} , given \mathbf{p} , i.e. the conditional p.d.f. (Fig. 2):

$$\theta(\mathbf{d}|\mathbf{p}). \quad (5-3)$$

We will see in Sect. 10 how to display such a conditional p.d.f. for actual problems.

In all generality, we will assume that any theoretical relationship may be represented by a joint density function:

$$\theta(\mathbf{x}) = \theta(\mathbf{d}, \mathbf{p}). \quad (5-4)$$

From the definition of conditional probability (see Appendix):

$$\theta(\mathbf{d}, \mathbf{p}) = \theta(\mathbf{d}|\mathbf{p}) \cdot \theta_p(\mathbf{p}) \quad (5-5)$$

where $\theta_p(\mathbf{p})$ is the marginal d.f. for \mathbf{P} . In the class of problems where the simplification (5-2) is used, the theory does not impose any constraint on \mathbf{P} but only in \mathbf{D} . Equation (5-5) may then be rewritten:

$$\theta(\mathbf{d}, \mathbf{p}) = \theta(\mathbf{d}|\mathbf{p}) \cdot \mu_p(\mathbf{p}) \quad (5-6)$$

where $\mu_p(\mathbf{p})$ is the null information d.f.

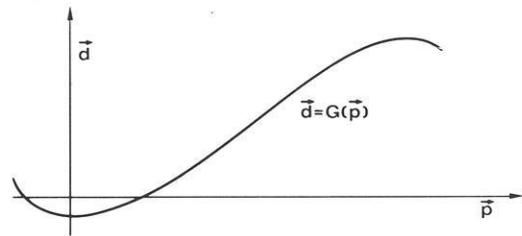


Fig. 1. An exact theory, viewed as a functional relationship

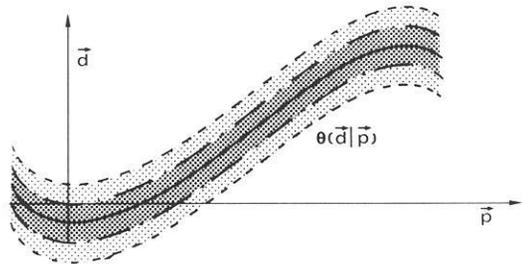


Fig. 2. Putting "error-bars" on the theoretical relation $\mathbf{d} = \mathbf{G}(\mathbf{p})$

The particular case of an exact theory (Eq. (5-2)) obviously corresponds to $\theta(\mathbf{d}|\mathbf{p}) = \delta(\mathbf{d} - \mathbf{G}(\mathbf{p}))$ where δ is the Dirac distribution. So, for an exact theory:

$$\theta(\mathbf{d}, \mathbf{p}) = \delta(\mathbf{d} - \mathbf{G}(\mathbf{p})) \cdot \mu_p(\mathbf{p}). \quad (5-7)$$

In cases where a rigorous computation of $\theta(\mathbf{d}|\mathbf{p})$ cannot be made, but where we have an idea of the theoretical "error-bar" σ_T , choices of $\theta(\mathbf{d}, \mathbf{p})$ of a form similar to

$$\theta(\mathbf{d}|\mathbf{p}) = \text{const.} \exp \left\{ -\frac{1}{2} \frac{\|\mathbf{d} - \mathbf{G}(\mathbf{p})\|^2}{\sigma_T^2} \right\} \quad (5-8)$$

may be good enough to take into account this theoretical error.

In any case, we assume that theoretical relationships are in general represented by the joint density function of Eq. (5-4) which will be named the *theoretical density function*.

6. Statement and Solution of Inverse Problems

Let \mathcal{S} be a physical system, and let \mathbf{X} be a parametrization of \mathcal{S} . In Sect. 3 we have defined the density function $\mu(\mathbf{x})$ representing the state of null information on the system. In Sect. 4 we have defined the density function $\rho(\mathbf{x})$ representing all a priori information on the system, in particular the results of the measurements and a priori constraints on parameters. In Sect. 5 we have defined the density function $\theta(\mathbf{x})$ representing the theoretical relationships between parameters.

The conjunction of $\rho(\mathbf{x})$ and $\theta(\mathbf{x})$ gives a new state of information, which will be named the *a posteriori state of information*. The corresponding d.f. will be denoted $\sigma(\mathbf{x})$ and is given, using Eq. (2-19), by

$$\sigma(\mathbf{x}) = \frac{\rho(\mathbf{x}) \cdot \theta(\mathbf{x})}{\mu(\mathbf{x})}. \quad (6-1)$$

To examine inverse problems we separate our set of parameters \mathbf{X} into the subsets $\mathbf{X} = (\mathbf{D}, \mathbf{P})$ representing data and parameters strictu sensu respectively. Equation (6-1) may then

be rewritten:

$$\sigma(\mathbf{d}, \mathbf{p}) = \frac{\rho(\mathbf{d}, \mathbf{p}) \cdot \theta(\mathbf{d}, \mathbf{p})}{\mu(\mathbf{d}, \mathbf{p})}. \quad (6-2)$$

From this equation we may compute the a posteriori marginal density functions:

$$\sigma_{\mathbf{d}}(\mathbf{d}) = \int \frac{\rho(\mathbf{d}, \mathbf{p}) \cdot \theta(\mathbf{d}, \mathbf{p})}{\mu(\mathbf{d}, \mathbf{p})} d\mathbf{p}, \quad (6-3)$$

$$\sigma_{\mathbf{p}}(\mathbf{p}) = \int \frac{\rho(\mathbf{d}, \mathbf{p}) \cdot \theta(\mathbf{d}, \mathbf{p})}{\mu(\mathbf{d}, \mathbf{p})} d\mathbf{d}. \quad (6-4)$$

Equation (6-4) performs the task of transferring to parameters, via theoretical correlations, the information contained in the data set. This is, by definition, the solution to an *inverse problem* (see Fig. 3).

Equation (6-3) solves what could be named a generalized forward problem.

In most cases the a priori information on \mathbf{D} is independent from the a priori information on \mathbf{P} :

$$\rho(\mathbf{d}, \mathbf{p}) = \rho_{\mathbf{d}}(\mathbf{d}) \cdot \rho_{\mathbf{p}}(\mathbf{p}) \quad (6-5)$$

and the theoretical density function is obtained in the form of a conditional density function (Eq. (5-6)):

$$\theta(\mathbf{d}, \mathbf{p}) = \theta(\mathbf{d}|\mathbf{p}) \cdot \mu_{\mathbf{p}}(\mathbf{p}). \quad (6-6)$$

Equation (6-4) may then be simplified to:

$$\sigma_{\mathbf{p}}(\mathbf{p}) = \rho_{\mathbf{p}}(\mathbf{p}) \cdot \int \frac{\rho_{\mathbf{d}}(\mathbf{d}) \cdot \theta(\mathbf{d}|\mathbf{p})}{\mu_{\mathbf{d}}(\mathbf{d})} d\mathbf{d}. \quad (6-7)$$

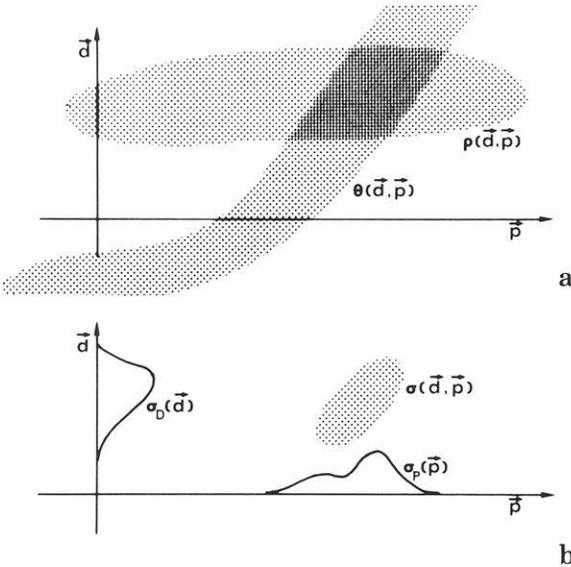


Fig. 3. **a** The theoretical model is often non exact (simplified, rough parameterization, etc.). We can then introduce the theoretical relationship between parameters as a density function $\theta(\mathbf{d}, \mathbf{p})$ (see Sect. 5). **b** The solution of the problem is then defined by $\sigma(\mathbf{x}) = \frac{\rho(\mathbf{x}) \cdot \theta(\mathbf{x})}{\mu(\mathbf{x})}$. If the a priori density function contains small variances for data and great variances for parameters, the marginal density function $\sigma_{\mathbf{p}}(\mathbf{p})$ solves an "inverse problem". On the contrary, if in $\rho(\mathbf{x})$ the data have large variances and the parameters have small variances, $\sigma_{\mathbf{d}}(\mathbf{d})$ solves the "forward problem"

If, furthermore, the theoretical relationship may be considered as exact (i.e. we can write $\mathbf{d} = \mathbf{G}(\mathbf{p})$), then using Eq. (5-7).

$$\theta(\mathbf{d}|\mathbf{p}) = \delta(\mathbf{d} - \mathbf{G}(\mathbf{p})). \quad (6-8)$$

Equation (6-7) may be easily integrated to:

$$\sigma_{\mathbf{p}}(\mathbf{p}) = \rho_{\mathbf{p}}(\mathbf{p}) \cdot \frac{\rho_{\mathbf{d}}(\mathbf{G}(\mathbf{p}))}{\mu_{\mathbf{d}}(\mathbf{G}(\mathbf{p}))}. \quad (6-9)$$

This last equation solves the inverse problem for an exact, non-linear theory with arbitrary a priori constraints on parameters ($\rho_{\mathbf{p}}$), and an arbitrary probabilistic distribution of data ($\rho_{\mathbf{d}}$).

Returning to the general solution (6-4) let us answer the question of displaying the information contained in $\sigma_{\mathbf{p}}(\mathbf{p})$. If we are interested in a particular parameter, say P_1 , all the information on P_1 is contained in the marginal d.f. (Eq. (2-5)):

$$\sigma_1(p_1) = \int \sigma_{\mathbf{p}}(\mathbf{p}) \cdot dp_2 \cdot dp_3 \cdot \dots \cdot dp_s \quad (6-10)$$

As far as we are interested in the parameter P_1 and not in the correlations between this and other parameters, $\sigma_1(p_1)$ exhibits *all* the available information about P_1 . For example, from $\sigma_1(p_1)$ we can precisely answer questions such as the probability that P_1 lies between two values. Alternatively, from $\sigma_1(p_1)$ it is possible to extract the mean value, the median value, the maximum-likelihood value, the standard deviation, the mean deviation, or any estimator we need.

Let us remark that it is possible to compute from the general solution $\sigma_{\mathbf{p}}(\mathbf{p})$ the *a posteriori* mathematical expectation:

$$E(\mathbf{P}) = \int \mathbf{p} \cdot \sigma_{\mathbf{p}}(\mathbf{p}) \cdot d\mathbf{p} \quad (6-11)$$

or the a posteriori covariance matrix:

$$\begin{aligned} \mathbf{C} &= E\{(\mathbf{P} - E(\mathbf{P})) \cdot (\mathbf{P} - E(\mathbf{P}))^T\} \\ &= \int (\mathbf{p} - E(\mathbf{P})) \cdot (\mathbf{p} - E(\mathbf{P}))^T \sigma_{\mathbf{p}}(\mathbf{p}) d\mathbf{p} \\ &= \int \mathbf{p} \cdot \mathbf{p}^T \sigma_{\mathbf{p}}(\mathbf{p}) d\mathbf{p} - E(\mathbf{P}) \cdot E(\mathbf{P})^T. \end{aligned} \quad (6-12)$$

Estimators such as $E(\mathbf{P})$ and \mathbf{C} are similar to what is obtained in traditional approaches to inverse problems, but here they can be obtained without any linear approximation.

7. Existence, Uniqueness, Consistency, Robustness, Resolution

In inverse problem theory, it is not always possible to prove the existence or the uniqueness of the solution. With our approach, the existence of the solution is merely the existence of the a posteriori density function $\sigma(\mathbf{x})$ of Eq. (6-1) and results trivially from the assumption of the existence of $\rho(\mathbf{x})$ and $\theta(\mathbf{x})$. Of course we can obtain a solution $\sigma(\mathbf{x})$ with some pathologies (non-normalizable, infinite variance, not unique maximum likelihood point, etc.), but the solution is *the* a posteriori density function, with all the pathologies it may present.

Consistency is warranted because Eq. (6-1) is consistent, i.e. the function $\sigma(\mathbf{x})$ is a density function. Let us suppose again that we use velocity as a parameter and obtain as solution the d.f. $\sigma(v)$, then using the same data in the same problem (with the same discretization) but using slowness as parameter we obtain as a solution the d.f. $\sigma'(n)$. Since Eq. (6-1) is consistent, $\sigma'(n)$ will be related to $\sigma(v)$ by the usual

formula (2-3) $\sigma'(n) = \sigma(v) \cdot \left| \frac{dv}{dn} \right|$: $\sigma'(n)$ and $\sigma(v)$ represent exactly the same state of information.

In those approaches where all the information contained in the data is condensed into the form of central estimators (mean, median, etc.), the notion of robustness must be carefully examined if we suspect that there may be blunders in the data set (Clearbout and Muir, 1973). In our approach, the suspicion of the presence of blunders in a data set may be introduced using long-tailed density functions in $\rho_{\mathbf{a}}(\mathbf{d})$, decreasing much more slowly than gaussian functions as, for example, exponential functions. Our experience shows that with such long-tailed functions, the solution $\sigma_{\mathbf{p}}(\mathbf{p})$ is rather insensitive to one blunder.

The concept of resolution must be considered under two different aspects: to what extent a given parameter has been "resolved" by the data? and what is the "spatial resolution" attained with our data set?

For the first aspect, let us consider a parameter P_i whose value does not influence the values of the data. This means that the theoretical density function $\theta(\mathbf{d}, \mathbf{p})$ does not depend on P_i . Even in this case we can obtain a certain amount of information on this parameter, if the other parameters are resolved by the data, and if the a priori density function $\rho_{\mathbf{p}}(\mathbf{p})$ introduces some correlation between parameters. Furthermore, let us assume that no correlation is introduced by $\rho_{\mathbf{p}}(\mathbf{p})$ between P_i and the other parameters. This is the worst case of non-resolution we can imagine for a parameter. Under these assumptions Eq. (6-7) can be written:

$$\sigma_{\mathbf{p}}(\mathbf{p}) = \rho_i(p_i) \cdot \rho_{\mathbf{q}}(\mathbf{q}) \cdot \int \rho(\mathbf{d}) \cdot \frac{\theta(\mathbf{d}|\mathbf{q})}{\mu_{\mathbf{a}}(\mathbf{d})} d\mathbf{d} \quad (7-1)$$

where \mathbf{q} is the vector $(p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_s)$. After integration over the set of \mathbf{q} , we find (dropping the multiplicative constant):

$$\sigma_i(p_i) = \rho_i(p_i) \quad (7-2)$$

which means that for a completely unresolved parameter, the a posteriori marginal density function equals the a priori one. The more $\sigma_i(p_i)$ differs from $\rho_i(p_i)$, the more the parameter P_i has been resolved by the data set.

The concept of spatial resolution applies to a different problem: Assume that P_1, \dots, P_j form a set of parameters spatially (or temporally) distributed as for example when the parameters represent the seismic velocities of successive geologic layers (or values from the sampling of some continuous geophysical record). Assume that we are not interested in obtaining the a posteriori d.f. for each parameter, but only the a posteriori mean values, as given by Eq. (6-11). There are two reasons for $E(\mathbf{P})$ to be a smooth vector (i.e. to have small variations between consecutive values $E(P_i)$ and $E(P_{i+1})$). The first reason may be the type of data used; it is well known for instance that long period surface wave data only give a smoothed vision of the Earth. The second reason for obtaining a smoothed solution may simply be that we decide a priori to impose such smoothness introducing non null a priori covariances in $\rho_{\mathbf{p}}(\mathbf{p})$ (see Sect. 4).

This question of spatial resolution has been clearly pointed out, and extensively studied by Backus and Gilbert (1970). From our point of view, this problem must be solved by studying the a posteriori correlations between parameters. From Eq. (6-12):

$$C_{ij} = \int p_i \cdot p_j \cdot \sigma(p_i, p_j) \cdot dp_i \cdot dp_j - E(p_i) \cdot E(p_j). \quad (7-3)$$

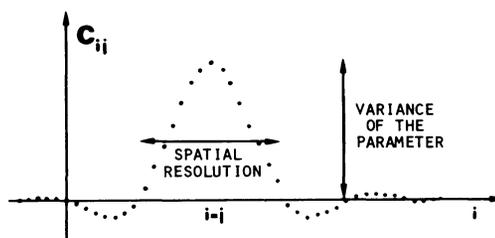


Fig. 4. Rows (or columns) of the a posteriori covariance matrix showing both the length of spatial resolution and the a posteriori variance of the parameter

If, for a given j , we plot the "curve" C_{ij} versus i we simultaneously obtain information on the a posteriori variance of the parameter (C_{ii}) and the spatial resolution (the length of correlation) (see Fig. 4).

In applications of Backus and Gilbert's point of view on inverse problems it is usual to study the *trade-off* between variance and resolution in order to choose the desired solution. In our approach, such a trade-off also exists: modifying the a priori variances or a priori correlations of parameters (in $\rho(\mathbf{p})$) results in a change of the a posteriori variances and resolutions. But, in our opinion, the a priori information on the values of parameters, as contained in $\rho(\mathbf{p})$, must *not* be stated in order to obtain a pleasant solution, but in order to closely correspond to the *actual* a priori information.

8. Computational Aspects

For linear problems, all the integrations of Sect. 6 can often be performed analytically, and the most general solution can sometimes be reached easily (see for example the next section). Linear procedures may of course be used to obtain adequate approximations for the solution of weakly non-linear problems.

For non-linear problems, the solution is less straightforward. Often the integrations in Eq. (6-4) or (6-7) can be performed analytically, no matter what the degree of non-linearity (see for example Sect. 10). The computation of the density of probability $\sigma_{\mathbf{p}}(\mathbf{p})$ at a given point \mathbf{p} then involves mainly the solution of a forward problem. If the number of parameters is small we can then explicitly compute the marginal p.d.f. for each one of the parameters, using a grid in the parameter space ordinary methods of numerical integration. If the number of parameters is great, Monte-Carlo methods of numerical integration should be used. The possibility of conveniently solving non-linear inverse problems will then depend on the possibility of solving the forward problem a large enough number of times. Let us remark that if we are not able to compute the marginal p.d.f. for each one of the parameters of interest, we can limit ourselves to the computation of mean values and covariances (Eqs. 6-11 and 6-12).

In problems where the solution of the forward problem is so costly that either the explicit computation of the density of probability in the parameter space and the computation of mean values and covariances cannot be performed in a reasonable computer time, we suggest to restrict the problem to the search of the maximum likelihood point in the parameter space (point at which the density of probability is maximum). This computation is often very easy to perform, and classical methods can be used for particular assumptions about the form of the p.d.f.'s representing experimental data and a priori assumptions on parameters. For example, it is easy to

see that with gaussian assumptions, the search of the maximum likelihood point simply becomes a classical least-squares problem (Tarantola and Valette, 1982). With exponential assumptions, the search of the maxima likelihood point becomes a L^1 -norm problem (which shows that the exponential assumption gives a result more *robust* than the gaussian one). With the use of step functions, the a posteriori p.d.f. in the parameter space is constant inside a given bounded domain. The point of this domain the maximizes some function of the parameters can be reached using the linear (or non-linear) programming techniques. We see thus that ordinary methods for solving parameterized inverse problems can be deduced as particular cases of our approach, and we want to emphasize that such methods should only be used when the explicit computation of the density of probability in the parameter space or the non-linear computation of mean values and covariances would be too much time consuming.

9. Gaussian Case

Since the gaussian linear problem is widely used, we will show how the usual formulae may be derived from our results.

By gaussian problem we mean that the a priori density function has a gaussian form for all parameters:

$$\rho(\mathbf{x}) = \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \mathbf{C}_0^{-1} \cdot (\mathbf{x} - \mathbf{x}_0)\right\} \quad (9-1)$$

where \mathbf{x}_0 is the a priori expected value and \mathbf{C}_0 is the a priori covariance matrix.

By the linear problem we mean that if the theory may be assumed to be exact, the theoretical relationship between parameters takes the general linear form:

$$\mathbf{F} \cdot \mathbf{x} = \mathbf{0}. \quad (9-2)$$

On the other hand, if theoretical errors may not be neglected, we assume that the theoretical density function also has a gaussian form:

$$\theta(\mathbf{x}) = \exp\left\{-\frac{1}{2}(\mathbf{F} \cdot \mathbf{x})^T \cdot \mathbf{C}_T^{-1} \cdot (\mathbf{F} \cdot \mathbf{x})\right\} \quad (9-3)$$

where the covariance matrix \mathbf{C}_T describes theoretical errors (and tends to vanish for an exact theory).

We finally assume that for the parameters chosen, the null information is represented by a constant function:

$$\mu(\mathbf{x}) = \text{const}. \quad (9-4)$$

The a posteriori density function (Eq. 6-1) is then given by:

$$\begin{aligned} \sigma(\mathbf{x}) &= \rho(\mathbf{x}) \cdot \theta(\mathbf{x}) \\ &= \exp\left\{-\frac{1}{2}[(\mathbf{x} - \mathbf{x}_0)^T \cdot \mathbf{C}_0^{-1} \cdot (\mathbf{x} - \mathbf{x}_0) + \mathbf{x}^T \mathbf{F}^T \cdot \mathbf{C}_T^{-1} \cdot \mathbf{F} \cdot \mathbf{x}]\right\} \end{aligned} \quad (9-5)$$

and after some matrix manipulations (see appendix) we obtain:

$$\sigma(\mathbf{x}) = \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{x}_*)^T \mathbf{C}_*^{-1} \cdot (\mathbf{x} - \mathbf{x}_*)\right\} \quad (9-6)$$

where:

$$\mathbf{x}_* = \mathbf{P} \cdot \mathbf{x}_0, \quad (9-7)$$

$$\mathbf{C}_* = \mathbf{P} \cdot \mathbf{C}_0 \quad (9-8)$$

and:

$$\mathbf{P} = \mathbf{I} - \mathbf{Q} \quad (9-9)$$

$$\mathbf{Q} = \mathbf{C}_0 \cdot \mathbf{F}^T \cdot (\mathbf{F} \cdot \mathbf{C}_0 \cdot \mathbf{F}^T + \mathbf{C}_T)^{-1} \mathbf{F}.$$

Equation (9-6) shows that the a posteriori density function is gaussian, centered in \mathbf{x}_* and with the covariance matrix \mathbf{C}_* .

If theoretical errors may be neglected, i.e. if Eq. (9-2) holds, we just drop the term \mathbf{C}_T in Eq. (9-9) to obtain the corresponding solution.

To compare our results to those published in the literature, let us assume that the separation of \mathbf{X} into the sets \mathbf{D} and \mathbf{P} is made:

$$\mathbf{x} = \begin{bmatrix} \mathbf{d} \\ \mathbf{p} \end{bmatrix}, \quad \mathbf{x}_0 = \begin{bmatrix} \mathbf{d}_0 \\ \mathbf{p}_0 \end{bmatrix}, \quad \mathbf{x}_* = \begin{bmatrix} \mathbf{d}_* \\ \mathbf{p}_* \end{bmatrix}, \quad \mathbf{C}_0 = \begin{bmatrix} \mathbf{C}_{dd} & \mathbf{C}_{dp} \\ \mathbf{C}_{pd} & \mathbf{C}_{pp} \end{bmatrix}. \quad (9-10)$$

We also assume that Eq. (9-2) simplifies to:

$$\mathbf{F} \cdot \mathbf{x} = [\mathbf{I} - \mathbf{G}] \cdot \begin{bmatrix} \mathbf{d} \\ \mathbf{p} \end{bmatrix} = \mathbf{d} - \mathbf{G} \cdot \mathbf{p} = \mathbf{0}. \quad (9-11)$$

Substituting Eqs. (9-10, 11) in Eqs. (9-7, 8, 9) we obtain the solution published by Franklin (1970) for the parametrized problem, which was obtained using the classical least squares approach. Our equations (9-7, 8, 9) are more compact than those of Franklin because we use the parameter space \mathcal{E}^m , and more general because we allow theoretical errors \mathbf{C}_T .

Let us emphasize that in traditional approaches \mathbf{x}_* is interpreted as the *best estimator of the "true" solution* and \mathbf{C}_* is interpreted as the covariance matrix of the estimator. Our approach demonstrates that the a posteriori density function is gaussian, and that \mathbf{x}_* and \mathbf{C}_* are, respectively, the center and the dispersion of the density function.

The results shown here only apply to the linear least squares problem. For the non-linear problem, the reader should refer to Tarantola and Valette (1982).

10. Example with Actual Data: Hypocenter Location

The data for a hypocenter location are the arrival times of phases at stations. The basic unknowns of the problem are the spatio-temporal coordinates of the focus. Some of the parameters which may be relevant to the problem are: the coordinates of seismic stations, the parameters describing the velocity model, etc. We will assume that the coordinates of the stations are accurately enough known to treat them as constants and not as parameters. The parameters describing the velocity model would be taken into account if we were performing a simultaneous inversion for hypocenter location and velocity determination, but this is not the case in this simple illustration.

In the example treated below we will then consider the parameters of the velocity model as constants, and we will assume that the only effect of our imprecise knowledge of the medium is not to allow an exact theoretical computation of the arrival times at the stations from a known source. Let:

$$\mathbf{t} = \mathbf{g}(X, Y, Z, T) \quad (10-1)$$

be the theoretical (and not exact) relationship between arrival times and the spatio-temporal coordinates of the focus, derived from the wave propagation theory and the velocity model. Let \mathbf{C}_T be a covariance matrix which is a reasonable estimation of the errors made when theoretically computing the arrival time at stations from a source at (X, Y, Z) . If we assume that the theoretical errors are gaussian, the theoretical

relationship between data and parameters will be written:

$$\theta(\mathbf{t}|X, Y, Z, T) = \exp\left\{-\frac{1}{2}[\mathbf{t}-\mathbf{g}(X, Y, Z, T)]^T \cdot \mathbf{C}_T^{-1} \cdot [\mathbf{t}-\mathbf{g}(X, Y, Z, T)]\right\} \quad (10-2)$$

which correspond to Eq. (5-8).

The next simplifying hypothesis is to assume that our data possess a gaussian structure. Let \mathbf{t}_0 be their vector of mean values and \mathbf{C}_t , their covariance matrix:

$$\rho(\mathbf{t}) = \exp\left\{-\frac{1}{2}(\mathbf{t}-\mathbf{t}_0)^T \cdot \mathbf{C}_t^{-1} \cdot (\mathbf{t}-\mathbf{t}_0)\right\}. \quad (10-3)$$

As all our data and parameters consist in cartesian space-time coordinates, the null information function is constant and need not be considered (see Sect. 3).

The a posteriori density function for parameters is directly given by Eq. (6-7) and after analytical integration we obtain (see appendix):

$$\sigma(X, Y, Z, T) = \rho(X, Y, Z, T) \cdot \exp\left\{-\frac{1}{2}(\mathbf{t}_0 - \mathbf{g}(X, Y, Z, T))^T \cdot (\mathbf{C}_t + \mathbf{C}_T)^{-1} \cdot (\mathbf{t}_0 - \mathbf{g}(X, Y, Z, T))\right\}. \quad (10-4)$$

The a posteriori density function (10-4) gives the general solution for the problem of spatio-temporal hypocenter location in the case of gaussian data. We emphasize that this solution does not contain any "linear approximation"

We are sometimes interested in the *spatial* location of the quake focus, and not in its *temporal* location. The density function for the spatial coordinates is obtained, of course, by the marginal density function:

$$\sigma(X, Y, Z) = \int_{-\infty}^{+\infty} \sigma(X, Y, Z, T) dT \quad (10-5)$$

where we integrate over the range of the origin time T .

Classical least squares computations of hypocenter are based on the maximization of $\sigma(X, Y, Z, T)$. It is clear that if we are only interested in the spatial location we must maximize $\sigma(X, Y, Z)$ given by (10-5) instead of maximizing $\sigma(X, Y, Z, T)$. Let us show how the integration in (10-5) can be performed analytically.

We will assume that while we may sometimes have a priori information about the spatial location of the focus (from tectonic arguments, or from past experience in the region, etc.) it is generally impossible to have a priori information (independent from the data) about the origin time T . We will thus assume an a priori density function uniform on T :

$$\rho(X, Y, Z, T) = \rho(T) \cdot \rho(X, Y, Z) = \rho(X, Y, Z). \quad (10-6)$$

The computed *arrival time* at a station i , $g_i(X, Y, Z, T)$ can be written:

$$g_i(X, Y, Z, T) = h_i(X, Y, Z) + T \quad (10-7)$$

where h_i is the *travel time* between the point (X, Y, Z) and the station i .

With (10-6) and (10-7), Eq. (10-5) can be integrated (see appendix) any yields:

$$\sigma(X, Y, Z) = K \rho(X, Y, Z) \cdot \exp\left\{-\frac{1}{2}[\tilde{\mathbf{t}}_0 - \tilde{\mathbf{h}}(X, Y, Z)]^T \cdot \mathbf{P} \cdot [\tilde{\mathbf{t}}_0 - \tilde{\mathbf{h}}(X, Y, Z)]\right\}. \quad (10-8)$$

Here

$$\mathbf{P} = (\mathbf{C}_t + \mathbf{C}_T)^{-1} \quad (10-9)$$

Table 1. Coordinates of stations (km), arrival times and errors (s)

Station	x	y	z	t	σ_t
1	49.58	9.54	-0.80	13.35	0.02
2	48.07	7.74	-0.80	13.30	0.02
3	49.67	4.22	-1.50	13.79	0.02
4	52.34	14.37	-0.30	13.70	0.02
5	43.17	5.70	-0.80	13.90	0.02
6	46.79	17.87	-0.30	14.35	0.02
7	48.82	-1.87	-0.90	14.20	0.10
8	33.31	8.72	-0.60	15.41	0.02
9	23.10	9.54	-0.50	17.09	0.02
10	13.42	15.30	-0.40	19.00	0.10
11	-2.19	11.41	-1.20	18.73	0.02

is a "weight matrix",

$$p_i = \sum_j P_{ij} \quad (10-10)$$

are "weights", and

$$K = \sum_i p_i = \sum_{ij} P_{ij}. \quad (10-11)$$

Moreover, \tilde{t}_0^i is the observed arrival time minus the weighted mean of observed arrival times,

$$\tilde{t}_0^i = t_0^i - \frac{\sum_j p_j t_0^j}{\sum_j p_j} \quad (10-12)$$

and $\tilde{h}^i(X, Y, Z)$ is the computed travel time minus the weighted mean of computed travel times

$$\tilde{h}^i = h^i - \frac{\sum_j p_j h^j}{\sum_j p_j}. \quad (10-13)$$

(Note that \mathbf{C}_T may depend on (X, Y, Z) and therefore P_{ij} , p_i , and K also.)

Equation (10-8) gives the general solution for the *spatial* location of a quake focus in the gaussian case.

Table 1 shows the observed arrival times and their standard deviation. We have assumed that the theoretical errors are of the form:

$$[\mathbf{C}_T(X, Y, Z)]_{ij} = \sigma_T^2 \cdot \exp\left\{-\frac{1}{2} \frac{D_{ij}^2}{\Delta^2}\right\} \quad (10-14)$$

where D_{ij} is the distance between the station i and the station j , σ_T is some theoretical error, and Δ is the correlation length of errors (the wavelength or the length of lateral heterogeneities of the medium). By comparison of the layered model of velocities for the Western Pyrenees (Gagnepain et al. 1980) with data from refraction profiles (Gallart 1980) we have chosen $\sigma_T = 0.2$ s and $\Delta = 0.1$ km.

We also assumed that no a priori information is known about the epicenter, but that we know that the depth of the hypocenter is greater than -0.5 km (the mean topography):

$$\rho(X, Y, Z) = \rho(Z) = \begin{cases} 1 & \text{if } Z \geq -0.5 \text{ km} \\ 0 & \text{if } Z < -0.5 \text{ km}. \end{cases} \quad (10-15)$$

We have then computed numerically from (10-8) the a posteriori marginal density functions for the epicenter and for the depth:

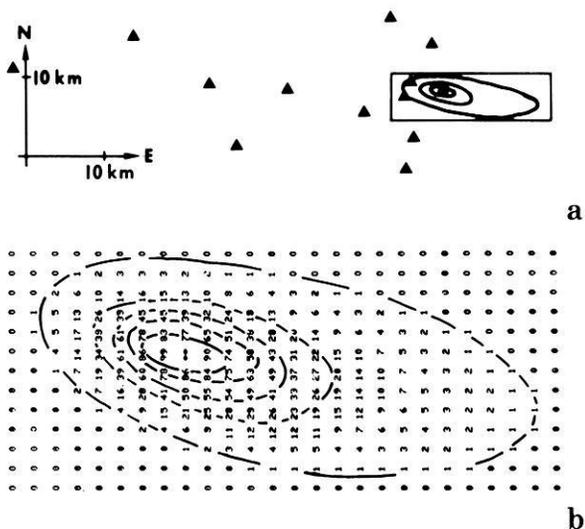


Fig. 5a and b. Results of the inverse problem of hypocenter computation. **a** shows the position of the stations and the probability density function obtained for the epicentral coordinates. **b** shows the computer output. Curves are visual interpolations

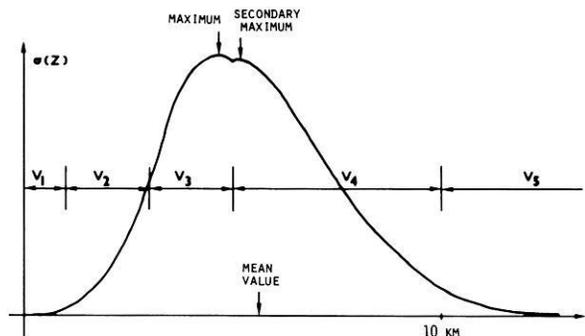


Fig. 6. The probability density function for depth. The layered velocity model is also shown. Note the existence of a secondary maximum likelihood point

$$\sigma(X, Y) = \int_{-0.5 \text{ km}}^{+\infty} \sigma(X, Y, Z) dZ, \quad (10-16)$$

$$\sigma(Z) = \int_{-\infty}^{+\infty} dX \int_{-\infty}^{+\infty} dY \sigma(X, Y, Z). \quad (10-17)$$

The results are shown in Figs. 5 and 6.

We have also computed mean values and variances. The corresponding results are, in the local frame of Fig. 5:

$$E(X) = 51.7 \text{ km} \quad \sigma_x = 1.5 \text{ km}$$

$$E(Y) = 7.8 \text{ km} \quad \sigma_y = 0.9 \text{ km}$$

$$E(Z) = 5.6 \text{ km} \quad \sigma_z = 2.1 \text{ km}.$$

We wish to make the following remarks about these results.

First, they have been obtained exactly without the use of linear approximations. We have used numerical integration instead of matrix algebra and the computation of partial derivatives. The results shown in Figs. 5 and 6 represent the most general knowledge which can be obtained for the hypocenter coordinates from the arrival times, from the given velocity model, and from the given theoretical model (of wave propagation).

Since the velocity model is discontinuous in Z the a posteriori density functions have discontinuities in slope, as it

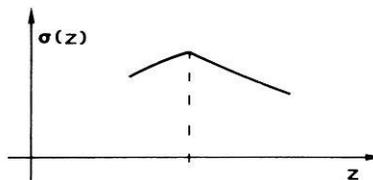


Fig. 7. Example of discontinuity of slope leading to oscillations in maximum likelihood algorithms. The effect will be an artificial accumulation at the interfaces between layers

is clearly seen in Fig. 6 for $\sigma(Z)$. To the extent that the discontinuities in the velocity model are artificial, the discontinuities of slope are of course also artificial. From Fig. 6 it is easy to visualize some of the problems which may affect the maximum likelihood approach. If the discontinuity of slope is similar to the one at 5 km depth, we will have secondary maxima. We can also have a discontinuity of slope of the type drawn in Fig. 7. In that case, algorithms searching for the maximum likelihood point will oscillate around the point of slope discontinuity, leading to the well-known artificial situation in which hypocenters accumulate at the interface between layers of constant velocity.

11. Conclusion

Our informational approach to probability calculus allows us to formulate inverse problems in such a way that all necessary constraints (see Sect. 1) are satisfied. Essentially, we propose to work with the probability density functions for parameters rather than with central estimators, as it is usually done.

The general solution of inverse problems is expressed by the simple formula (6-1). We emphasize that inverse problems cannot be correctly stated until the three density functions $\rho(\mathbf{x})$ (data and a priori information about parameters), $\theta(\mathbf{x})$ (theory and theoretical errors), and $\mu(\mathbf{x})$ (null information) have been precisely defined.

We have demonstrated that the ideas developed in this paper give new insights into the oldest and best known inverse problem in geophysics: the hypocenter location. Of course our theory also applies to more difficult inverse problems. The only practical limitation comes from problems where the solution of the forward problem is very time-consuming and the number of parameters is high.

Acknowledgements. We would like to thank our colleagues of the I.P.G. for very helpful discussions, and Professor G. Jobert, Dr. A. Nercessian, T. Van der Pyl, Dr. J.C. Houard, Dr. Piednoir, and Professor C.J. Allègre for their suggestions. This work has been partially supported by the "Recherche Coopérative sur Programme no. 264, Etude Interdisciplinaire des Problèmes Inverses". Contribution I.P.G. no. 363.

Appendix

Some Remarks on Probability Calculus

Let \mathbf{X} be a parametrization of a physical system \mathcal{S} and let $\mathbf{X}_I = \{X_1, \dots, X_r\}$ and $\mathbf{X}_{II} = \{X_{r+1}, \dots, X_m\}$ be a partition of \mathbf{X} . For any p.d.f. $f(\mathbf{x}) = f(\mathbf{x}_I, \mathbf{x}_{II})$ we can define the marginal p.d.f.

$$f_i(\mathbf{x}_I) = \int f(\mathbf{x}_I, \mathbf{x}_{II}) d\mathbf{x}_{II}. \quad (A-1)$$

The interpretation of this definition is as follows: if we admit that $f(\mathbf{x}_I, \mathbf{x}_{II})$ represents all the knowledge that we possess on the whole set of parameters and if we disregard the parameters \mathbf{X}_{II} , then all the information on \mathbf{X}_I is contained in $f_i(\mathbf{x}_I)$.

The conditional p.d.f. for \mathbf{X}_I , given $\mathbf{X}_{II} = \mathbf{x}_{II}^0$ may be defined, in our approach, as the *conjunction* of a general state of information (represented by a p.d.f. $f_i(\mathbf{x}) = f_i(\mathbf{x}_I, \mathbf{x}_{II})$) with the information $\mathbf{X}_{II} = \mathbf{x}_{II}^0$.

The information $\mathbf{X}_{II} = \mathbf{x}_{II}^0$ clearly corresponds to the p.d.f.

$$f_j(\mathbf{x}_I, \mathbf{x}_{II}) = \mu_I(\mathbf{x}_I) \cdot \delta(\mathbf{x}_{II} - \mathbf{x}_{II}^0) \quad (\text{A-2})$$

because f_j does not contain information on \mathbf{X}_I and gives null probability for all values of \mathbf{X}_{II} different from \mathbf{x}_{II}^0 . $\mu_I(\mathbf{x}_I)$ represents the null information on \mathbf{X}_I . Admitting that null informations are independent: $\mu(\mathbf{x}_I, \mathbf{x}_{II}) = \mu_I(\mathbf{x}_I) \cdot \mu_{II}(\mathbf{x}_{II})$ and using Eq. (2-19) we obtain the combined probability:

$$f(\mathbf{x}_I, \mathbf{x}_{II}) = \frac{f_i(\mathbf{x}_I, \mathbf{x}_{II}) \cdot \mu_I(\mathbf{x}_I) \cdot \delta(\mathbf{x}_{II} - \mathbf{x}_{II}^0)}{\mu_I(\mathbf{x}_I) \cdot \mu_{II}(\mathbf{x}_{II})}. \quad (\text{A-3})$$

Using definition (A-1) we obtain

$$f_i(\mathbf{x}_I) = \frac{f_i(\mathbf{x}_I, \mathbf{x}_{II}^0)}{\int f_i(\mathbf{x}_I, \mathbf{x}_{II}^0) d\mathbf{x}_I}, \quad (\text{A-4})$$

which corresponds to what is ordinarily named the *conditional* p.d.f. for \mathbf{X}_I given $f_i(\mathbf{x}_I, \mathbf{x}_{II})$ and $\mathbf{X}_{II} = \mathbf{x}_{II}^0$. To follow the usual notation we will write this solution $f_i(\mathbf{x}_I | \mathbf{x}_{II}^0)$ rather than $f_i(\mathbf{x}_I)$:

$$f_i(\mathbf{x}_I | \mathbf{x}_{II}^0) = \frac{f_i(\mathbf{x}_I, \mathbf{x}_{II}^0)}{\int f_i(\mathbf{x}_I, \mathbf{x}_{II}^0) d\mathbf{x}_I}. \quad (\text{A-5})$$

The Bayes problem may be states as follows: Let $f(\mathbf{x}_I, \mathbf{x}_{II})$ be the joint p.d.f. representing all the available information on \mathbf{X}_I and \mathbf{X}_{II} . If we learn that $\mathbf{X}_{II} = \mathbf{x}_{II}$ we obtain $g(\mathbf{x}_I | \mathbf{x}_{II})$ using Eq. (A-4). To the contrary, if we learn $\mathbf{X}_I = \mathbf{x}_I$ we obtain $g(\mathbf{x}_{II} | \mathbf{x}_I)$. Which is the relation between $g(\mathbf{x}_I | \mathbf{x}_{II})$ and $g(\mathbf{x}_{II} | \mathbf{x}_I)$?

We have

$$g(\mathbf{x}_I | \mathbf{x}_{II}) = \frac{f(\mathbf{x}_I, \mathbf{x}_{II})}{\int f(\mathbf{x}_I, \mathbf{x}_{II}) d\mathbf{x}_I} = \frac{f(\mathbf{x}_I, \mathbf{x}_{II})}{f_{II}(\mathbf{x}_{II})} \quad (\text{A-6})$$

$$g(\mathbf{x}_{II} | \mathbf{x}_I) = \frac{f(\mathbf{x}_I, \mathbf{x}_{II})}{\int f(\mathbf{x}_I, \mathbf{x}_{II}) d\mathbf{x}_{II}} = \frac{f(\mathbf{x}_I, \mathbf{x}_{II})}{\int g(\mathbf{x}_I | \mathbf{x}_{II}) \cdot f_{II}(\mathbf{x}_{II}) \cdot d\mathbf{x}_{II}}$$

and hence

$$g(\mathbf{x}_{II} | \mathbf{x}_I) = \frac{g(\mathbf{x}_I | \mathbf{x}_{II}) \cdot f_{II}(\mathbf{x}_{II})}{\int g(\mathbf{x}_I | \mathbf{x}_{II}) \cdot f_{II}(\mathbf{x}_{II}) \cdot d\mathbf{x}_{II}} \quad (\text{A-7})$$

which corresponds to Bayes theorem.

We have thus shown that well known theorems may be obtained using the concept of the conjunction of states of information. Many other problems may be solved using this concept. Consider for example n independent measurements of a given parameter X . In the particular case where the null information density is uniform ($\mu(x) = \text{const.}$), and each measurement gives a gaussian p.d.f. $f_i(x)$ centered at x_i and with variance σ^2

$$f_i(x) = \frac{1}{\sqrt{2\pi} \sigma} \exp \left\{ -\frac{1}{2} \frac{(x - x_i)^2}{\sigma^2} \right\}, \quad (\text{A-8})$$

the iteration of Eq. (2-19) at each new measurement gives:

$$f(x) = \frac{1}{\sqrt{2\pi} \Sigma} \exp \left\{ -\frac{1}{2} \frac{(x - \bar{x})^2}{\Sigma^2} \right\} \quad (\text{A-9})$$

where:

$$\bar{x} = \frac{\Sigma x_i}{n} \quad \Sigma = \frac{\sigma}{\sqrt{n}} \quad (\text{A-10})$$

which are well known results in statistics.

Demonstrations for Sect. 9

Let us first demonstrate two useful identities. If \mathbf{C}_1 and \mathbf{C}_2 are two positive definite matrices respectively of order $(n \times n)$ and $(m \times m)$, and \mathbf{M} an arbitrary $(n \times m)$ matrix, then:

$$(\mathbf{M}^T \mathbf{C}_1^{-1} \mathbf{M} + \mathbf{C}_2^{-1})^{-1} \mathbf{M}^T \mathbf{C}_1^{-1} = \mathbf{C}_2 \mathbf{M}^T (\mathbf{C}_1 + \mathbf{M} \mathbf{C}_2 \mathbf{M}^T)^{-1}, \quad (\text{A-11})$$

$$(\mathbf{M}^T \mathbf{C}_1^{-1} \mathbf{M} + \mathbf{C}_2^{-1})^{-1} = \mathbf{C}_2 - \mathbf{C}_2 \mathbf{M}^T (\mathbf{C}_1 + \mathbf{M} \mathbf{C}_2 \mathbf{M}^T)^{-1} \mathbf{M} \mathbf{C}_2. \quad (\text{A-12})$$

The first equation follows from the following obvious identities

$$\begin{aligned} \mathbf{M}^T + \mathbf{M}^T \mathbf{C}_1^{-1} \mathbf{M} \mathbf{C}_2 \mathbf{M}^T &= \mathbf{M}^T \mathbf{C}_1^{-1} (\mathbf{C}_1 + \mathbf{M} \mathbf{C}_2 \mathbf{M}^T) \\ &= (\mathbf{M}^T \mathbf{C}_1^{-1} \mathbf{M} + \mathbf{C}_2^{-1}) \mathbf{C}_2 \mathbf{M}^T \end{aligned} \quad (\text{A-13})$$

since $\mathbf{M}^T \mathbf{C}_1^{-1} \mathbf{M} + \mathbf{C}_2^{-1}$ and $\mathbf{C}_1 + \mathbf{M} \mathbf{C}_2 \mathbf{M}^T$ are definite positive and thus regular matrices.

Furthermore (A-11) leads to

$$\begin{aligned} \mathbf{C}_2 - \mathbf{C}_2 \mathbf{M}^T (\mathbf{C}_1 + \mathbf{M} \mathbf{C}_2 \mathbf{M}^T)^{-1} \mathbf{M} \mathbf{C}_2 \\ &= \mathbf{C}_2 - (\mathbf{M}^T \mathbf{C}_1^{-1} \mathbf{M} + \mathbf{C}_2^{-1})^{-1} \mathbf{M}^T \mathbf{C}_1^{-1} \mathbf{M} \mathbf{C}_2 \\ &= (\mathbf{M}^T \mathbf{C}_1^{-1} \mathbf{M} + \mathbf{C}_2^{-1})^{-1} \{ (\mathbf{M}^T \mathbf{C}_1^{-1} \mathbf{M} + \mathbf{C}_2^{-1}) \mathbf{C}_2 - \mathbf{M}^T \mathbf{C}_1^{-1} \mathbf{M} \mathbf{C}_2 \} \\ &= (\mathbf{M}^T \mathbf{C}_1^{-1} \mathbf{M} + \mathbf{C}_2^{-1})^{-1} \end{aligned} \quad (\text{A-14})$$

which proves (A-12).

Now from Eq. (9-5) we obtain:

$$\begin{aligned} \sigma(\mathbf{x}) &= \exp \left\{ -\frac{1}{2} [(\mathbf{x} - \mathbf{x}_0)^T \mathbf{C}_0^{-1} (\mathbf{x} - \mathbf{x}_0) + \mathbf{x}^T \mathbf{F}^T \mathbf{C}_T^{-1} \mathbf{F} \mathbf{x}] \right\} \\ &= \exp \left\{ -\frac{1}{2} [\mathbf{x}^T (\mathbf{C}_0^{-1} + \mathbf{F}^T \mathbf{C}_T^{-1} \mathbf{F}) \mathbf{x} - 2 \mathbf{x}^T \mathbf{C}_0^{-1} \mathbf{x}_0 + \mathbf{x}_0^T \mathbf{C}_0^{-1} \mathbf{x}_0] \right\} \end{aligned} \quad (\text{A-15})$$

then defining:

$$\mathbf{P} = \mathbf{I} - \mathbf{C}_0 \mathbf{F}^T (\mathbf{F} \mathbf{C}_0 \mathbf{F}^T + \mathbf{C}_T)^{-1} \mathbf{F}, \quad (\text{A-16})$$

$$\mathbf{C}_* = \mathbf{P} \mathbf{C}_0, \quad (\text{A-17})$$

$$\mathbf{x}_* = \mathbf{P} \mathbf{x}_0. \quad (\text{A-18})$$

We obtain, using Eq. (A-12)

$$\begin{aligned} \mathbf{C}_* &= \mathbf{P} \mathbf{C}_0 = \mathbf{C}_0 - \mathbf{C}_0 \mathbf{F}^T (\mathbf{F} \mathbf{C}_0 \mathbf{F}^T + \mathbf{C}_T)^{-1} \mathbf{F} \mathbf{C}_0 \\ &= (\mathbf{F}^T \mathbf{C}_T^{-1} \mathbf{F} + \mathbf{C}_0^{-1})^{-1} \end{aligned} \quad (\text{A-19})$$

and

$$\mathbf{x}_* = \mathbf{P} \mathbf{x}_0 = \mathbf{C}_* \mathbf{C}_0^{-1} \mathbf{x}_0. \quad (\text{A-20})$$

Thus Eq. (A-15) becomes:

$$\begin{aligned} \sigma(\mathbf{x}) &= \exp \left\{ -\frac{1}{2} (\mathbf{x}^T \mathbf{C}_*^{-1} \mathbf{x} - 2 \mathbf{x}^T \mathbf{C}_*^{-1} \mathbf{x}_* + \mathbf{x}_0^T \mathbf{C}_*^{-1} \mathbf{x}_*) \right\} \\ &= \exp \left\{ -\frac{1}{2} [(\mathbf{x} - \mathbf{x}_*)^T \mathbf{C}_*^{-1} (\mathbf{x} - \mathbf{x}_*) - (\mathbf{x}_* - \mathbf{x}_0)^T \mathbf{C}_0^{-1} \mathbf{x}_0] \right\}. \end{aligned} \quad (\text{A-21})$$

From (A-18), we deduce:

$$\mathbf{x}_* - \mathbf{x}_0 = -\mathbf{C}_0 \mathbf{F}^T (\mathbf{F} \mathbf{C}_0 \mathbf{F}^T + \mathbf{C}_T)^{-1} \mathbf{F} \mathbf{C}_0 \mathbf{x}_0 \quad (\text{A-22})$$

and then:

$$\begin{aligned} \sigma(\mathbf{x}) &= \exp \left\{ -\frac{1}{2} \mathbf{x}_0^T \mathbf{F}^T (\mathbf{F} \mathbf{C}_0 \mathbf{F}^T + \mathbf{C}_T)^{-1} \mathbf{F} \mathbf{x}_0 \right\} \\ &\cdot \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{x}_*)^T \mathbf{C}_*^{-1} (\mathbf{x} - \mathbf{x}_*) \right\} \\ &= \text{const.} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{x}_*)^T \mathbf{C}_*^{-1} (\mathbf{x} - \mathbf{x}_*) \right\} \end{aligned} \quad (\text{A-23})$$

which demonstrates Eq. (9-6).

Demonstrations for Sect. 10

Let us now evaluate the sum:

$$\begin{aligned} I &= \int \exp \left\{ -\frac{1}{2} [(\mathbf{d} - \mathbf{d}_0)^T \mathbf{C}_d^{-1} (\mathbf{d} - \mathbf{d}_0) \right. \\ &\left. + (\mathbf{d} - \mathbf{g}(\mathbf{p}))^T \mathbf{C}_T^{-1} (\mathbf{d} - \mathbf{g}(\mathbf{p})) \right\} d\mathbf{d}. \end{aligned} \quad (\text{A-24})$$

The separation of the quadratic terms from the linear terms leads to:

$$I = \int \exp \left\{ -\frac{1}{2} (\mathbf{d}^T \mathbf{A} \mathbf{d} - 2 \mathbf{B}^T \mathbf{d} + \mathbf{C}) \right\} d\mathbf{d} \quad (\text{A-25})$$

where:

$$\begin{aligned} \mathbf{A} &= \mathbf{C}_d^{-1} + \mathbf{C}_T^{-1} \\ \mathbf{B}^T &= \mathbf{d}_0^T \mathbf{C}_d^{-1} + \mathbf{g}(\mathbf{p})^T \mathbf{C}_T^{-1} \\ \mathbf{C} &= \mathbf{d}_0^T \mathbf{C}_d^{-1} \mathbf{d}_0 + \mathbf{g}(\mathbf{p})^T \mathbf{C}_T^{-1} \mathbf{g}(\mathbf{p}). \end{aligned} \quad (\text{A-26})$$

Since \mathbf{A} is positive definite there follows:

$$\begin{aligned}
I &= \int \exp \left\{ -\frac{1}{2} [(\mathbf{d} - \mathbf{A}^{-1} \mathbf{B})^T \mathbf{A} (\mathbf{d} - \mathbf{A}^{-1} \mathbf{B}) + (\mathbf{C} - \mathbf{B}^T \mathbf{A}^{-1} \mathbf{B})] \right\} d\mathbf{d} \\
&= \exp \left\{ -\frac{1}{2} (\mathbf{C} - \mathbf{B}^T \mathbf{A}^{-1} \mathbf{B}) \right\} \int \exp \left\{ -\frac{1}{2} (\mathbf{d} - \mathbf{A}^{-1} \mathbf{B})^T \mathbf{A} (\mathbf{d} - \mathbf{A}^{-1} \mathbf{B}) \right\} d\mathbf{d} \\
&= (2\pi)^{n/2} (\det \mathbf{A})^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{C} - \mathbf{B}^T \mathbf{A}^{-1} \mathbf{B}) \right\}. \tag{A-27}
\end{aligned}$$

By substitution of (A-26) we obtain:

$$\begin{aligned}
\mathbf{C} - \mathbf{B}^T \mathbf{A}^{-1} \mathbf{B} &= \mathbf{d}_0^T (\mathbf{C}_d^{-1} - \mathbf{C}_d^{-1} (\mathbf{C}_d^{-1} + \mathbf{C}_T^{-1}) \mathbf{C}_d^{-1}) \mathbf{d}_0 \\
+ \mathbf{g}(\mathbf{p})^T (\mathbf{C}_T^{-1} - \mathbf{C}_T^{-1} (\mathbf{C}_d^{-1} + \mathbf{C}_T^{-1}) \mathbf{C}_T^{-1}) \mathbf{g}(\mathbf{p}) \\
- 2\mathbf{g}(\mathbf{p}) \mathbf{C}_T^{-1} (\mathbf{C}_d^{-1} + \mathbf{C}_T^{-1}) \mathbf{C}_d^{-1} \mathbf{d}_0. \tag{A-28}
\end{aligned}$$

Thus, by using the two identities (A-11-A-12) we get:

$$\begin{aligned}
\mathbf{C} - \mathbf{B}^T \mathbf{A}^{-1} \mathbf{B} &= \mathbf{d}_0 (\mathbf{C}_d + \mathbf{C}_T)^{-1} \mathbf{d}_0 \\
+ \mathbf{g}(\mathbf{p})^T (\mathbf{C}_d + \mathbf{C}_T)^{-1} \mathbf{g}(\mathbf{p}) - 2\mathbf{g}(\mathbf{p}) (\mathbf{C}_d + \mathbf{C}_T)^{-1} \mathbf{d}_0 \\
= (\mathbf{d}_0 - \mathbf{g}(\mathbf{p}))^T (\mathbf{C}_d + \mathbf{C}_T)^{-1} (\mathbf{d}_0 - \mathbf{g}(\mathbf{p})). \tag{A-29}
\end{aligned}$$

Finally we obtain:

$$\begin{aligned}
I &= (2\pi)^{n/2} [\det(\mathbf{C}_d^{-1} + \mathbf{C}_T^{-1})]^{-1/2} \\
&\cdot \exp \left\{ -\frac{1}{2} (\mathbf{d}_0 - \mathbf{g}(\mathbf{p}))^T (\mathbf{C}_d + \mathbf{C}_T)^{-1} (\mathbf{d}_0 - \mathbf{g}(\mathbf{p})) \right\} \tag{A-30}
\end{aligned}$$

which demonstrates Eq. (10-4).

Let us define

$$\mathbf{P} = (\mathbf{C}_d + \mathbf{C}_T)^{-1}. \tag{A-31}$$

Using 10-4 and 10-7, the sum 10-5 becomes:

$$\begin{aligned}
I &= \int \exp \left\{ -\frac{1}{2} \sum_{ij} [t_i^0 - h_i - T] P_{ij} [t_j^0 - h_j - T] \right\} dT \\
&= \int \exp \left\{ -\frac{1}{2} (dT^2 - 2bT + c) \right\} dT \tag{A-32}
\end{aligned}$$

where:

$$\begin{aligned}
a &= \sum_{ij} P_{ij} \\
b &= \sum_{ij} P_{ij} \cdot (t_j^0 - h_j) \\
c &= \sum_{ij} (t_i^0 - h_i) \cdot P_{ij} \cdot (t_j^0 - h_j). \tag{A-33}
\end{aligned}$$

This yields:

$$\begin{aligned}
I &= \int \exp \left\{ -\frac{1}{2} \left[a \left(T - \frac{b}{a} \right)^2 + \left(c - \frac{b^2}{a} \right) \right] \right\} dT \\
&= \left(\frac{2\pi}{a} \right)^{1/2} \exp \left\{ -\frac{1}{2} \left(c - \frac{b^2}{a} \right) \right\}. \tag{A-34}
\end{aligned}$$

By substitution of a , b and c given in (A-33) in the above expression, we obtain:

$$\begin{aligned}
I &= \left(\frac{2\pi}{\sum_{ij} P_{ij}} \right)^{1/2} \cdot \exp \left\{ -\frac{1}{2} \left[\sum_{ij} (t_i^0 - h_i) \cdot P_{ij} \cdot (t_j^0 - h_j) \right. \right. \\
&\quad \left. \left. - \frac{\left[\sum_{ij} P_{ij} \cdot (t_j^0 - h_j) \right]^2}{\sum_{ij} P_{ij}} \right] \right\} \tag{A-35}
\end{aligned}$$

which can also be written:

$$\begin{aligned}
I &= \left(\frac{2\pi}{\sum_{ij} P_{ij}} \right)^{1/2} \cdot \exp \left\{ -\frac{1}{2} \sum_{ij} \left[t_i^0 - h_i - \frac{\sum_{kl} P_{kl} (t_l^0 - h_l)}{\sum_{kl} P_{kl}} \right] \right. \\
&\quad \left. \cdot P_{ij} \cdot \left[t_j^0 - h_j - \frac{\sum_{kl} P_{kl} (t_l^0 - h_l)}{\sum_{kl} P_{kl}} \right] \right\} \tag{A-36}
\end{aligned}$$

or

$$I = \left(\frac{2\pi}{\sum_{ij} P_{ij}} \right)^{1/2} \cdot \exp \left\{ -\frac{1}{2} \sum_{ij} (\hat{t}_i^0 - \hat{h}_i) \cdot P_{ij} \cdot (\hat{t}_j^0 - \hat{h}_j) \right\} \tag{A-37}$$

where

$$\hat{t}_i^0 = t_i^0 - \frac{\sum_{kl} P_{kl} \cdot t_l^0}{\sum_{kl} P_{kl}} \quad \hat{h}_i = h_i - \frac{\sum_{kl} P_{kl} \cdot h_l}{\sum_{kl} P_{kl}} \tag{A-38}$$

which is the expression (10-8).

References

- Backus, G., Gilbert, F.: Numerical applications of a formalism for geophysical inverse problems. *Geophys. J. R. Astron. Soc.* **13**, 247-276, 1967
- Backus, G., Gilbert, F.: The resolving power of gross earth data - *Geophys. J. R. Astron. Soc.* **16**, 169-205, 1968
- Backus, G., Gilbert, F.: Uniqueness in the inversion of inaccurate gross earth data. *Philos. Trans. R. Soc. London* **266**, 123-192, 1970
- Backus, G.: Inference from inadequate and inaccurate data, *Mathematical problems in the Geophysical Sciences: Lecture in applied Mathematics*, **14**, American Mathematical Society, Providence, Rhode Island, 1971
- Clearbout, J.F., Muir, F.: Robust modelling with erratic data. *Geophysics* **38**, 5, 826-844, 1973
- Descombes, R.: *Integration. Corollary 6.2*, p. 106. Paris: Hermann 1972
- Franklin, J.N.: Well posed stochastic extension of ill posed linear problems. *J. Math. Anal. Applic.* **31**, 682-716, 1970
- Gagnepain, J., Modiano, T., Cisternas, A., Ruegg, J.C., Vadell, M., Hatzfeld, D., Mezcuca, J.: Sismicité de la région d'Arette (Pyrénées-Atlantiques) et mécanismes au foyer. *Ann. Géophys.* **36**, 499-508, 1980
- Gallart, J.: Structure crustale des Pyrénées d'après les études de sismologie expérimentale. Thesis of 3rd cycle. 132 p. Université Paris VI, 1980
- Jackson, D.D.: Interpretation of inaccurate, insufficient and inconsistent data. *Geophys. J. R. Astron. Soc.* **28**, 97-110, 1972
- Jaynes, E.T.: Prior probabilities. *I.E.E.E. Transactions on systems, Science and cybernetics*. Vol. SSC-4, No. 3, 227-241, 1968
- Jeffreys, H.: *Theory of probability*, Oxford: Clarendon Press 1939
- Jeffreys, H.: *Scientific Inference*, London: Cambridge University Press 1957
- Keilis-Borok, V.I., Yanovskaya: Inverse problems in seismology, *Geophys. J. R. Astron. Soc.* **13**, 223-234, 1967
- Parker, R.L.: The theory of ideal bodies for gravity interpretation. *Geophys. J. R. Astron. Soc.* **42**, 315-334, 1975
- Rietsch, E.: The maximum entropy approach to Inverse Problems. *J. Geophys.* **42**, 489-506, 1977
- Sabatier, P.C.: On geophysical inverse problems and constraints. *J. Geophys.* **43**, 115-137, 1977
- Shannon, C.E.: A mathematical theory of communication - *Bell System Tech. J.* **27**, 379-423, 1948
- Tarantola, A., Valette, B.: Generalized non linear inverse problems solved using the least squares criterion. *Rev. Geophys. Space Phys.*, **19**, No. 2, 1982 (in press)
- Wiggins, R.A.: The general inverse problem: implication of surface waves and free oscillations for earth structure. *Rev. Geophys. Space Phys.* **10**, 251-285, 1972

Reflection and Transmission of Love Channel Waves at Coal Seam Discontinuities Computed with a Finite Difference Method*

M. Korn¹ and H. Stöckl²

¹ Institut für Meteorologie und Geophysik, Feldbergstr. 47, D-6000 Frankfurt, Federal Republic of Germany

² Fraunhofer-Institut für Werkstoffmechanik, Rosastr. 9, D-7800 Freiburg, Federal Republic of Germany

Abstract. Channel waves serve as a tool for the detection of discontinuities in coal seams caused, for example, by tectonic faulting. In this paper we study Love waves propagating along two-dimensional discontinuous coal seams. Synthetic seismograms, computed with an explicit finite difference method, are presented for three types of discontinuities: the seam end, the horizontal and the vertical offset. In all cases the discontinuity reflects mainly those waves with short wavelength and transmits those waves with large wavelength. An additional damping term is introduced into the finite difference formulation in order to prevent reflections from the edges of the computational grid. The boundary conditions at interfaces are all approximated with a truncation error of second order.

Key words: Synthetic seismograms – Finite difference method – Channel waves – Coal seam discontinuities

Introduction

Coal seams are often not continuous but are disturbed by micro-tectonic faults, sand channels or tectonic displacements. For an effective exploitation of a mine it is important to explore such discontinuities in advance. Channel waves are most commonly used to do this (Krey 1963). These waves are radiated from a source located within the coal seam and they are also recorded within the seam. Depending on the geophone setup, either the transmitted or the reflected channel wave is recorded. Marked losses in transmission or the occurrence of a strong reflection indicate the presence of a discontinuity.

For a better understanding of the properties of these waves, it is useful to investigate the reflection and transmission process theoretically. In the past, channel waves of the Rayleigh type were analyzed extensively by model seismic experiments (Dresen and Freystätter 1976; Freystätter and Dresen 1978). In real prospecting, however, mainly channel waves of the Love type are recorded, which are difficult to study by model seismic means. Therefore we construct numerically synthetic seismograms of Love channel waves propagating in discontinuous coal seams. The method used is an explicit finite difference method, which allows the numerical solution of the wave equation in two-dimensional inhomogeneous media. In the past this technique has been successfully applied to numerous wave propagation problems

(Alterman and Loewenthal 1972; Boore 1972; Kelly et al. 1976).

In this paper we use a new approximation of the boundary conditions at interfaces, which has some advantage compared to the usual approximation (Alterman and Rotenberg 1969), as will be explained later. In addition we include a damping term into the wave equation which allows us to suppress the artificial reflections from the edge of the computational grid. In the next section these extensions of the finite difference method will be described. In the subsequent section the propagation of a Love wave in a coal seam ending with a discontinuity perpendicular to the bedding plane is discussed and a simple analytical approximation of the frequency dependence of the reflectivity is derived. Some more complicated seam discontinuities are also discussed.

Method of Calculation

The equation of motion in heterogeneous elastic media can be solved with two different kinds of finite-difference schemes. In the so-called “homogeneous formulation”, the equation of motion for homogeneous media is solved numerically. At interfaces between media with different elastic parameters the boundary conditions for continuity of stress and displacement have to be fulfilled in addition. The “heterogeneous formulation” (Boore 1972; Kelly et al. 1976) uses the equation of motion in heterogeneous media, where the elastic parameters are arbitrarily space dependent. In this case there is no need to fulfil boundary conditions explicitly. Kelly et al. (1976) showed that these two formulations, when applied to the same problem, produce slightly different results. In our treatment we employ the homogeneous formulation, but use a different approximation of the boundary conditions with a truncation error of second order. With this approximation we are able to generalize the homogeneous formulation, so that it becomes equivalent to a heterogeneous formulation.

We start with the equation of motion for horizontally polarized shear (*SH*) waves in cartesian coordinates x and z :

$$\rho v_{tt} = \mu(v_{xx} + v_{zz}). \quad (1)$$

ρ is the density, μ is the shear modulus and $v(x, z, t)$ is the displacement. Subscripts t, x, z mean the partial derivatives $\partial/\partial t, \partial/\partial x, \partial/\partial z$.

The boundary conditions at a horizontal interface at depth $z=a$ are:

$$\mu^{(1)} v_z^{(1)}|_{z=a} = \mu^{(2)} v_z^{(2)}|_{z=a} \quad (2a)$$

$$v^{(1)}|_{z=a} = v^{(2)}|_{z=a} \quad (2b)$$

Here, the indices in brackets denote the different media.

* Contribution No. 241, Geophysikalisches Institut, Universität Karlsruhe

^{1,2} Formerly at: Geophysikalisches Institut, Universität Karlsruhe, Hertzstr. 16, D-7500 Karlsruhe, Federal Republic of Germany

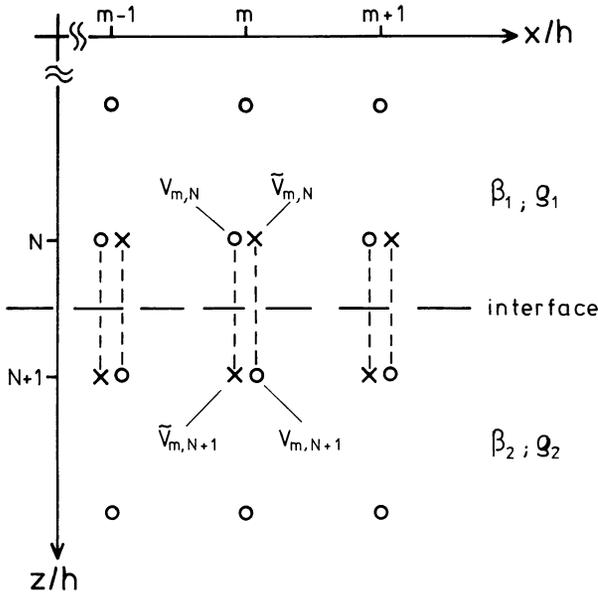


Fig. 1. Arrangement of grid points at a horizontal interface at $z = (N + 1/2)h$. Real points are denoted by \circ , fictitious points by \times . The fictitious points below the interface belong to the upper medium and vice versa

For the discretization, a rectangular grid with equal spacings $\Delta x = \Delta z = h$ is employed. The time step is Δt . Using the notation $v(mh, nh, p\Delta t) = v_{m,n}^p$, the well-known finite difference approximation of Eq. (1) with a truncation error of second order is (Boore 1970):

$$v_{m,n}^{p+1} = -v_{m,n}^{p-1} + \left(\frac{\beta \Delta t}{h}\right)^2 (v_{m+1,n}^p + v_{m-1,n}^p + v_{m,n+1}^p + v_{m,n-1}^p) + 2 \left[1 - 2 \left(\frac{\beta \Delta t}{h}\right)^2\right] v_{m,n}^p \quad (3)$$

where $\beta = (\mu/\rho)^{1/2}$ is the shear velocity:

This scheme is numerically stable, if the condition $\frac{\beta \Delta t}{h} \leq 1/\sqrt{2}$ is fulfilled.

The boundary conditions, Eq. (2) are usually approximated by introducing one line of fictitious grid points at a distance h from the boundary and approximating the derivatives in Eq. (2a) by finite differences with a truncation error of first order (Alterman and Loewenthal 1972). This obviously causes some additional error in the numerical solution.

In our approximation we put an interface between two grid lines N and $N+1$ and add a line of fictitious points on either side of the interface (see Fig. 1). The boundary condition Eq. (2a) is approximated using central differences with respect to the interface at $N + \frac{1}{2}$ resulting in a truncation error of second order:

$$\frac{\mu^{(1)}}{h} \cdot (\tilde{v}_{m,N+1}^p - v_{m,N}^p) = \frac{\mu^{(2)}}{h} \cdot (v_{m,N+1}^p - \tilde{v}_{m,N}^p). \quad (4a)$$

Fictitious points are denoted by a tilde.

Equation (2b) is approximated by linear interpolation between neighbouring grid points:

$$\frac{1}{2}(v_{m,N}^p + \tilde{v}_{m,N+1}^p) = \frac{1}{2}(\tilde{v}_{m,N}^p + v_{m,N+1}^p). \quad (4b)$$

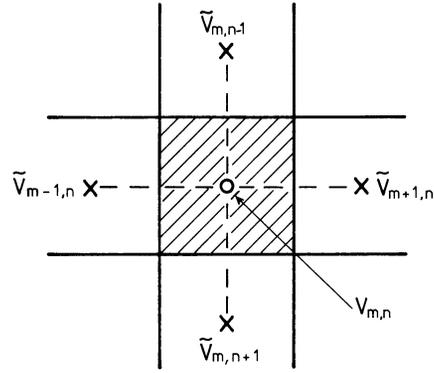


Fig. 2. Arrangement of fictitious grid points (\times) in a heterogeneous medium in the neighbourhood of a real point (\circ) at $x = mh$, $y = nh$. The hatched area indicates the homogeneous region with density $\rho_{m,n}$ and shear velocity $\beta_{m,n}$ surrounding the point (m, n)

Equations (4a) and (4b) can be solved for the unknown fictitious points:

$$\begin{cases} \tilde{v}_{m,N}^p = (2v_{m,N}^p + (M-1)v_{m,N+1}^p)/(M+1) \\ \tilde{v}_{m,N+1}^p = (2v_{m,N+1}^p + (M-1)v_{m,N}^p)/(M+1) \end{cases} \quad (5)$$

with $M = \mu^{(2)}/\mu^{(1)}$.

This formulation can be analogously applied to vertical interfaces and so any interface consisting of horizontal and vertical sections can be modeled.

We now consider a completely heterogeneous medium with velocity $\beta(x, z)$ and density $\rho(x, z)$. In order to apply the above concept of boundary conditions to this case, we assume that each grid point (m, n) is located in the center of a homogeneous quadratic region with parameters $\beta_{m,n}$ and $\rho_{m,n}$ (see Fig. 2). Between every two neighbouring points there exists an interface, where the boundary conditions of Eq. (2) have to be satisfied. The finite difference approximation of the homogeneous wave equation for the grid point (m, n) is, according to Eq. (3):

$$v_{m,n}^{p+1} = -v_{m,n}^{p-1} + \left(\frac{\beta_{m,n} \Delta t}{h}\right)^2 (\tilde{v}_{m+1,n}^p + \tilde{v}_{m-1,n}^p + \tilde{v}_{m,n+1}^p + \tilde{v}_{m,n-1}^p) + 2 \left[1 - 2 \left(\frac{\beta_{m,n} \Delta t}{h}\right)^2\right] v_{m,n}^p \quad (6)$$

Note that fictitious points have been used for all neighbouring points. The displacements at the fictitious points can be expressed in terms of real points by using Eq. (5) appropriately for all four boundaries. Equation (6) then becomes

$$\begin{aligned} v_{m,n}^{p+1} = & -v_{m,n}^{p-1} + 2v_{m,n}^p + 2 \left(\frac{\beta_{m,n} \Delta t}{h}\right)^2 \left[M_1 \cdot (v_{m+1,n}^p - v_{m,n}^p) \right. \\ & - M_2 (v_{m,n}^p - v_{m-1,n}^p) + M_3 (v_{m,n+1}^p - v_{m,n}^p) \\ & \left. - M_4 (v_{m,n}^p - v_{m,n-1}^p) \right] \end{aligned} \quad (7)$$

with

$$\begin{aligned} M_1 &= \frac{\mu_{m+1,n}}{\mu_{m,n} + \mu_{m+1,n}}; & M_2 &= \frac{\mu_{m-1,n}}{\mu_{m,n} + \mu_{m-1,n}}; \\ M_3 &= \frac{\mu_{m,n+1}}{\mu_{m,n} + \mu_{m,n+1}}; & M_4 &= \frac{\mu_{m,n-1}}{\mu_{m,n} + \mu_{m,n-1}} \end{aligned}$$

We have now obtained a heterogeneous formulation, because Eq. (7) is valid in a medium, in which the elastic parameters

may vary from grid point to grid point. In contrast to the heterogeneous formulation of Boore (1972) and Kelly et al. (1976), Eq. (7) is completely equivalent to the homogeneous formulation (Eqs. (3) and (4)). Furthermore, second order approximations are used throughout the whole scheme.

It is interesting to note that Eq. (7) can also be obtained following the method of Tikhonov and Samarskii (Mitchell 1969), which was also discussed in Boore (1972):

Their approach to the equation of motion in heterogeneous media depends on the detailed variation of $\mu(x, z)$. In this case the coefficients M_1 to M_4 in Eq. (7) have to be replaced by integral formulas, for example

$$M_1 = \frac{h}{2\mu_{m,n}} \cdot \left(\int_{x_m}^{x_{m+1}} \frac{dx}{\mu(x, z)} \right)^{-1} \quad (8)$$

If μ jumps from $\mu_{m,n}$ to $\mu_{m+1,n}$ at $x_{m+\frac{1}{2}}$, Eq. (8) can be integrated and M_1 is the same as in Eq. (7).

A problem which is common to all finite difference computations is the occurrence of reflections from the boundaries of the model. Several methods have been developed to attenuate these edge reflections (Smith 1974; Clayton and Engquist 1977; Reynolds 1978), but none is able to prevent them completely. In this paper our approach is to modify the equation of motion in such a way that the energy is dissipated within the medium. For the models in our study this proved to be a practicable way to suppress disturbing reflections almost completely.

First, Eq. (1) is transformed into the frequency domain and the frequency ω is substituted by the complex frequency $\omega - i\sigma$. Transformation back into the time domain yields a modified equation of motion:

$$v_{tt} + 2\sigma v_t + \sigma^2 v = \beta^2 (v_{xx} + v_{zz}) \quad (9)$$

Solutions of Eq. (9) describe the propagation of waves with a dissipation factor $e^{-\sigma t}$, or, what is equivalent, $e^{-\frac{\sigma}{\beta} r}$ (r =distance traveled by the wave). This means that after a distance $r_0 = \beta/\sigma$ the amplitude of a plane wave has decreased by a factor of e^{-1} . Suppressing of edge reflections is now achieved by including a zone with nonvanishing σ along the boundaries of the grid. This zone has to be so large that the amplitudes of waves passing it twice become negligible. In order to avoid additional reflections from the boundary separating the elastic medium with $\sigma=0$ and the dissipating medium with $\sigma>0$ we made σ space dependent and let it increase linearly with the distance from the boundary of the elastic medium.

It should be noted that Eq. (9) implies a specific dissipation function Q proportional to frequency and hence a strong frequency dependence of dissipation. Therefore, it is normally not suited for realistic modeling of dissipative wave propagation in seismic media, but only for the suppression of unphysical reflections.

The finite difference approximation of Eq. (9) in a heterogeneous medium is obtained in the same way as shown before for the elastic wave equation. The difference equation corresponding to Eq. (7) becomes

$$\begin{aligned} v_{m,n}^{p+1} = & (\sigma_{m,n} \Delta t + 1)^{-1} \left\{ (\sigma_{m,n} \Delta t - 1) v_{m,n}^{p-1} + 2 \left(\frac{\beta_{m,n} \Delta t}{h} \right)^2 \right. \\ & [M_1 (v_{m+1,n}^p - v_{m,n}^p) - M_2 (v_{m,n}^p - v_{m-1,n}^p) \\ & + M_3 (v_{m,n+1}^p - v_{m,n}^p) - M_4 (v_{m,n}^p - v_{m,n-1}^p)] \\ & \left. + 2[1 - \sigma_{m,n} \Delta t] v_{m,n}^p \right\} \quad (10) \end{aligned}$$

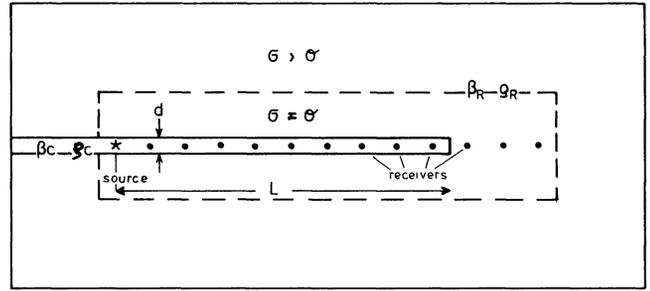


Fig. 3. Geometry of the computational model. The dashed line indicates the boundary between the elastic medium ($\sigma=0$) and the dissipating medium ($\sigma>0$). The grid size was 250×80 grid points. The grid spacing h was 0.5 m

With this method the amplitudes of the edge reflections are attenuated to about 3% of the incident amplitudes, if the zone with slowly increasing σ is about $25-30h$ thick.

Results

Reflection and Transmission of Love Waves at the End of a Coal Seam

The first model which we will study in detail is a coal seam with thickness $d=2.5$ m, shear velocity $\beta_c=1$ km/s and density $\rho_c=1.5$ g/cm³, embedded in homogeneous rock with $\beta_R=2$ km/s and $\rho_R=3$ g/cm³. A line source is located in the middle of the seam. After the horizontal distance $L=20.5d$ from the source the seam ends with a discontinuity perpendicular to the bedding plane. The geometry of the computational model is given in Figure 3. The source radiates a pulse $s(t) = \sin(2\pi t/T_s) - \frac{1}{2} \sin(4\pi t/T_s)$ for $0 < t < T_s$. T_s is the pulse duration.

Figure 4 shows seismogram sections for two different pulse lengths T_s . The receivers are located at the same depth as the source and at various horizontal distances x from the source. In Fig. 4a the dominant wavelength in the seam is four times the seam thickness d , in Fig. 4b it is reduced to $2.5d$. The dispersion of the direct seam wave is clearly seen, resulting in an Airy-phase with large amplitudes in the second example.

Having passed the discontinuity at $x=L$, the wave propagates in the rock as a body wave without further dispersion. The reflected seam wave shows strong frequency dependence. In Fig. 4a, only a weak reflection can be seen, whereas for the higher frequencies in Fig. 4b the reflected amplitudes become fairly large, especially in that part of the seismogram containing the Airy-phase. A reflection coefficient R and a transmission coefficient T , which are defined as follows, show this in more detail. R (T) is the maximum peak-to-peak amplitude of the reflected (transmitted) wave, measured at a distance $x=0.5L$ ($x=1.5L$). Both R and T are normalized with respect to the amplitude of the wave coming directly from the source, measured at $x=0.5L$. Figure 5 shows R and T as a function of the dominant wavelength of the direct wave. Obviously R increases for higher frequencies, while T dominates at low frequencies. Waves which are very long compared to the seam thickness, are not reflected at all. Figure 5 gives rough information about the frequency range that should be used for the optimal detection of discon-

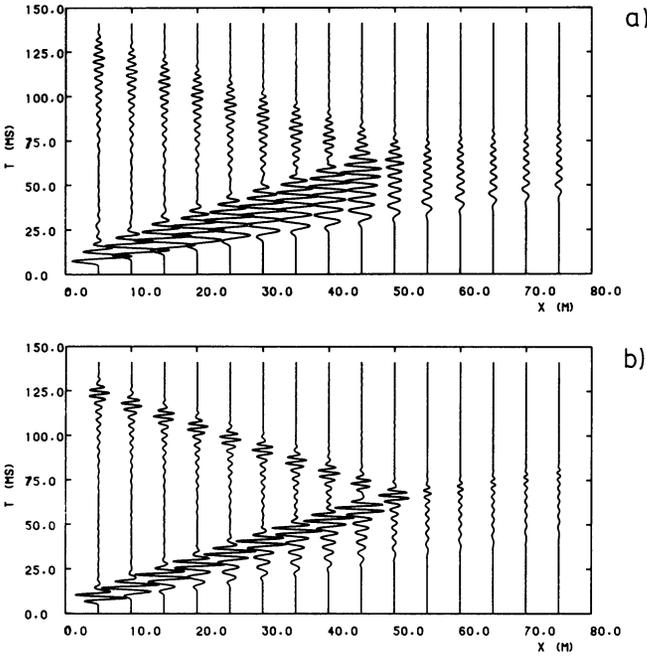


Fig. 4a, b. Seismogram sections for the model of Fig. 3. The duration T_s of the source pulse is a 10 ms and b 6.25 ms. $L = 50$ m

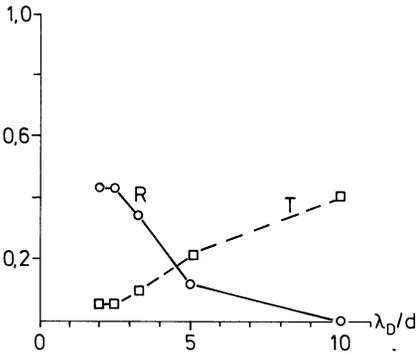


Fig. 5. Reflection coefficient R and transmission coefficient T of the seam end (model of Fig. 3) as defined in the text versus the dominant wavelength λ_D in the seam

tinuities. In this case, wavelengths of up to $2-3d$ produce maximal reflection and minimal transmission.

In order to get more detailed information about the frequency dependence of the reflection and transmission process, we compute the reflectivity function $r(\omega)$ and the transmissivity function $t(\omega)$ from the amplitude spectra of the three waves taken at the points mentioned above. $r(\omega)$ and $t(\omega)$ are shown in Fig. 6, $r(\omega)$ behaving like a high-pass filter. It approaches zero for low frequencies and goes up to a maximum of about 0.55 for high frequencies. The transmissivity $t(\omega)$, in contrast, drops to a small but constant value at high frequencies.

In addition to these numerical results we derive in the following a simple analytical approximation for the reflectivity $r(\omega)$. It shows on which parameters the reflection process mainly depends.

Figure 7 gives the amplitude distribution $A(z)$ of a plane harmonic fundamental-mode Love wave in a layer between two identical homogeneous half-spaces for various frequencies. $z=0$ denotes the middle of the layer and d is its thickness. $A(z)$ is given by

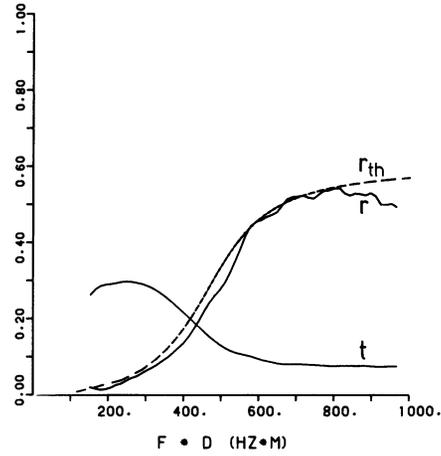


Fig. 6. Reflectivity r together with its analytical approximation r_{th} and transmissivity t versus fd for the model of a seam end

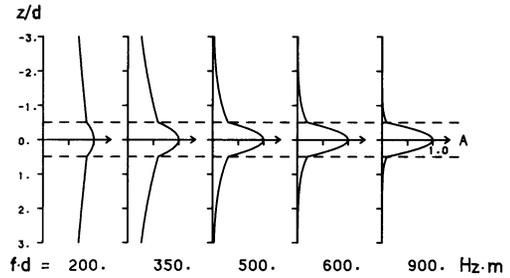


Fig. 7. Amplitude distribution for plane harmonic Love waves (fundamental mode) in a layer between two identical halfspaces for various frequencies

$$A(z) = \begin{cases} A_0 \cos(k\gamma_1 z) & |z| \leq d/2 \\ A_0 \cos(k\gamma_1 d/2) \exp[-k\gamma_2(|z| - d/2)] & |z| > d/2 \end{cases} \quad (11)$$

with

$$\gamma_1 = (c^2/\beta_c^2 - 1)^{1/2}; \quad \gamma_2 = (1 - c^2/\beta_R^2)^{1/2}$$

$c(\omega)$ is the phase velocity and k is the wave number. For higher frequencies, the amplitudes become more and more concentrated within the layer, as shown in Fig. 7. We now assume that only that part of the wave propagating inside the layer is reflected at the discontinuity with the reflection coefficient for plane waves $r_0 = (\rho_c \beta_c - \rho_R \beta_R) / (\rho_c \beta_c + \rho_R \beta_R)$.

Moreover, the amplitudes in the neighbouring rock are supposed to be not affected by the discontinuity. Under this assumption, an approximation for the reflectivity $r(\omega)$ can be written as

$$r_{th}(\omega) = r_0 \cdot \frac{\int_{-d/2}^{+d/2} A(z) dz}{\int_{-\infty}^{+\infty} A(z) dz} \quad (12)$$

Inserting Eq. (11) into Eq. (12) and making use of the dispersion relation for Love waves in a layer of thickness $d/2$ over a halfspace,

$$\tan \left[\omega \frac{d}{2} \cdot (\beta_c^{-2} - c^{-2})^{1/2} \right] = \mu_R \gamma_2 / \mu_c \gamma_1$$

the final result is

$$r_{th}(\omega) = r_0 \left/ \left(1 + \frac{\mu_c (\beta_c^{-2} - c^{-2})}{\mu_R (c^{-2} - \beta_R^{-2})} \right) \right. \quad (13)$$

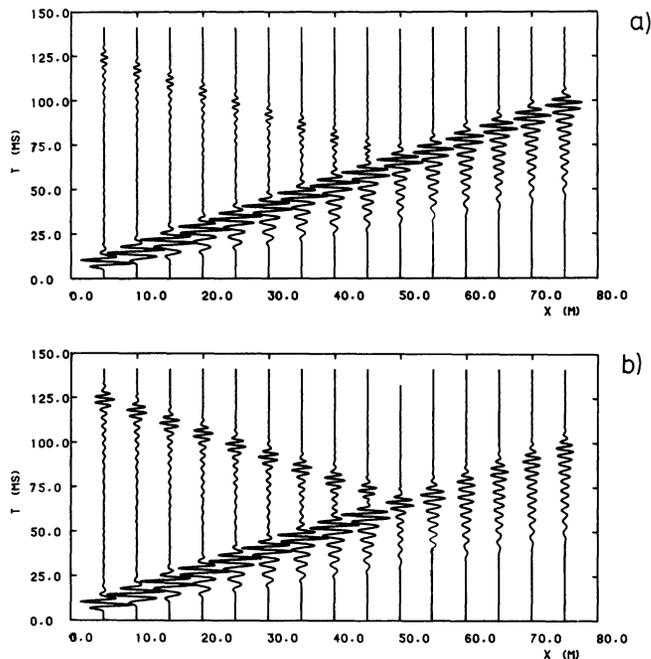


Fig. 8a, b. Seismogram sections for the model with a vertical seam offset b at $x=50$ m. The duration of the source pulse is 6.25 ms. Seam thickness $d=2.5$ m. **a** $b=0.4d$, **b** $b=d$

$r_{th}(\omega)$ is shown in Fig. 6. It agrees with the numerically obtained result to within a few per cent over the whole range of frequencies. From this we may conclude that the amplitude of the reflected Love wave depends mainly on the amplitude partition of the incident wave between rock and coal, and that other effects, such as diffractions at the seam corners, are not important in this case.

Horizontal and Vertical Offsets

In this section we will discuss the reflection and transmission of Love waves at more realistic seam discontinuities. The intrusion of a sand channel, for example, may be simulated by a horizontal offset. Similarly, a tectonic fault may result in a vertical offset of the seam. We compute seismogram sections for different values of the horizontal offset a and vertical offset b . From the seismograms the reflection and transmission coefficients R and T and the reflectivity and transmissivity functions $r(\omega)$ and $t(\omega)$ are obtained as described before. The elastic parameters and the seam thickness are the same as in the previous section. The source time function is the same in all cases with a dominant wavelength $\lambda_D = \beta_c \cdot T_s = 2.5d$.

In Fig. 8, two seismogram sections with different values of the vertical offset b can be compared: $b=0.4d$ and $b=d$, respectively. In the first case the reflected seam wave is weak, and the transmitted wave is nearly unaffected by the discontinuity. In the second example, however, the reflected wave is quite strong, whereas the transmitted wave is relatively weak, particularly in that part of the seismograms dominated by the high-frequency Airy-phase. From this example we may conclude that offsets of less than one half of the seam thickness can barely be detected by seam wave observations.

The reflection and transmission coefficients for both horizontal and vertical offsets are drawn in Fig. 9. (We do not show seismogram sections for horizontal offsets, because they

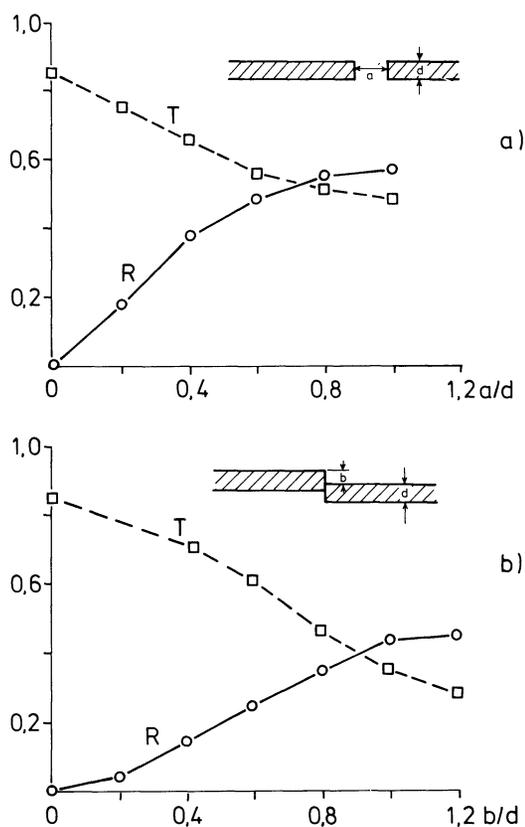


Fig. 9a, b. Reflection coefficient R and transmission coefficient T versus **a** the horizontal offset a with $b=0$ and **b** the vertical offset b with $a=0$

look very similar to those in Fig. 8). For horizontal offsets the reflection coefficients are always greater than for the corresponding vertical offsets. The transmission coefficients are not so different in the two cases. Only for offsets of about one seam thickness or more, are the transmission coefficients for a vertical offset significantly smaller than those for a horizontal offset. Apparently, reflected seam waves are more sensitive to discontinuities with a horizontal offset than to discontinuities with a vertical offset. Suppose a reflection coefficient of about 0.5 is regarded as sufficient for a clear identification of a reflected seam wave, then a horizontal offset of about $0.5d$ could be detected, whereas a vertical offset of $1d$ would be needed to produce the same reflection coefficient. The seam wave transmitted through the discontinuity, however, is more attenuated in the case of a vertical offset, if the offset is at least $1d$. Therefore a large vertical offset may be more easily detected with the aid of transmission measurements.

Finally we look at the reflectivity and transmissivity functions for some special cases. Figure 10 gives $r(\omega)$ and $t(\omega)$ for the seismogram sections with a vertical offset already shown in Fig. 8. For the large offset of one seam thickness in Fig. 10b the reflectivity grows with frequency up to a maximum of about 0.6, while the transmissivity drops sharply for frequencies of more than $450/d$ Hz. This is due to the fact that for higher frequencies the amplitudes of the incident wave become more and more concentrated within the seam and therefore a higher percentage of the incident energy is reflected at the discontinuity. If the vertical offset is less than $1d$, as in Fig. 10a, the two parts of the seam overlap at the discontinuity and the incident energy propagating inside the

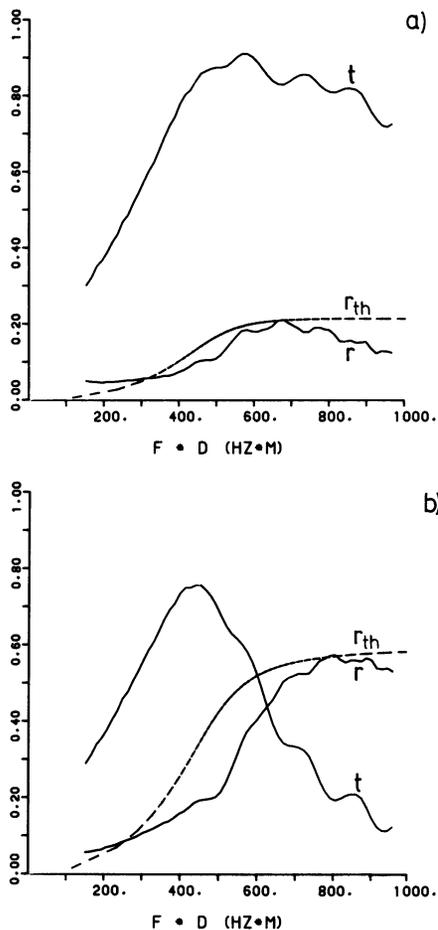


Fig. 10a, b. Reflectivity r together with its analytical approximation r_{th} and transmissivity t versus fd for the model with a vertical seam offset **a** $b=0.4d$ and **b** $b=d$

seam is only partially reflected. Therefore the reflectivity is small over the whole range of frequencies and the transmissivity is only slightly reduced at high frequencies.

For the models with a horizontal offset we obtained reflectivity and transmissivity functions which differ only slightly from those for vertical offsets. The general feature in all cases is an increase of the reflectivity and a decrease of the transmissivity with frequency. As a consequence it seems not to be possible to use spectral analysis of the reflected and transmitted seam waves to distinguish between different types of discontinuities.

For the models with a vertical offset it is possible to compute the analytical approximation $r_{th}(\omega)$ for the reflectivity function. r_{th} is obtained according to Eq. (12) with the difference that integration over $A(z)$ in the numerator is carried out not from $-d/2$ to $+d/2$ but over that depth range, where a discontinuity is actually present in the special model. The approximation r_{th} is additionally given in Fig. 10. It follows that even for these rather complicated models r_{th} predicts correctly the order of magnitude of the reflectivity.

Summary and Conclusions

An explicit finite difference scheme has been presented for the solution of the equation of motion in heterogeneous media, which has the following advantages compared to commonly used schemes:

(1) The boundary conditions at interfaces are all approximated with a truncation error of second order. Furthermore, this scheme is developed into a heterogeneous formulation with the same accuracy.

(2) A damping mechanism is included which attenuates waves approaching the edges of the computational grid and thus almost completely prevents edge reflections.

With this scheme the propagation of Love waves in coal seams with discontinuities was calculated. We restricted ourselves to some simple discontinuities such as horizontal and vertical offsets, although it is possible to treat more complicated models with the same finite difference scheme.

The reflection and transmission process at a discontinuity depends strongly on the frequency range of the incident Love wave. As a general result, it was found that low frequencies are mainly transmitted and high frequencies are mainly reflected. This is explained qualitatively by the amplitude partition of Love waves between rock and coal as a function of frequency. For the vertical offset we derive an analytical approximation for the reflectivity as a function of frequency, which agrees satisfactorily with the numerical results.

The study predicts the amount of the offset which may be detected for a given detection level and a given dominant wavelength. The best results are obtained with dominant wavelengths up to about 3 seam thicknesses. The reflectivity and transmissivity functions differ slightly for different types of discontinuities, but probably not enough to allow for a discrimination in practice.

Acknowledgements. We thank G. Müller for helpful comments and for reading the manuscript and I. Hörnchen for typing it. The computations were performed at the computing centers of the Universities of Karlsruhe and Frankfurt.

References

- Alterman, Z., Rotenberg, A.: Seismic waves in a quarter plane. Bull. Seismol. Soc. Am. **59**, 347–368, 1969
- Alterman, Z., Loewenthal, D.: Computer generated seismograms. In: Methods in computational physics, Vol. 12, B. Bolt, B. Alder, S. Fernbach, M. Rotenberg, eds. New York: Academic Press 1972
- Boore, D.M.: Love waves in nonuniform waveguides: finite difference calculations. J. Geophys. Res. **75**, 1512–1527, 1970
- Boore, D.M.: Finite difference methods for seismic wave propagation in heterogeneous materials. In: Methods in computational physics, Vol. 11, B. Alder, S. Fernbach, M. Rotenberg, eds. New York: Academic Press 1972
- Clayton, R., Engquist, B.: Absorbing boundary conditions for acoustic and elastic wave equations. Bull. Seismol. Soc. Am. **67**, 1529–1540, 1977
- Dresen, L., Freystätter, S.: Rayleigh channel waves for the in-seam seismic detection of discontinuities. J. Geophys. **42**, 111–129, 1976
- Freystätter, S., Dresen, L.: The influence of oblique-dipping discontinuities on the use of Rayleigh channel waves for the in-seam seismic reflection method. Geophys. Prosp. **26**, 1–15, 1978
- Kelly, K.R., Ward, R.W., Treitel, S., Alford, R.M.: Synthetic seismograms: a finite difference approach. Geophysics **41**, 2–27, 1976
- Krey, T.C.: Channel waves as a tool of applied geophysics in coal mining. Geophysics **28**, 701–714, 1963
- Mitchell, A.R.: Computational methods in partial differential equations. New York: J. Wiley 1969
- Reynolds, A.C.: Boundary conditions for the numerical solution of wave propagation problems. Geophysics **43**, 1099–1110, 1978
- Smith, W.D.: A nonreflecting plane boundary for wave propagation problems. J. Comp. Phys. **15**, 492–508, 1974

Received September 16, 1981; Revised version October 26, 1981

Accepted October 27, 1981

Numerical Simulation of Vertical Seismic Profiling

P. Temme and G. Müller

Institute of Meteorology and Geophysics, Feldbergstr. 47, 6000 Frankfurt, Federal Republic of Germany

Abstract. Methods for the calculation of synthetic seismograms along vertical profiles through horizontally layered media are presented; the aim is to simulate numerically the technique of vertical seismic profiling (VSP), which is sometimes used for improved resolution of deep primary reflections. We present the theory for vertically travelling plane waves and for spherical waves, generated by a single force or an explosive point source. The solution is restricted to the acoustic case. The theoretical seismograms are complete, i.e., they contain all surface and internal multiples. Anelastic effects are taken into account with the method of complex velocities. In order to save computing time, seismogram segments are calculated, using the method of complex frequencies.

These methods are applied to a layered model of the coal-bearing carboniferous in the Ruhr district of W. Germany. Different assumptions on overburden structure, recording geometry and anelasticity are made, and the VSP seismograms are compared with seismograms calculated for horizontal profiles at the surface. It is suggested that VSP field surveys be preceded by similar numerical studies for an estimate of the information that can be obtained by this expensive technique.

Key words: Synthetic seismograms — Vertical seismic profiling — Coal prospecting

Introduction

Overburden layers like those in the Ruhr district of the Federal Republic of Germany, have a great influence on seismic prospecting of coal-seam sequences. In the usual line or areal seismic surveys, made at the surface, multiples within the overburden and conversion phases may mask primary reflections from the coal bearing carboniferous, strong frequency dependent absorption changes the excitation signal and the frequency content of reflected waves, passing through the top layers and scattering at inhomogeneities close to the surface results in a reduced signal-to-noise ratio.

These problems, of course, are not restricted to seismic prospecting for coal. A possible reduction of the disturbing influence of the overburden can be obtained by applying vertical seismic profiling (VSP), where receivers are arranged vertically in boreholes. Wuenschel (1976) and Kennett et al. (1980) describe several advantages of this recording geometry, including improvement of signal-to-noise ratio, higher band-

width, especially better recording of high frequencies, better identification of primary and multiple reflections due to separation of up- and downgoing waves, determination of deconvolution operators from the direct downward travelling wave, better knowledge of the absorption properties of the overburden.

Because VSP is an especially expensive technique, it is desirable to perform numerical simulations of this method prior to the measurements, using all information on the reflecting structure that is already available. Numerical simulation means the calculation of synthetic seismograms for receivers on profiles extending vertically through a layered model of the ground from the surface to arbitrary depth. The synthetic seismograms should be realistic, i.e., they should include all internal and surface multiples, complicated layering should be tractable, and the field geometry and absorption of seismic energy in the layers should be taken into account. Under these circumstances the numerical simulation of VSP may make a valuable contribution to the design of field experiments. The purpose of this paper is to present the theoretical background for the calculation of synthetic VSP seismograms and to illustrate their potential by examples for a model which is typical of the layering in the Ruhr district.

The theoretical approach presented here is restricted to the acoustic case. We present the theory for vertically propagating plane waves and for spherical waves in some detail, since we believe that these methods deserve wider distribution among interpreters and researchers in reflection seismics, at least the methods for the acoustic case, where really complicated layering with arbitrary elastic and anelastic properties can be treated at reasonable computational cost. Of course, the limitation to the acoustic case yields only an approximation of the true elastic wavefield, especially at larger horizontal distances between source and receivers, and Kennett (1979) has emphasized this point. However, Kennett's seismogram calculations for both the acoustic and the elastic case (which includes $P-SV$ interaction) illustrate, in our opinion, that from a practical point of view, the acoustic approximation is often sufficient. The theory, which is presented here with mainly the modelling of vertical seismic profiles in mind, can also be applied to horizontal profiles; the differences between the corresponding computer programs are small.

In the following we start with the theory for vertical plane-wave propagation in a layered half space. Very recently, papers on this special subject have appeared (Wyatt 1981; Ganley 1981), and our method is, in fact, similar to that used by Ganley; we therefore present only a compact outline. We

then show how calculated plane-wave seismograms can subsequently be corrected for spherical divergence in order to get an approximation of point-source excitation. Exact theoretical seismograms for explosive point sources as well as for vertical single forces at the surface can be obtained from the theory developed in the appendix. A few remarks are added on the implementation of absorption effects by using complex velocities and on the calculation of seismogram segments with the method of complex frequencies, a method by which aliasing in the time domain may be suppressed. Synthetic VSP seismogram sections are then presented for a model of the carboniferous, typical of the Ruhr district, overlain by two different tertiary overburden structures. Different assumptions on the recording geometry and Q structure of the models are made. We finally compare the synthetic VSP sections with horizontal profile sections which simulate the usual line-survey technique at the surface.

Theory

Synthetic Seismograms for Plane-Wave Excitation

We assume a model of $n-1$ layers on top of a half space, which is layer n (Fig. 1). Each layer is described by its parameters d_i =thickness, α_i = P -wave velocity, ρ_i =density, and Q_i =quality factor. At the free surface $z_1=0$ a seismic source, exciting vertically travelling plane waves, is located. We describe the incident wave by its potential $\phi_0(z, t)$

$$\phi_0(z, t) = F\left(t - \frac{z}{\alpha_1}\right).$$

For many purposes it is easier to work in the frequency domain. The Fourier transform of ϕ_0 is given by

$$\bar{\phi}_0(z, \omega) = \bar{F}(\omega) e^{-j\omega \frac{z}{\alpha_1}},$$

where $\bar{F}(\omega)$ is the Fourier transform of the excitation function $F(t)$. For the potential $\bar{\phi}_i(z, \omega)$ in layer i we assume similarly

$$\bar{\phi}_i(z, \omega) = \bar{F}(\omega) \{A_i e^{-j\omega \frac{z-z_i}{\alpha_i}} + B_i e^{+j\omega \frac{z-z_i}{\alpha_i}}\} \quad (1)$$

with so far unknown coefficients A_i and B_i . Downward travelling waves are represented by the first term of (1), upward

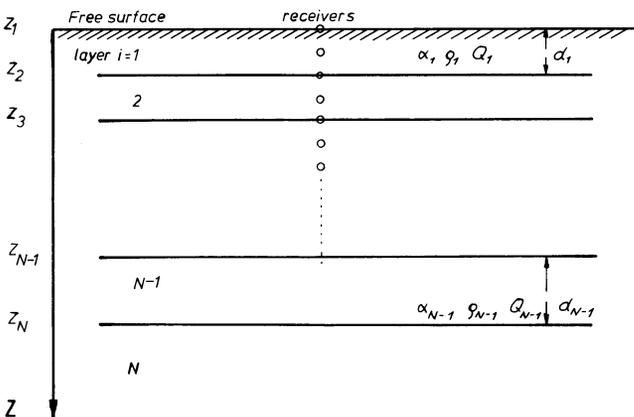


Fig. 1. The general model for computing synthetic seismograms. Circles indicate receiver positions. The parameters of layer i are: α_i — P velocity, ρ_i density, Q_i quality factor, d_i layer thickness. z_i ($i=1, 2, \dots, n$) is the depth of interface i

travelling waves by the second. The amplitude coefficients A_i and B_i are calculated from the boundary conditions at the interfaces (continuity of vertical displacement $\partial\phi/\partial z$ and normal stress $\rho\partial^2\phi/\partial t^2$):

$$\bar{\phi}'_i = 0 \quad (\text{free surface at } z=z_1). \quad (2)$$

$$\frac{\partial\bar{\phi}_i}{\partial z} = \frac{\partial\bar{\phi}_{i-1}}{\partial z} \quad (\text{at } z=z_i \text{ (} i=2, 3, \dots, n)) \quad (3)$$

$$\rho_i\bar{\phi}_i = \rho_{i-1}\bar{\phi}_{i-1}$$

$\bar{\phi}'_i = \bar{\phi}'_1 - \bar{\phi}'_0$ in Eq. (2) represents the potential of all upgoing and downgoing waves in layer 1 except for the incident wave $\bar{\phi}'_0$. If we assume an incident wave of unit amplitude, Eq. (2) results in

$$A_1 = 1 - B_1.$$

Equations (3) give for A_i, B_i in layer $i > 1$ the matrix relation

$$\begin{pmatrix} A_i \\ B_i \end{pmatrix} = m_i \begin{pmatrix} A_{i-1} \\ B_{i-1} \end{pmatrix} \quad (4)$$

with the 2×2 layer matrix

$$m_i = \frac{e^{j\omega d_{i-1}/\alpha_{i-1}}}{2\rho_i/\alpha_i} \begin{pmatrix} (\rho_i/\alpha_{i-1} + \rho_{i-1}/\alpha_i) e^{-2j\omega d_{i-1}/\alpha_{i-1}} & (-\rho_i/\alpha_{i-1} + \rho_{i-1}/\alpha_i) \\ (-\rho_i/\alpha_{i-1} + \rho_{i-1}/\alpha_i) e^{-2j\omega d_{i-1}/\alpha_{i-1}} & (\rho_i/\alpha_{i-1} + \rho_{i-1}/\alpha_i) \end{pmatrix}. \quad (5)$$

Successive applications of Eq. (4) and additional use of the relation $B_n=0$, which means that there are no upgoing waves in the halfspace (index n), yields

$$\begin{pmatrix} A_n \\ 0 \end{pmatrix} = M \begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = M \begin{pmatrix} 1 - B_1 \\ B_1 \end{pmatrix} \quad (6)$$

with the layer matrix product

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = m_n \cdot m_{n-1} \cdot \dots \cdot m_2.$$

Solution of Eq. (6) yields

$$B_1 = \frac{M_{21}}{M_{21} - M_{22}} = \frac{R}{1 + R} \quad (7)$$

and

$$A_1 = 1 - B_1 = \frac{1}{1 + R},$$

where we have introduced the reflectivity $R = -M_{21}/M_{22}$ of the layer stack below $z=z_2$. As shown in the appendix, the reflectivity R can also be calculated by a recursive procedure; this calculation is faster than the calculation via the matrix M .

The amplitude coefficients for $i=2, 3, \dots, n$ are determined by the matrix relation

$$\begin{pmatrix} A_i \\ B_i \end{pmatrix} = M_i \begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = M_i \begin{pmatrix} 1 - B_1 \\ B_1 \end{pmatrix}, \quad (8)$$

where the layer matrix product M_i only has to be taken down to layer i :

$$M_i = \begin{pmatrix} M_{i11} & M_{i12} \\ M_{i21} & M_{i22} \end{pmatrix} = m_i \cdot m_{i-1} \cdot \dots \cdot m_2.$$

From Eqs. (7) and (8) we obtain:

$$A_i = \frac{M_{i11} + RM_{i12}}{1 + R}$$

$$B_i = \frac{M_{i21} + RM_{i22}}{1 + R}.$$
(9)

Finally, in order to obtain the displacement spectrum, we have to take the derivative of the potential $\bar{\phi}_i$ in Eq. (1) with respect to z . After inverse Fourier transform we find the displacement in layer i ,

$$u_i(z, t) = \frac{1}{2\pi} \frac{\alpha_1}{\alpha_i} \int_{-\infty}^{+\infty} \bar{G}(\omega) \{A_i e^{-j\omega \frac{z-z_1}{\alpha_i}} - B_i e^{+j\omega \frac{z-z_1}{\alpha_i}}\} e^{j\omega t} d\omega,$$
(10)

where $\bar{G}(\omega) = -\bar{F}(\omega) \frac{j\omega}{\alpha_1}$ is the spectrum of the displacement of the incident wave at $z_1 = 0$.

Equation (10) allows fast wavefield calculations for arbitrary receiver depth in the layered medium. The seismograms are complete, i.e., all internal and surface multiples are included.

Synthetic Seismograms for Point-Source Excitation

In realistic field experiments point sources which generate spherical waves are used. Thus amplitudes of reflected waves are not only affected by reflection coefficients and absorption, but also by spherical divergence. In the appendix the theory of point source excitation dealing with the explosive and vertical single force case is presented. In principle the Sommerfeld integral is used to decompose the direct spherical wave into plane waves, and the corresponding plane wave responses of the medium are superposed to give the spherical-wave response. Because typically several hundred plane-wave cases have to be solved, the computing time is much longer than that needed for the evaluation of Eq. (10).

In the following we present an approximation of the spherical-wave case, which can easily be applied to seismograms calculated from the plane-wave theory. Thus we can simulate VSP in the special case where receivers are buried *vertically beneath the point source*. The method has been described by Müller (1971) for receivers at the surface; the plane-wave seismogram at the arrival times of primary reflections is multiplied by the divergence factors of these reflections and by interpolated values in between. Suppose now, a receiver is buried at depth z in layer i . Thus, the primary reflection from interface k ($k = i + 1, i + 2, \dots, n$) has the travel-time

$$t_k = \sum_{j=1}^{k-1} \frac{2d_j}{\alpha_j} - \sum_{j=1}^{i-1} \frac{d_j}{\alpha_j} - \frac{z-z_i}{\alpha_i}.$$

The divergence factor by which the plane-wave seismogram is multiplied at this time is

$$G_k = \frac{\alpha_1}{\sum_{j=1}^{k-1} 2d_j \alpha_j - \sum_{j=1}^{i-1} d_j \alpha_j - (z-z_i) \alpha_i}.$$

In the case of receivers at depth, the direct wave has to be corrected accordingly:

$$t_1 = \sum_{j=1}^{i-1} \frac{d_j}{\alpha_j} + \frac{z-z_i}{\alpha_i},$$

$$G_1 = \frac{\alpha_1}{\sum_{j=1}^{i-1} d_j \alpha_j + (z-z_i) \alpha_i}.$$

Between the times $t_1, t_{i+1}, t_{i+2}, \dots, t_n$ we apply a linear interpolation, and for $t > t_n$ the seismogram is damped according to $G = G_n \cdot t_n/t$. The direct wave and primary reflections are corrected exactly by this procedure; it turns out that multiple reflections are suppressed too much. However, the comparison with seismograms calculated by the exact spherical theory shows that this simple method often is an acceptable approximation.

Causal Absorption

When synthetic seismograms are used to simulate wave propagation in realistic media, the computational methods should take account of absorption, particularly at the high frequencies used in coal-seam prospecting where absorption strongly influences the waves. Anelasticity can be incorporated into the methods described in this paper by making the wave velocities complex. Like O'Neill and Hill (1979) we use the following complex velocity law:

$$\alpha(\omega) = \alpha_r \left(1 + \frac{1}{\pi Q} \ln \frac{\omega}{\omega_r} + \frac{j}{2Q} \right).$$
(11)

Here Q is the quality factor, assumed to be frequency independent, ω_r is a reference frequency and α_r is the (real) velocity at the reference frequency, assumed to be known. In calculations for seismic prospecting one would choose $f_r = \omega_r/2\pi$ roughly around 100 Hz. The velocity law (11) corresponds to causal absorption and hence implies a slight velocity dispersion.

Aliasing in the Time Domain

In numerical calculations of theoretical seismograms the seismogram length strongly influences the computing time, since usually the spectra are calculated for equidistant frequencies and the frequency interval is the reciprocal value of this length. For complicated models, especially those having a free surface, the duration of the complete seismogram can be very long, and the number of frequencies in a specified frequency window is proportional to duration. If we choose a time window shorter than the response of the medium in order to save computing time, those parts of the response beyond the end of the seismogram will not be suppressed, but appear in the early parts of the seismogram and interfere with early arrivals. This situation is completely analogous to aliasing in the frequency domain and therefore is called aliasing in the time domain.

An effective method to avoid or reduce aliasing in the time domain has been in use in seismology for quite some time (Rosenbaum 1974; Bouchon 1979). The basic idea is to calculate first the *damped* seismogram $u_\tau(t) = u(t) e^{-t/\tau}$ instead of $u(t)$ by evaluating the Fourier transform $\bar{u}(\omega)$ of $u(t)$ at complex frequencies $\omega - j/\tau$ and using the damping theorem of Fourier transforms: $\bar{u}_\tau(\omega) = \bar{u}(\omega - j/\tau)$. Evidently, $u_\tau(t)$ is less disturbed by aliasing than $u(t)$. Multiplication of $u_\tau(t)$ by $e^{+t/\tau}$ in a second step gives the desired seismogram $u(t)$. Experiments with this method, with τ equal 20–40% of the desired length of $u(t)$, show good results.

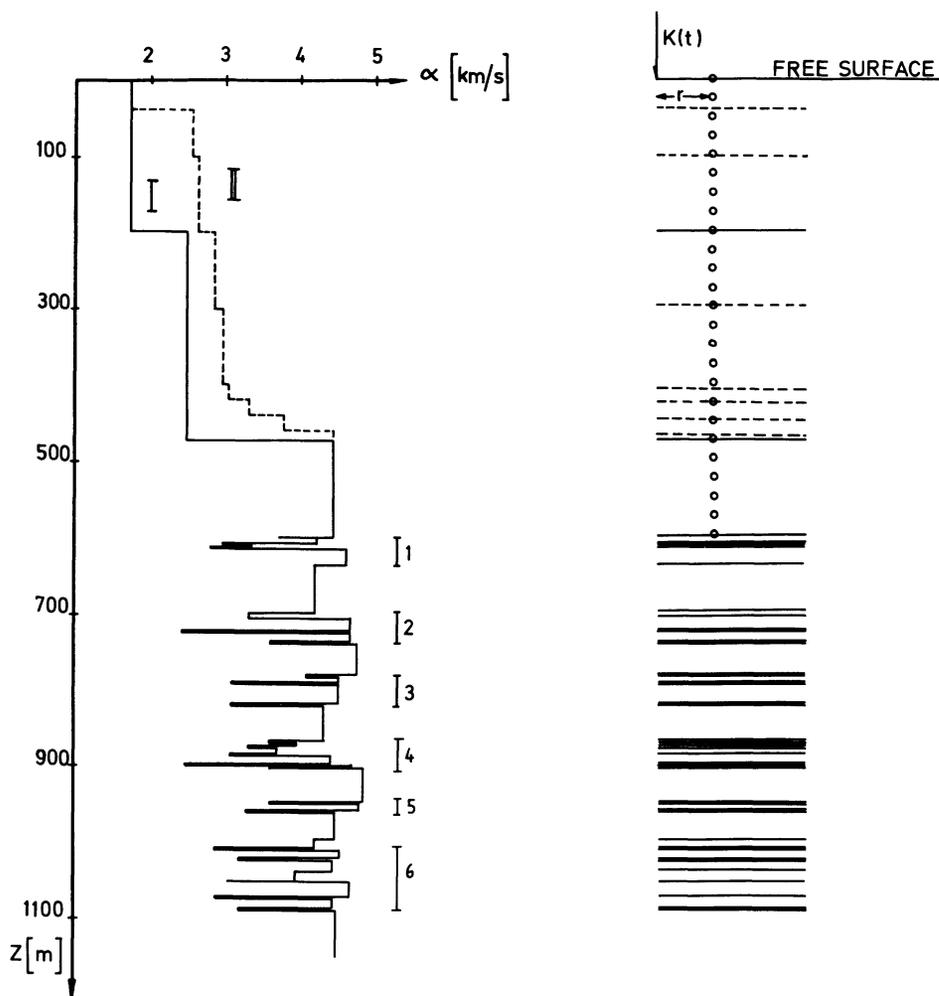


Fig. 2. *Left:* A typical velocity-depth model from the Ruhr district, W. Germany. The carboniferous is subdivided into 6 seam groups, indicated by vertical bars and combined with two different overburden models, a simple 3-layer overburden (model I) and a complicated 9-layer overburden (model II, dashed line). *Right:* Corresponding layer sequences; circles indicate vertical seismic profile $K(t)$, single force; r , offset from source to receiver array

Synthetic Seismograms

In the following we present synthetic seismograms, calculated for the complicated layering shown in Figure 2. A continuous borehole profile from the Ruhr district in W. Germany was approximated by a discretized velocity-depth function, where the coal-seam sequence extends in the depth range 600–1100 m, showing strong velocity inversions with a maximum reflection coefficient of 0.4 at depth 900 m. The density of each layer was calculated from the relation $\rho = 1.7 + 0.2\alpha$. We subdivided the seam sequence into 6 seam groups, denoted by vertical bars. The carboniferous was combined with two different types of overburden structure. Model I in Fig. 2 has a simple 3-layer overburden with strong reflection coefficients, whereas in model II there is a more gradual transition from the surface to the carboniferous, consisting of 9 layers. The receivers are arranged on a vertical array extending from the surface to the level 600 m, with a depth interval of 25 m. A seismic source, exciting vertically travelling plane waves or spherical waves, is located at the free surface. In the spherical-wave case the source is a vertical single force, and the receivers may have an arbitrary offset r .

Model I

Figure 3 shows the calculated VSP seismograms for plane-wave excitation with a dominant frequency of 100 Hz in the case of relatively weak absorption ($Q=1000$ in all layers). In

order to save computing time, the response, which has a duration of nearly 8 s, is only calculated for a time segment of 1 s using the method of complex frequencies described above; it took a computing time of about 2.5 min on a DEC1091 to calculate this section. Downgoing waves with positive velocities are clearly separated in Fig. 3 from upcoming waves with negative velocities. The first arrival in each seismogram is the direct, downward travelling wave. The interfaces of the 3-layer overburden can be identified at the depths z_2 and z_3 , where the direct wave coincides with primary reflections from the overburden. At the surface the amplitudes of reflections are doubled. From there the upcoming waves are reflected back as surface multiples into the structure. Within each layer, the amplitudes remain constant, as we expect for plane waves in elastic media. At interfaces the transmitted waves are affected only by the transmission coefficient, primary reflections from the seam groups, indicated by bars in Fig. 2, are affected only by the transmission coefficient. Primary reflections from the groups 1, 2 and 3 can be identified clearly. Each reflection is a complicated interference signal. The deepest traces, in the depth range 425–600 m, contain seam-group reflections, which are undisturbed by overburden multiples.

Figure 4 demonstrates the influence of the frequency content of the excitation signal. We have now assumed a dominant frequency of 50 Hz; all other parameters are the same as in the calculation for Fig. 3. This seismic section shows

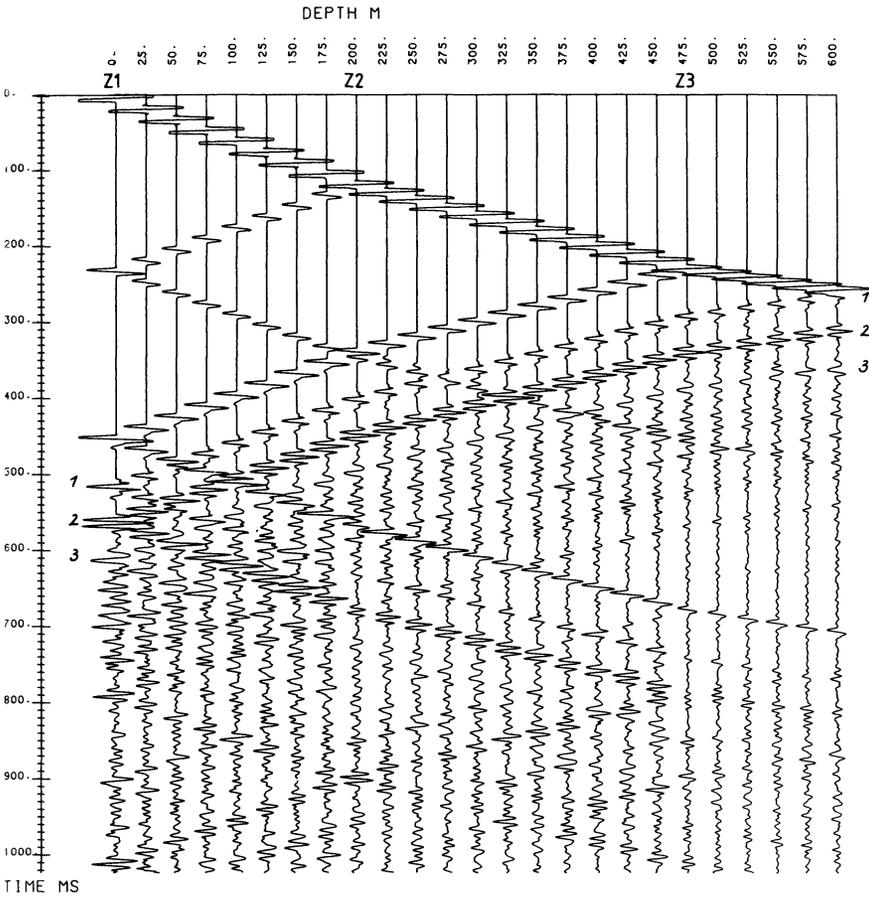


Fig. 3. VSP section for model I and plane-wave excitation with a dominant frequency of 100 Hz; seismograms are normalized to the maximum of the deepest trace. z_1 , z_2 and z_3 indicate overburden interfaces. Seam-group reflections are indicated by 1, 2, 3

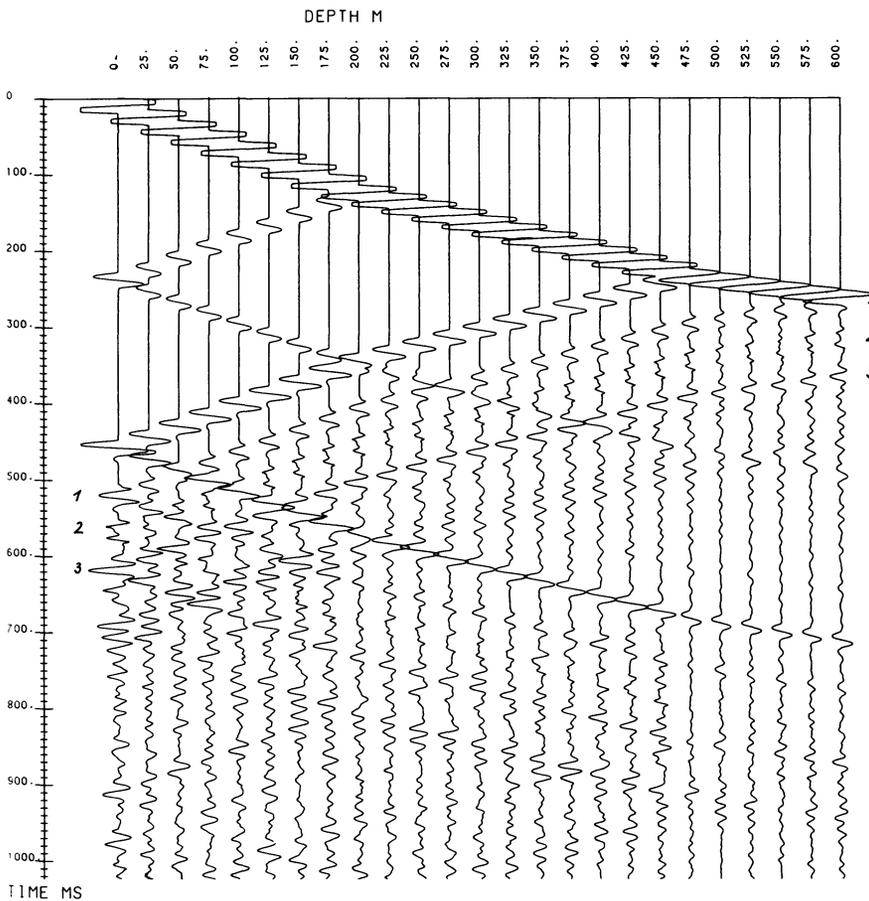


Fig. 4. VSP section for model I and plane-wave excitation with a dominant frequency of 50 Hz

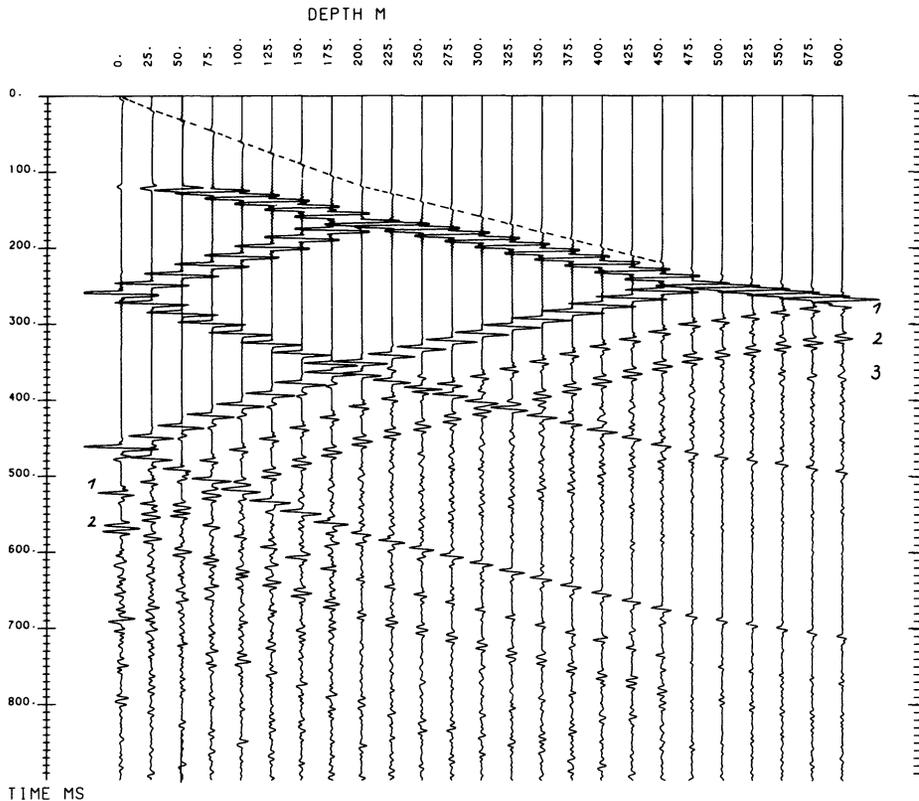


Fig. 5. VSP section for model I and spherical-wave excitation at the surface; offset $r=200$ m. The *dashed line* indicates a numerical phase without physical significance. Seam-group reflections are indicated by 1, 2, 3

strongly modified primary reflections from the seam groups. The amplitude ratio between them is changed, particularly in that the reflection from seam group 2 becomes much weaker relative to the reflections 1 and 3. Thus, a shift in dominant frequency of seismic waves used for coal prospecting may strongly enhance or reduce a particular seam-group reflection. Broadband recording and subsequent bandpass filtering in different frequency bands may therefore be quite helpful for the identification of seam groups.

If we use a vertical single force at the free surface, which excites spherical waves with again 100 Hz dominant frequency, we obtain the VSP synthetic seismograms in Fig. 5, calculated for an offset of $r=200$ m. Compared to the plane-wave case in Fig. 3 the traveltime curves of all phases are now curved, and the amplitudes of up- and downgoing waves are no longer constant within the layers. Relative to the direct wave, the primary reflections from the seam groups become weaker. Only seam groups 1 and 2 can be identified clearly and show nearly the same interference signals as in Fig. 3; the seam group reflection 3 is much weaker. The computing time to calculate this section was about 2 h.

The theory of spherical waves allows also the calculation of synthetic seismograms on a horizontal profile. For receivers located on a line extending from 100–1100 m, we obtain the section shown in Fig. 6. The first arrival D is the strongly attenuated direct wave which travels along the top of the first layer. The primary reflection R_1 from the first overburden interface arrives at 230 ms. Beyond the critical distance of about 600 m, we then record the refracted wave from the second layer as a first arrival. At 440 ms the second primary reflection R_2 from depth level 475 m arrives, followed immediately by a multiple reflection of R_1 . Within the time window of 500 ms to 650 ms, we can identify the primary reflections from the seam groups 2 and 3. Beyond a distance

of 500 m they interfere strongly with multiples and primary reflections from the overburden. Compared to the VSP seismograms in Figs. 3 and 5, multiple events cannot easily be separated from primary reflections of the seam groups.

Model II

In order to study the influence of multiple reflections and adsorption in the overburden, VSP seismograms for a more complicated layering (model II) were calculated. Figure 7 shows the synthetic section for $Q=1000$ everywhere and for spherical-wave simulation, as described above. The direct wave D is followed by a strong multiple M of the reflection from the first overburden interface at depth 35 m. The primary reflections from the seam groups 1, 2 and 3 stand out clearly, but they begin to interfere with the multiple reflection M in the depth range below 500 m. As already mentioned, the spherical-wave approximation corrects exactly the primary reflections only, but attenuates multiple reflections too much. Thus, one could expect a small influence of multiples on seam reflections in the depth range 150–400 m.

If we want to model strong anelastic effects in the overburden by assuming $Q=100$, we obtain the section shown in Fig. 8. The amplitudes of the seismograms are now normalized with respect to the maximum of the seismogram at the depth 600 m in Fig. 7; thus seismograms of Figs. 7 and 8 are directly comparable. Wave amplitudes are strongly attenuated due to anelastic effects, and the accompanying dispersion can be seen. Primary reflections from the seam groups are clearly visible in the depth range 300–600 m only. The disturbing influence of overburden multiples on primary reflections from the seam sequence is reduced.

Figure 9 shows the synthetic seismograms for the exact spherical-wave case; they were calculated for an offset of $r=200$ m and $Q=1000$ throughout model II. A weak, non-

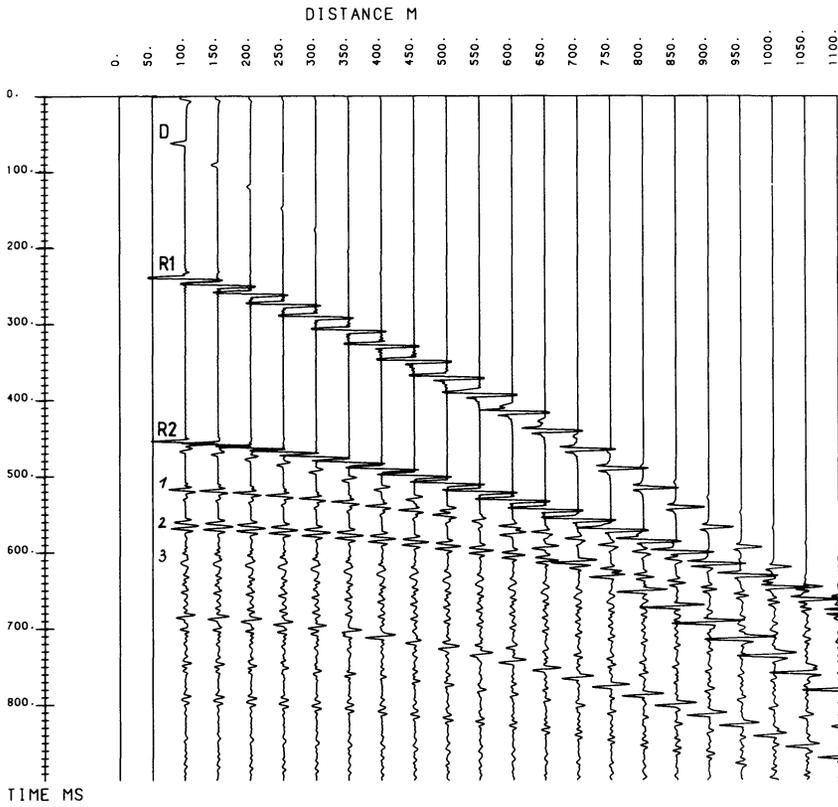


Fig. 6. Horizontal-profile section for spherical-wave excitation at the surface of model I, seismograms are normalized to the maximum of the most distant trace. *D*, direct wave; *R1*, reflection from the first overburden interface; *R2*, reflection from the second overburden interface. Seam-group reflections indicated by 1, 2, 3

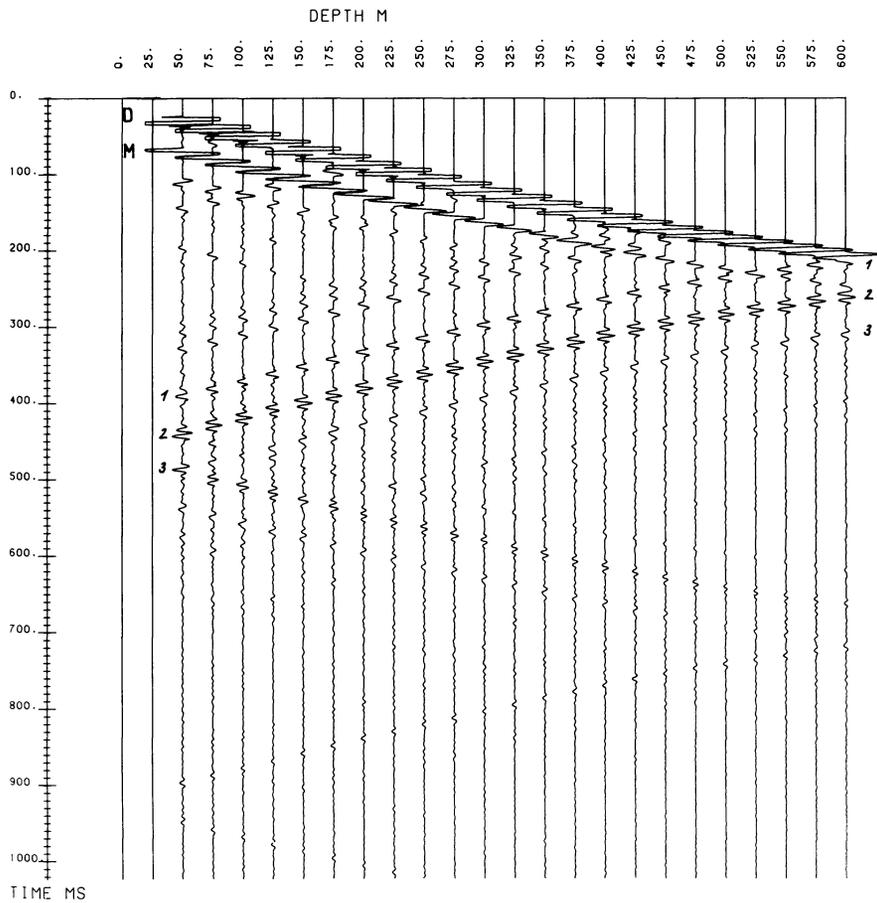


Fig. 7. VSP section for model II and spherical-wave simulation with offset $r=0$; $Q=1000$ everywhere. *D*, direct wave; *M*, surface multiple from the first overburden interface. Traces at 0 and 25 m are set zero for plotting reasons

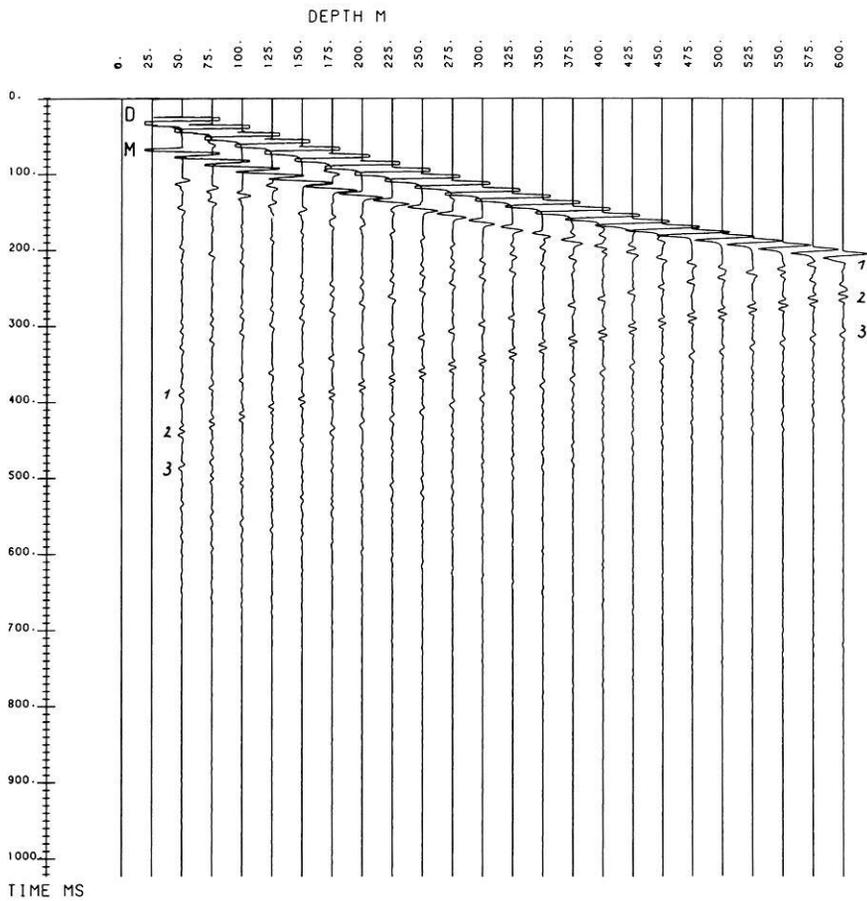


Fig. 8. VSP section for model II and spherical-wave simulation with offset $r = 0$; $Q = 100$ throughout the overburden and $Q = 1000$ elsewhere. Section is normalized with respect to the maximum of the deepest trace in Fig. 7

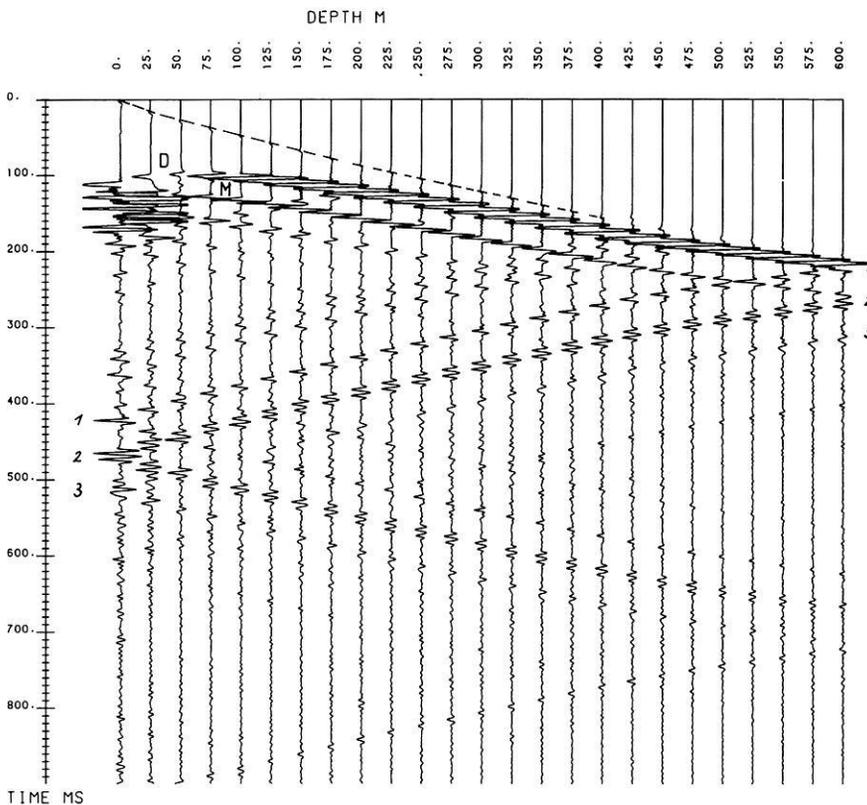


Fig. 9. VSP section for model II and spherical-wave excitation by a single force at the surface; offset $r = 200$ m, D , direct wave; M , surface multiple from the first overburden interface. The dashed line indicates a numerical phase

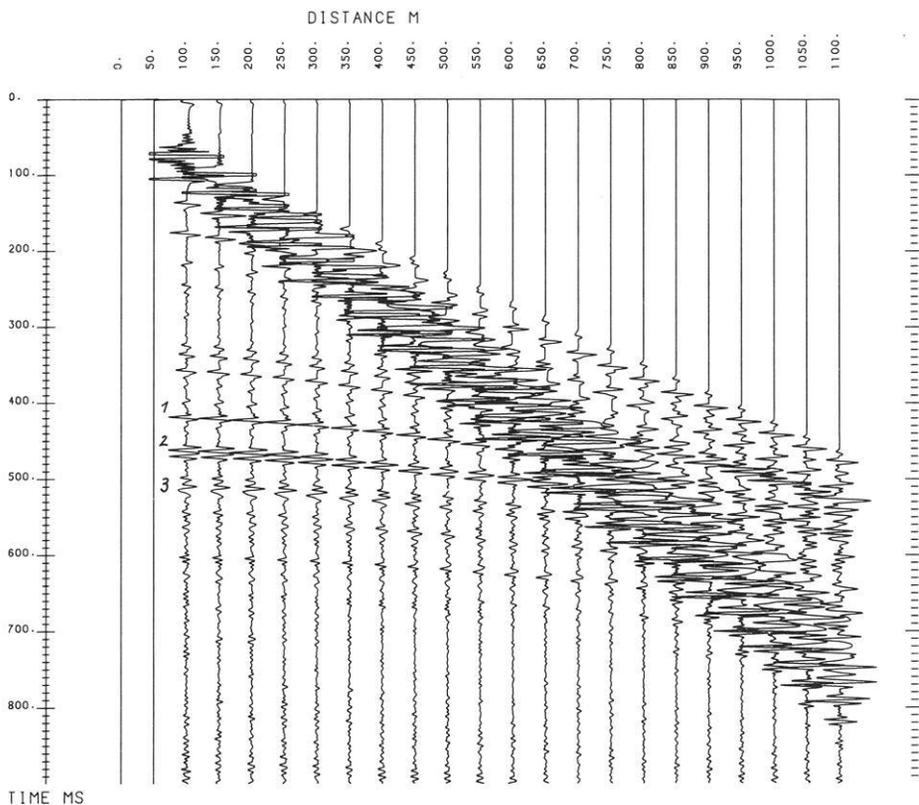


Fig. 10. Horizontal-profile section at the surface of model II for spherical-wave excitation (single force). Seismograms are normalized with respect to the maximum of the most distant trace

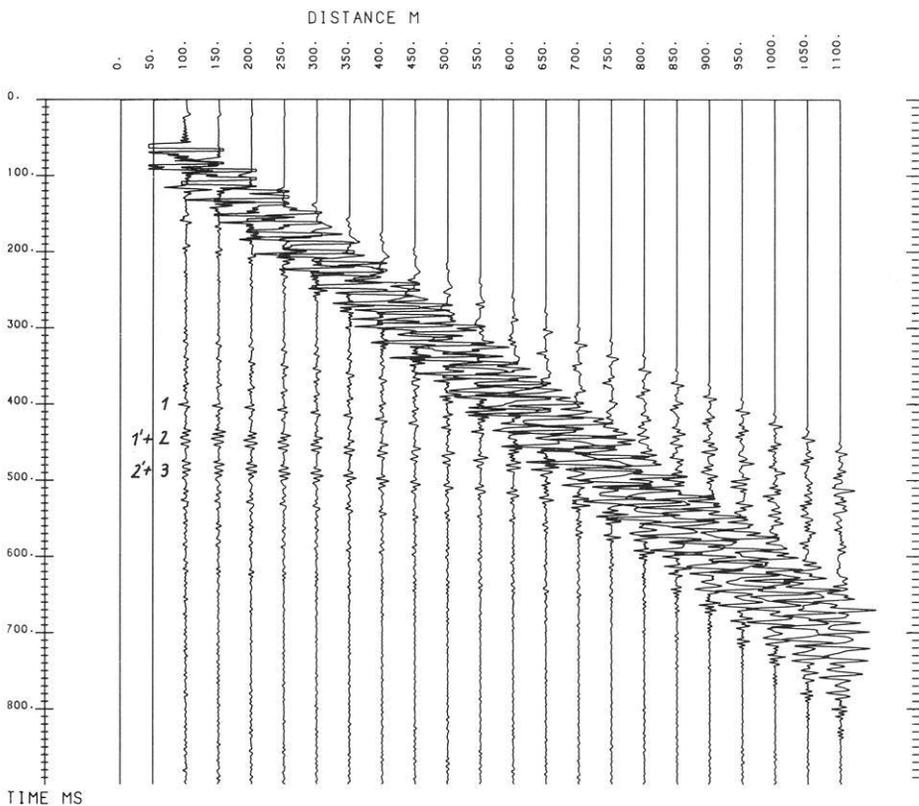


Fig. 11. Horizontal-profile section at the surface of model II for spherical-wave excitation (explosion source at depth 30 m). Seismograms are normalized with respect to the maximum of the most distant trace. The main band of energy in this section is stronger than in the section of Fig. 10 because horizontal radiation is stronger for an explosion than for a vertical single force. Seam-group reflections 1' and 2' are due to the ghost reflection at the source

disturbing numerical arrival is followed by the direct downgoing wave *D*; the first arrival at the depths 0 and 25 m is the refracted wave from the first overburden interface. Within the first layer supercritical reflections produce strong amplitudes in the time interval 100–200 ms. Many features in

Fig. 9 are quite similar to those in Fig. 7, except that multiple phases (which are correctly modelled in Fig. 9) are considerably stronger.

The synthetic seismogram section for a horizontal profile at the surface of model II is shown in Fig. 10; the correspond-

ing section for model I was given in Fig. 6. Strong amplitudes of multiple reflections in the first layer are visible in the whole distance range from 100–1100 m and cancel primary seam-group reflections beyond a distance of 500 m. Multiple reflections within the whole overburden disturb these reflections at shorter distances more than in the VSP seismograms of Fig. 9. For instance, only a small portion of energy in the wave group 3 in Fig. 10 is really caused by the seam-group reflection 3. Hence, it appears that the disturbing effects of a complicated overburden are more pronounced on horizontal than on vertical seismic profiles.

Our final example of synthetic seismograms (Fig. 11) corresponds to an explosive point source at a depth of 30 m in model II and to a horizontal profile at the surface. Now the surface reflection at the source (ghost reflection) entails doubling of all arrivals in comparison with Fig. 10, such that a really complicated section results. For instance, the seam-group reflections 1' and 2', which are a consequence of the ghost reflection, interfere with the direct seam-group reflections 2 and 3, respectively. We mention in passing that the computing time for Fig. 11 was much reduced in comparison with that for Fig. 10, since the reflectivity of model II, of course, had not to be recalculated.

Conclusions

With the method described above it is possible to calculate realistic seismograms on vertical as well as on horizontal seismic profiles through complicated layering at reasonable computational cost. The case of vertical seismic profiling can be modelled very quickly by using simple plane-wave theory and correcting approximately for spherical divergence. In this way the effects due to multiple reflections and absorption can be estimated, and some idea can be obtained about the depths at which receivers should be buried and what frequencies should be used in a field experiment, provided that some information on the structure is available in advance. The more sophisticated spherical wave theory yields additional information on the influence of the offset between source and receiver array. It also allows the simulation of the usual horizontal-profile observations at the surface and thus helps in deciding whether or not the expensive VSP-technique can really be expected to improve the identification of primary and multiple reflections and to reduce significantly the influence of absorption. In the case studied in this paper it appears that this is actually so, but this result cannot be generalized. Hence, a planned field survey should probably be preceded by numerical studies with the methods developed above.

Acknowledgements. This work was supported financially by Ruhrkohle AG and the government of the state of Nordrhein-Westfalen, Federal Republic of Germany. The computations were performed at the computing center, University of Frankfurt. We are grateful to R. Kind for drawing our attention to the possibility of reducing time-domain aliasing by the use of complex frequencies and for discussions on the subject of this paper. Thanks for discussions go also to J. Fertig. We thank I. Hörnchen for typing the manuscript.

Appendix. Synthetic Seismograms for Point-Source Excitation

We treat the case of an explosive point source in the first layer of a horizontally layered medium. This case simulates reasonably well the excitation of seismic waves by explosions and similar sources in

reflection measurements on land or sea. For simplicity we consider only acoustic waves. The geometry of the medium is the same as in Fig. 1; the source depth is z_s with $0 < z_s < z_2$, and the receiver can be at any depth and horizontal offset. We work with displacement potentials and use cylindrical coordinates (r, φ, z) . Because of symmetry reasons there is independence of the azimuth φ ; r is the horizontal source-receiver distance.

For the Fourier transformed potential of the direct wave in layer 1 we use the Sommerfeld integral

$$\bar{\phi}_0 = \bar{F}(\omega) \int_0^\infty \frac{k}{j l_1} J_0(kr) e^{-j l_1 |z - z_s|} dk, \quad (\text{A-1})$$

where $\bar{F}(\omega)$ is the spectrum of the excitation function, k the horizontal wavenumber, $l_1 = (\omega^2/\alpha_1^2 - k^2)^{1/2}$ the vertical wavenumber, and $J_0(kr)$ the Bessel function of first kind and order zero.

Similar representations are sought for the potential in an arbitrary layer with index i :

$$\begin{aligned} \bar{\phi}_i = \bar{F}(\omega) \int_0^\infty \frac{k}{j l_i} J_0(kr) \\ \cdot \{A_i e^{-j l_i (z - z_i)} + B_i e^{+j l_i (z - z_i)}\} dk \\ l_i = (\omega^2/\alpha_i^2 - k^2)^{1/2} \quad (l_i > 0 \text{ or } \text{Im} l_i < 0). \end{aligned} \quad (\text{A-2})$$

This integral solves the wave equation in layer i . As in the plane-wave case with normal incidence, treated in the main part of the paper, the first term in the curly brackets of (A-2) represents all downgoing waves in layer i and the second term all upgoing waves, and the amplitude coefficients A_i and B_i are calculated from the boundary conditions at the interfaces (continuity of normal displacement and normal stress):

$$\left. \begin{aligned} \frac{\partial \bar{\phi}_i}{\partial z} = \frac{\partial \bar{\phi}_{i-1}}{\partial z} \\ \rho_i \bar{\phi}_i = \rho_{i-1} \bar{\phi}_{i-1} \end{aligned} \right\} \text{at } z = z_i \quad (i = 2, 3, \dots, n) \quad (\text{A-3})$$

$$\bar{\phi}_1 = 0 \quad \text{at } z = z_1 = 0. \quad (\text{A-4})$$

Equation (A-3) yields the matrix relation

$$\begin{pmatrix} A_i \\ B_i \end{pmatrix} = m_i \begin{pmatrix} A_{i-1} \\ B_{i-1} \end{pmatrix} \quad (\text{A-5})$$

with the 2×2 layer matrix

$$m_i = \frac{e^{+j l_i - 1 d_{i-1}}}{2 \rho_i l_i} \begin{pmatrix} (\rho_i l_{i-1} + \rho_{i-1} l_i) e^{-2 j l_i - 1 d_{i-1}} (-\rho_i l_{i-1} + \rho_{i-1} l_i) \\ (-\rho_i l_{i-1} + \rho_{i-1} l_i) e^{-2 j l_i - 1 d_{i-1}} (\rho_i l_{i-1} + \rho_{i-1} l_i) \end{pmatrix}. \quad (\text{A-6})$$

The layer matrix in the main part, Eq. (5), follows from (A-6) for $k = 0$.

Layer 1 requires a special treatment, because the source potential $\bar{\phi}_0$ in (A-1) has to be included in $\bar{\phi}_1$. Above the source ($z < z_s$) we use the denotation $\bar{\phi}_1 = \bar{\phi}_1^+$ with amplitude coefficients A_1^+ and B_1^+ , and below the source ($z > z_s$) we have the quantities $\bar{\phi}_1^-$, A_1^- and B_1^- . The relation between the two pairs of coefficients is

$$\begin{pmatrix} A_1^- \\ B_1^- \end{pmatrix} = \begin{pmatrix} A_1^+ \\ B_1^+ \end{pmatrix} + \begin{pmatrix} A_s \\ -1/A_s \end{pmatrix}, \quad \text{where } A_s = e^{+j l_1 z_s}. \quad (\text{A-7})$$

We then apply (A-5) successively, use additionally the relations $B_n = 0$ (no upgoing wave in the half-space with index n) and $B_1^+ = -A_1^+$ (from boundary condition (A-4)), and introduce (A-7):

$$\begin{pmatrix} A_n \\ 0 \end{pmatrix} = M \begin{pmatrix} A_1^- \\ B_1^- \end{pmatrix} = M \begin{pmatrix} A_1^+ \\ -A_1^+ \end{pmatrix} + M \begin{pmatrix} A_s \\ -1/A_s \end{pmatrix} \quad (\text{A-8})$$

with the layer-matrix product

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = m_n \cdot m_{n-1} \cdot \dots \cdot m_2. \quad (\text{A-9})$$

The second of the two equations (A-8) can be solved for A_1^+ :

$$A_1^+ = \frac{M_{21}A_s - M_{22}/A_s}{M_{22} - M_{21}}, \quad B_1^+ = -A_1^+. \quad (\text{A-10})$$

Introducing these results into (A-7) yields:

$$A_1^- = \frac{M_{22}}{M_{22} - M_{21}}(A_s - 1/A_s), \quad B_1^- = RA_1^-. \quad (\text{A-11})$$

The amplitude coefficients for $i=2, 3, \dots, n$ follow from the matrix relation

$$\begin{pmatrix} A_i \\ B_i \end{pmatrix} = M_i \begin{pmatrix} A_{i-1}^- \\ B_{i-1}^- \end{pmatrix},$$

where

$$M_i = \begin{pmatrix} M_{i11} & M_{i12} \\ M_{i21} & M_{i22} \end{pmatrix} = m_i \cdot m_{i-1} \cdot \dots \cdot m_2 \quad (\text{A-12})$$

is the layer-matrix product taken down to layer i . Using (A-11) we have:

$$A_i = \frac{M_{i11}M_{22} - M_{i12}M_{21}}{M_{22} - M_{21}}(A_s - 1/A_s),$$

$$B_i = \frac{M_{i21}M_{22} - M_{i22}M_{21}}{M_{22} - M_{21}}(A_s - 1/A_s). \quad (\text{A-13})$$

A computational advantage, both with respect to accuracy and computing time, is obtained by observing that the only quantity, related to matrix M and needed in (A-10), (A-11) and (A-13), is the plane-wave reflectivity $R = -M_{21}/M_{22}$ of the layer stack below layer 1:

$$A_1^+ = -\frac{RA_s + 1/A_s}{1 + R}, \quad B_1^+ = -A_1^+, \quad (\text{A-14})$$

$$A_1^- = \frac{A_s - 1/A_s}{1 + R}, \quad B_1^- = RA_1^-, \quad (\text{A-15})$$

$$A_i = \frac{M_{i11} + RM_{i12}}{1 + R}(A_s - 1/A_s),$$

$$B_i = \frac{M_{i21} + RM_{i22}}{1 + R}(A_s - 1/A_s) \quad (i=2, 3, \dots, n). \quad (\text{A-16})$$

The reflectivity R can be computed recursively (Brekhovskikh 1960). For this purpose we define the reflectivity with respect to the depth $z = z_i$ by the ratio $R_i = B_i/A_i$ of the amplitudes of all upgoing and all downgoing waves in layer i . Then we find the relation between R_{i-1} and R_i from (A-5):

$$R_{i-1} = \frac{r_i + R_i}{1 + r_i R_i} e^{-2jl_{i-1}d_{i-1}}$$

$$r_i = \frac{\rho_i l_{i-1} - \rho_{i-1} l_i}{\rho_i l_{i-1} + \rho_{i-1} l_i}$$

$$= \text{plane-wave reflection coefficient of interface } z = z_i. \quad (\text{A-17})$$

The recursion is started with $i=n$, $R_n=0$, and is stopped at $i=2$: the result R_1 of the last step in the recursion is the desired reflectivity R . The reflectivity for vertically travelling waves, which is needed in Eqs. (7) and (9) of the main paper, follows from (A-17) by setting $l_k = \omega/\alpha_k$.

Formulas (A-14) to (A-16), together with (A-2) allow wavefield calculations for arbitrary receiver depths and distances from the source in the layered medium. For applications it is better to work in terms of integrals over slowness $u = k/\omega$ rather than integrals over wavenumber k . The vertical wavenumbers transform according to $l_i = \omega q_i$, where $q_i = (\alpha_i^{-2} - u^2)^{1/2}$. The spectra of the horizontal and vertical displacements, $\bar{u}_{ri} = \partial \bar{\phi}_i / \partial r$ and $\bar{u}_{zi} = \partial \bar{\phi}_i / \partial z$, respectively, are:

$$\left. \begin{aligned} \bar{u}_{ri} &= -\bar{F}(\omega) \omega^2 \int_0^\infty \frac{u^2}{j q_1} J_1(u\omega r) \{A_i e^{-j\omega q_i(z-z_i)} + B_i e^{+j\omega q_i(z-z_i)}\} du \\ \bar{u}_{zi} &= -\bar{F}(\omega) \omega^2 \int_0^\infty \frac{u q_i}{q_1} J_0(u\omega r) \{A_i e^{-j\omega q_i(z-z_i)} - B_i e^{+j\omega q_i(z-z_i)}\} du \end{aligned} \right\} \quad (i=1, 2, \dots, n). \quad (\text{A-18})$$

We add a few brief remarks on some aspects of the numerical calculation of the integrals in (A-18):

1) The integration is performed by the trapezoidal rule and is restricted to the slowness interval of interest. Normally this is the interval from 0 to $1/\alpha_1$. The number of slownesses is typically 100 to several hundreds. A cosine taper at the largest slownesses may be useful to reduce the amplitudes of numerical arrivals without physical significance.

2) Anelasticity of the layers is taken into account by working with complex wave velocities according to Eq. (11) of the main paper. Besides the fact that this is a desirable feature in principle, especially when the high-frequency waves of seismic prospecting are modelled, it is also computationally favorable since it removes the surface-waves poles of the integrands away from the real u -axis and thus away from the integration path.

3) Time-domain aliasing because of very long physical responses can be avoided or reduced by using complex instead of real frequencies, as described in the main part.

4) The Bessel functions are replaced by the asymptotic form of the corresponding Hankel functions of the second kind, implying only waves propagating away from the source. Numerical tests showed that only for very accurate calculations at small distances should the original Bessel functions be retained. This may be connected with the occurrence of relatively strong numerical arrivals.

5) Fast Fourier Transform methods are used for going from the time domain to the frequency domain and vice versa.

Additional information on the numerical calculation of integrals of the form (A-18) can be found in papers by Fuchs and Müller (1971), Fertig and Müller (1978) and Kennett (1979; 1980). We conclude this appendix with the displacement spectra for a *vertical single force at the surface* of the layered medium:

$$\left. \begin{aligned} \bar{u}_{ri} &= -\frac{\bar{K}(\omega)\omega}{4\pi\rho_1} \int_0^\infty u^2 J_1(u\omega r) \{A_i e^{-j\omega q_i(z-z_i)} + B_i e^{+j\omega q_i(z-z_i)}\} du \\ \bar{u}_{zi} &= -\frac{\bar{K}(\omega)\omega}{4\pi\rho_1} \int_0^\infty j u q_i J_0(u\omega r) \{A_i e^{-j\omega q_i(z-z_i)} - B_i e^{+j\omega q_i(z-z_i)}\} du \end{aligned} \right\} \quad (i=1, 2, \dots, n). \quad (\text{A-19})$$

where

$\bar{K}(\omega) = \text{force spectrum,}$

$$A_1 = \frac{2}{1 + R}, \quad B_1 = RA_1,$$

$$A_i = 2 \frac{M_{i11} + RM_{i12}}{1 + R},$$

$$B_i = 2 \frac{M_{i21} + RM_{i22}}{1 + R} \quad (i=2, 3, \dots, n).$$

Formulas (A-19) are a useful supplement to the results (A-18) for an explosive point source, since these vanish for a source location at the surface because of exact cancellation of the direct wave and the surface reflection.

References

- Bouchon, M.: Discrete wavenumber representation of elastic wave fields in three space dimensions. *J. Geophys. Res.* **84**, 3609-3614, 1979
- Brekhovskikh, L.M.: *Waves in layered media*. New York: Academic Press 1960

- Fertig, J., Müller, G.: Computations of synthetic seismograms for coal seams with the reflectivity method. *Geophys. Prospecting* **26**, 868-883, 1978
- Fuchs, K., Müller, G.: Computations of synthetic seismograms with the reflectivity method and comparison with observations. *Geophys. J. R. Astron. Soc.* **23**, 417-433, 1971
- Ganley, D.C.: A method for calculating synthetic seismograms which include the effects of absorption and noise. *Geophysics* **46**, 1100-1107, 1981
- Kennett, B.L.N.: Theoretical reflection seismograms for elastic media. *Geophys. Prospecting* **27**, 301-321, 1979
- Kennett, B.L.N.: Seismic waves in a stratified medium II - theoretical seismograms. *Geophys. J. R. Astron. Soc.* **61**, 1-10, 1980
- Kennett, P., Ireson, R.L., Conn. P.J.: Vertical seismic profiles: Their application in exploration geophysics. *Geophys. Prospecting* **28**, 676-699, 1980
- Müller, G.: Direct inversion of seismic observations. *Z. Geophys.* **37**, 225-235, 1971
- O'Neill, M.E., Hill, D.P.: Causal absorption: its effect on synthetic seismograms computed by the reflectivity method. *Bull. Seismol. Soc. Am.* **69**, 17-25, 1979
- Rosenbaum, J.M.: Synthetic microseisms: logging in porous formations. *Geophysics* **39**, 14-32, 1974
- Wuenschel, P.C.: The vertical array in reflection seismology - some experimental results. *Geophysics* **41**, 219-232, 1976
- Wyatt, K.D.: Synthetic vertical seismic profiles. *Geophysics* **46**, 880-891, 1981

Received September 23, 1981; Revised version December 17, 1981

Accepted December 17, 1981

Energy Dispersion and Acceleration of Low Energy Protons and Their Relation to Electrons During an Auroral Breakup

A. Urban¹, K.M. Torkar¹, and K. Wilhelm²

¹ Space Research Institute of the Austrian Academy of Sciences, c/o Technical University Graz, A-8010 Graz, Inffeldgasse 12, Austria

² Max-Planck-Institut für Aeronomie, D-3411 Katlenburg-Lindau 3, Federal Republic of Germany

Abstract. This paper describes a “proton time-of-flight event” observed onboard a high altitude rocket in the auroral zone during a magnetospheric substorm expansion phase. The onset was defined by a rapid low energy electron influx enhancement. After 36–55 s three successive peaks at 13, 7 and 5.5 keV were measured by the low energy proton detectors and related to the increase of the intense electron precipitation. On the assumption that the spectral variations of the protons were caused by energy dispersion, a distance to the proton acceleration or precipitation region of $\sim 9 R_E$ was evaluated. Generally, the measured differential proton spectra for the flight could be fitted by a power law $F = F_0 E^{-\gamma}$ with a slope of $\gamma \sim 0.7$, except for the event mentioned above and two other events discussed in relation to the simultaneously measured electron precipitation. The pitch angle distribution of the protons was typically isotropic over the upper hemisphere.

Key words: Particle acceleration – Energy dispersion – Mono-energetic protons – Substorm expansion phase

Introduction

The fluxes of low energy protons and electrons in the keV-range show all sorts of different relations, and all combinations of flux and spectral variations can be found in limited regions. The most commonly occurring combinations in this energy range seem to be the correlation in flux variations and anti-correlations in the spectral hardness variation. When the precipitating proton flux increases, its spectrum softens, simultaneously the electron flux also increases and the electron spectrum hardens (Johnstone 1971; Bryant et al. 1977; Edwards et al. 1978; Urban 1981). However, Miller and Whalen (1976) reported the lack of any consistent correlation between electron and proton precipitation. Whalen et al. (1978) came to a similar conclusion stating that no definite relations between proton and electron fluxes exist.

Proton energy spectra with peaks in the keV range were observed by Whalen et al. (1971), Whalen and McDiarmid (1972) and by Miller and Whalen (1976). They attributed the spectral peaks to electrostatic potential drops, but found no clear relation to the electron measurements. Bryant et al. (1977) reported a similar behaviour where a spectral peak developed at ~ 6 keV together with a small proton flux increase without any obvious change in the electron population.

The pitch angle distributions of low energy protons in the keV-range have been found to be rather isotropic or, in some cases, to peak at 90° (Johnstone 1971; Miller and Whalen 1976).

Field aligned proton pitch angle distributions have been observed mainly by means of low-orbiting satellites (Hultqvist 1971; Hultqvist and Borg 1978), but also with rockets (Urban 1981; Brüning et al. 1981). The observed pitch angle distributions peaking at a few tens of degrees above 90° (“conical beams”, Whalen et al. 1978) are not discussed in this paper. Here the data from simultaneous rocket-borne proton and electron measurements in the keV-range observed during an auroral breakup are presented. They show a clear correlation in the proton and electron flux variations.

In particular, the measured proton spectral variations with a sequence of peaks in the energy range 13–5 keV are related to the intense electron precipitation mentioned above and interpreted in terms of energy dispersion of the protons over a distance of about 9 Earth radii.

Instrumentation

The payload F4C was launched on 30 January 1978 at 21:37 UT by a Skylark-12 motor from the Andøya Rocket Range, Norway, and reached an apogee of 530 km. The actual launch azimuth was 346° . The payload had an attitude control system so that the spin axis had an angular deviation of $20^\circ \pm 10^\circ$ with respect to the negative geomagnetic field direction. The spin period was approximately 350 ms.

The complex experimental equipment of the payload allowed particle and field observations in wide energy and frequency ranges. F4C was one of four high altitude sounding rocket payloads of the German IMS campaign “Substorm Phenomena” launched from the Andøya Rocket Range during the time period 13 October 1977–30 January 1978 (Wilhelm 1980).

The proton measurements in the energy range 0.5–30 keV were obtained by 80° spherical electrostatic analysers as energy selectors and channel electron multipliers operating in pulse counting modes as particle detectors. After the energy selection the protons were post-accelerated by a negative potential of 1.3 kV in order to compensate the lower efficiency of the channel electron multiplier for protons at low energies.

The energy spectra in the range 0.5–30 keV were scanned in 19 logarithmic levels with 20 ms sampling time each. A summary of the relevant parameters of the instrument is compiled in Table 1. Three of these proton detectors were mounted at 0° , 20° and 90° with respect to the spin axis. Each sensor was calibrated in order to determine the entrance characteristic, the geometric factor, and the efficiency and energy dependence of the channel electron multipliers. For technical reasons this calibration procedure was performed using a mono-energetic elec-

Table 1. Characteristics of the instruments

		Proton-Spectr.	Electron-Spectr.
Energy range	keV	0.5–30	0.05–25
Resolution $\Delta E/E$	%	7.5	7
Geometric factor	cm ² sr	4×10^{-3}	2×10^{-3}
Field of view	deg ²	8×10	$24 \times 4 \times 6$
Time for one spectrum	s	0.4	46 or 1.5

tron beam with varying acceleration potentials of the beam and horizontal and vertical viewing angles of the sensor. The different efficiencies of the channeltrons for protons compared to electrons were accounted for by use of conversion methods published in the literature (Egidi et al. 1969; Iglesias and McGarity 1971; Crandall and Ray 1975).

The electrons were measured by a differential energy and angle analyser with a field of view of $96^\circ \times 6^\circ$ subdivided into 24 angular sections. The energy selection was performed by two 80° spherical electrostatic analysers. Auxiliary electrostatic deflection devices selected the 24 angular channels. This scheme enabled the experiment to measure directional electron fluxes with a coverage of about 96° in pitch angle range without even utilizing the rocket spin. The instrument operated in slow and fast measurement modes. During the slow mode the viewing direction was controlled by the onboard magnetometer and the selection energy changed slowly in approximately 46 s over the full energy range. During the fast mode the selection energy swept quickly within 1.5 s.

Geophysical Conditions

The payload F4C was launched at 21:37 UT during the expansive phase of a magnetospheric substorm. The K_p value during the time interval from 21–24 UT was -5 . The disturbance to be discussed here was preceded by a 100 nT depression of the X -component of the geomagnetic field (Fig. 1). The activity of the northernmost of the two stable auroral arcs located above Andøya intensified, and later on, at 21:30 UT, a violent breakup occurred, followed by an increase in brightness of the southern arc. Four minutes later the whole sky was covered by auroral displays. The brightness values measured at 21:32 UT, although somewhat uncertain because of a high altitude cloud layer, were 100 kR for $\lambda=557.7$ nm, 3.6 kR for 630.0 nm, and 200 R for 486.1 nm. Unfortunately, an additional low altitude cloud cover caused a rapid deterioration of the visibility immediately after the launch of F4C preventing all sky and photometer observations. Recordings of the geomagnetic variations and pulsations during this event are shown in Figures 1 and 2. The X -component decreased by approximately 500 nT within several minutes concurrently with the fast expansion of the aurora. The large variations of the Y -component were associated with auroral forms directed south-east to north-west.

The geomagnetic pulsations (Fig. 2) showed large amplitudes in both H and D components between 21:30 UT. During the flight of F4C the amplitudes decreased from 30 to about 20 nT. Thereafter the substorm recovered rapidly, interrupted only by some activity increase after 22:30 UT.

Observations

A summary of the low energy proton and electron measurements is given in Figure 3. The different time resolutions apparent

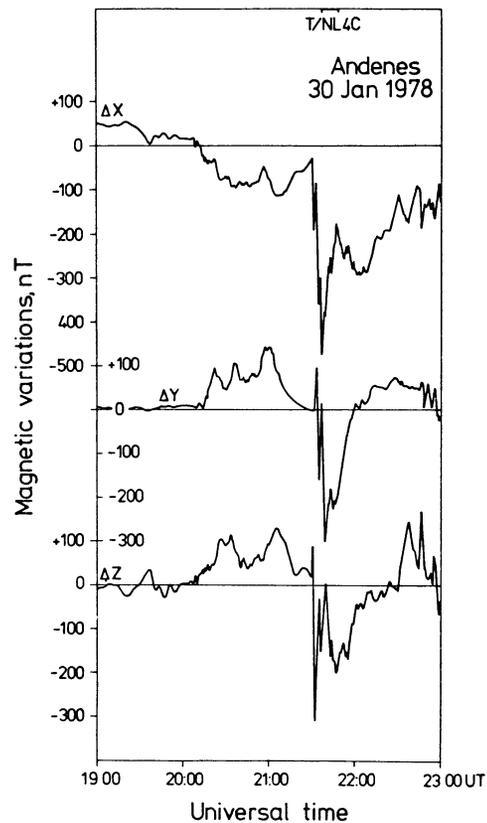


Fig. 1. Variations of the geomagnetic field at Andøya on 30 January 1978. The flight time of the rocket payload T/NL-F4C is indicated at the top of this figure. The variations of the north, east and vertical components are denoted ΔX , ΔY and ΔZ , respectively

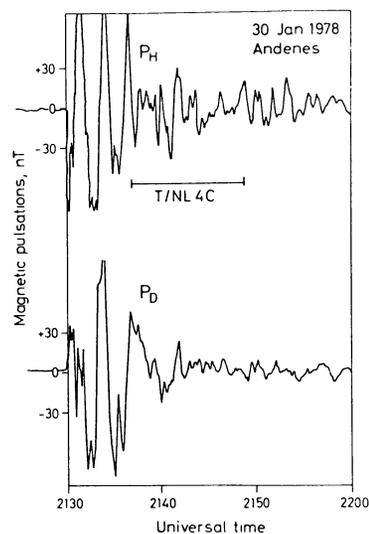


Fig. 2. Horizontal components of the magnetic pulsation activity at Andøya during the flight period of F4C

in the electron fluxes result from the different operating modes of the instrument. Hence the solid line between 160 and 285 s represents three differential energy spectra with high pitch angle resolution, whereas before that time differential energy spectra with high energy resolution have been measured.

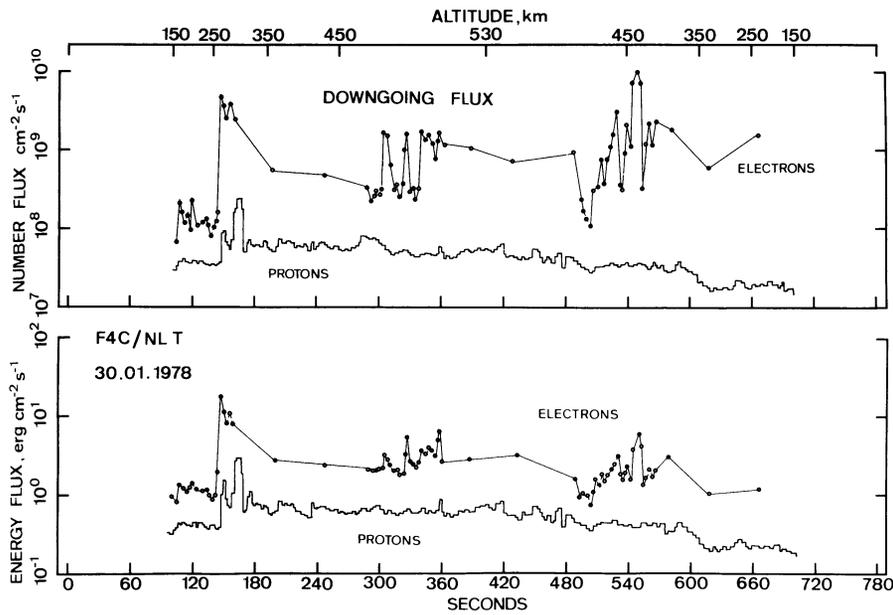


Fig. 3. Downward directed proton and electron energy fluxes and number fluxes during the flight F4C. The variable time resolution apparent in the electron fluxes results from different operating modes of the instrument. The observed proton spectral variations are related to the sudden flux increase at 151 s

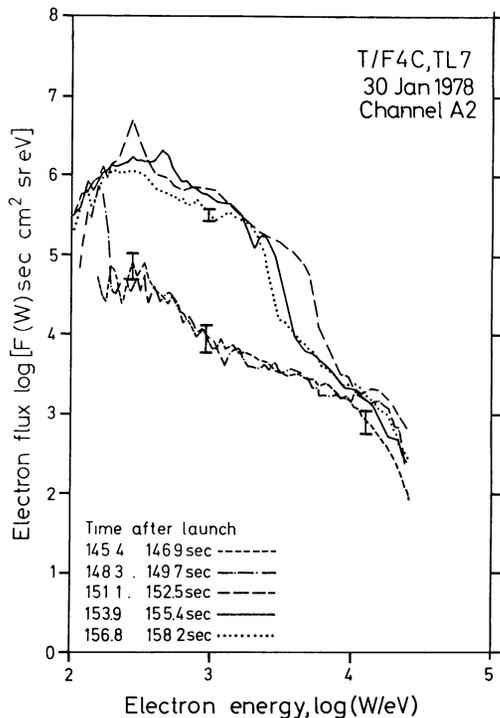


Fig. 4. Electron energy spectra before and after the flux increase at 151 s during flight F4C. Note the dramatic rise of the intensity of electrons below 5 keV only

At 151 s the downgoing proton energy and number fluxes in the energy range 0.5 to 30 keV increase simultaneously with the electron fluxes. Twenty seconds later a stronger precipitation of protons occurred for some seconds. During the rest of the flight the proton energy influx remained at a nearly constant level of $0.7 \text{ erg cm}^{-2} \text{ s}^{-1}$ at a height of 350 km, where charge exchange reactions become more and more important.

The above two time periods, and that after 500 s when the electron spectrometer is in the right mode to observe fast variations, are selected for comparisons with the proton data.

The electron energy spectra for the time interval near the first flux increase at 151 s have been compiled in Figure 4. The spectra just before the increase were relatively flat in the energy range 0.1 keV to approximately 10 keV. The high energy flux after 151 s was carried by electrons with energies between 0.1–5 keV. During the flux enhancement of nearly two orders of magnitude the high energy cutoff decreased from 5 to 2.4 keV within 6–7 s.

In the upper part of Figure 5 the energy channel 0.75 keV is plotted on an expanded time scale between 100 s and 220 s. At the onset of the intense electron precipitation (at 151 s flight time) a sudden increase of the intensity of the proton fluxes of 0.75 keV appeared. After 3 s the intensity decreased again and at 166 s the proton flux reached a new peak value one order of magnitude higher than the average flux. After 170 s the differential proton flux had an average value of $2 \times 10^6 \text{ protons cm}^{-2} \text{ sr}^{-1} \text{ s}^{-1} \text{ keV}^{-1}$. Three differential proton energy spectra from this event are shown in the bottom part of Figure 5. At 167 s, the intensity was approximately one order of magnitude higher in the full measured energy range 0.5–30 keV. The dashed line indicates the proton energy spectrum at 155 s in order to facilitate the comparison.

Interesting observations related to the enhanced electron precipitation near 151 s were not obtained only as spectral variations of the electrons, as illustrated in Figure 4, but also as a variation of the proton energy distribution (Fig. 6). Three successive peaks in the proton spectra were measured, the first at a peak energy of 13 keV, followed 13 s later by one at 7 keV and 6 s later by a 5.5 keV peak. These three peaks were the most remarkable and appeared for some seconds, whereas the dispersion of other energies could not be clearly identified due to some flux intensity variations during that time. A fourth peak at 9.3 keV fitting into the sequence was measured between 192 and 196 s, but has been omitted as it failed to reach statistical significance and fell within an energy range only coarsely covered by the instruments. These observations led us to assume that the change in the proton spectra was caused by energy dispersion. This will be discussed in the next section. The pitch angle distribution of the protons during these successive peaks was isotropic inside the loss cone.

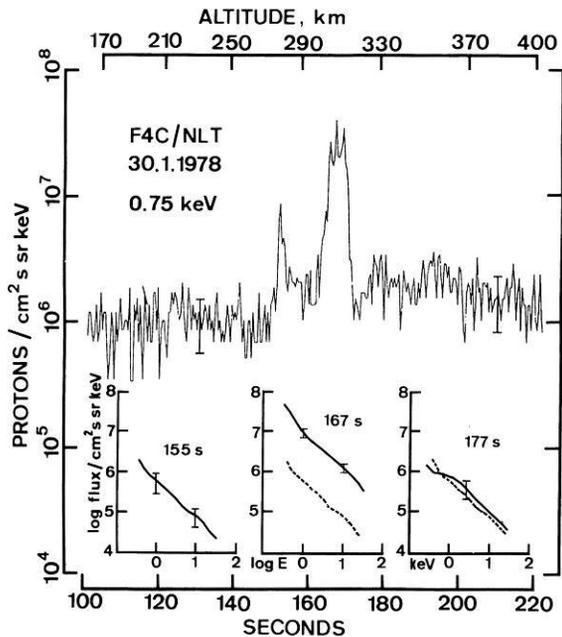


Fig. 5. The 0.75 keV proton channel during the intense electron precipitation in an expanded time scale. In the lower part three differential energy spectra of the protons are compiled demonstrating the intensity increasing in the full energy range

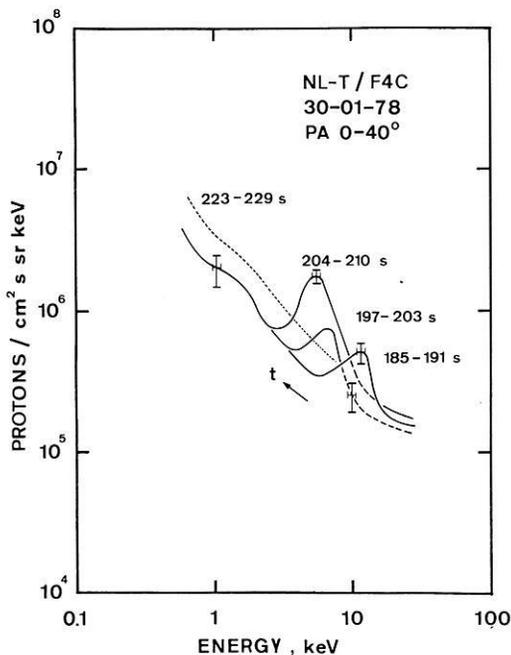


Fig. 6. Proton spectral variations showing energy dispersion effects in relation to the increase of electrons at 151 s. From the time of the appearance of the most marked energy peaks, the distance of the proton precipitation region was evaluated

Further peaked proton spectra were measured during this flight at 290 s and 485 s after launch. The differential energy spectrum in the time interval 285 to 295 s is plotted in Figure 7 and shows a remarkable peak at 3 keV. Below, the response of the detector to monoenergetic electrons, obtained from the calibration procedure, is indicated. The peak rose from a power law spectrum $F(E) = F_0 E^{-\gamma}$ typical for this flight with a spectral coefficient

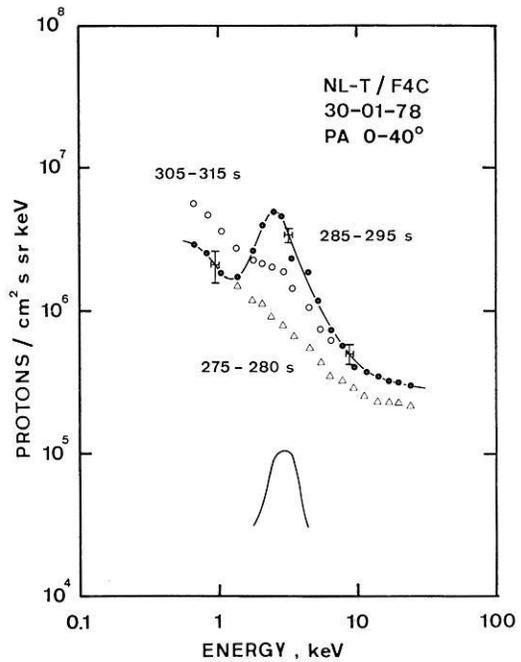


Fig. 7. Proton energy spectrum with the appearance of a mono-energetic peak before the large variations of the electron number flux at 300 s. Below the detector response to mono-energetic particles, as obtained from the calibration procedure of the sensor, is indicated

of $\gamma \sim 0.7$. The flux remained very constant during approximately 20 s and then dropped to a power law spectrum which is a factor 2 or 3 higher than prior to the peak formation. At the same time a large variation in the electron number flux was detected (see Fig. 3). The pitch angle distribution at the proton peak energy 2.9 keV was a typical loss cone distribution during this event.

The other isolated peak in the proton energy spectra was observed within the time interval 483–487 s after launch. The pitch angle distribution at the peak energy 5.36 keV was isotropic before, during, and after the peak.

Discussion and Conclusion

What are the conclusions that can be drawn from the reported observations with respect to the rapid electron energy flux enhancement at 151 s? Under the assumption that the electron enhancement defines the onset time (t_0) for an energy dispersion of the observed variations of the proton spectra (Fig. 6), we calculated the distance s of the proton precipitation region from the proton energy, E_p , and mass, m_p , using the basic formula (1) and

$$s = v_p \cdot t = \sqrt{\frac{2 E_p}{m_p}} \cdot t \quad (1)$$

the transit time t of the protons with the well developed peak energy (Table 2).

For the source distance s statistical uncertainties are calculated from the time where the flux intensity reached 70% of the maximum at the corresponding peak energy. The start time of the velocity dispersion event has been set at 150 s.

For all protons with peak energies we obtained a distance from the source region of approx. 9 Earth radii (R_E). Along a geomagnetic fieldline (dipole model) which is crossed by a rocket at an L -value of approx. 7 the distance from the Earth to the equatorial region is approx. 9 R_E .

Table 2. Compiled parameters transit time t at the maximum flux intensity E_p , velocity v_p and distance s evaluated from the energy dispersion, t_0 = onset time

t after $t_0 = 150$ s	E_p [keV]	v_p [m/s]	s [m]
$t_0 + 37$ s	13	1.58×10^6	$(58.6 \pm 2.5) \times 10^6$
$t_0 + 50$ s	7	1.16×10^6	$(58.1 \pm 2.4) \times 10^6$
$t_0 + 56$ s	5.5	1.03×10^6	$(58.7 \pm 2.1) \times 10^6$

At $t = 150$ s after launch an as yet unexplained diffusion mechanism precipitated electrons and protons in the equatorial region into the local loss cone. The transit time of the electrons (of the order of one second) was very short compared to the transit time of the protons (37 to 56 s) and their arrival (at 151 s) indicates the start of the event as being approx. one second before.

The dispersion calculation is based on the assumption that the ions are protons and not heavier ions (e.g. O^+). No velocity changes due to acceleration or deceleration in field aligned electric fields are included in the model because the field mentioned below disappeared before the protons of the primary source could be affected.

In addition to the sudden increase of the precipitation of the low energy electrons (with an energy < 5 keV) Wilhelm and Stüdemann (in press 1982) reported an energy shift in the electron spectrum together with a post-acceleration of energetic electrons by a 5 kV potential difference. As a mechanism for generating a potential drop, Hultqvist (1971) proposed an interaction between the hot magnetospheric plasma and the cold ionospheric plasma during intense electron precipitation.

No sufficient explanation can be given for the high intensities of the proton fluxes at around 167 s with a duration of some seconds. The electron data were measured in a slow sweep mode, so that the interpretation is rather difficult.

The large variations of the electron number flux at about 300 s were accompanied for 20 s by a well developed peak of the proton spectra at 3 keV. Bryant et al. (1977) reported a similar observation when the rocket crossed from a stable auroral arc into a diffuse aurora. Unfortunately, all-sky photographs or photometer recordings do not exist from flight F4C because of the overcast sky, so that no direct comparison can be made.

A further isolated peak of the proton energy spectra at 5.36 keV with a differential flux ratio of three above the background was observed later from 483–487 s. The assumption of a proton acceleration by a potential drop directed to the Earth and located at higher altitudes is supported by the strong decay of electrons at lower energies at that time. Comparing the strong decrease of the electron number flux at about 500 s to the continued, fairly constant electron energy flux in Figure 3 one may recognize a loss of electrons with lower energies, possibly retarded by a potential drop.

Any related proton energy flux increase one might expect is not observed. However, around that time there is a decreasing slope for both electrons and protons possibly due to a weakening of a common precipitation mechanism, which could counteract the effects of the potential drop.

Acknowledgements. The project "Substorm Phenomena" was supported financially by the German Bundesministerium für Forschung und Technologie. The Austrian contribution was made possible by grants of the Austrian Academy of Sciences and the Austrian Council for Scientific Research.

References

- Brüning, K., Baumjohann, W., Wilhelm, K., Stüdemann, W., Urban, A., Ott, W., Spenner, K., Schmidtke, G.L., Fischer, H.M.: Application of different methods for the determination of ionospheric conductivities from sounding rocket observations. *J. Geophys.* **49**, 74–81, 1981
- Bryant, D.A., Hall, D.S., Lepine, D.R., Mason, R.W.N.: Electrons and positive ions in an auroral arc. *Nature* **266**, 148–149, 1977
- Crandall, D.H., Ray, J.A.: Channeltron efficiency for counting of H^+ and H^- at low energy. *Rev. Sci. Instr.* **46**, 562–564, 1975
- Edwards, T., Bryant, D.A., Smith, M.J.: Preliminary results of particle measurements from Skylark SL 1421 and Fulmar F1. Paper presented at MIST, Southampton, 7 April, 1978
- Egidi, A., Marconero, R., Pizella, G., Sperli, F.: Channeltron fatigue and efficiency for protons and electrons. *Rev. Sci. Instr.* **40**, 88–91, 1969
- Hultqvist, B.: On the productions of magnetic field aligned electric fields by the interaction between the hot magnetospheric plasma and the cold ionosphere. *Planet. Space Sci.* **19**, 149–158, 1971
- Hultqvist, B., Borg, H.: Observations of energetic ions in inverted V-events. *Planet. Space Sci.* **26**, 673–689, 1978
- Iglesias, G.E., McGarity, J.O.: Channel electron multiplier efficiency for protons of 0.2 to 10 keV. *Rev. Sci. Instr.* **42**, 1728–1729, 1971
- Johnstone, A.D.: Correlation between electron and proton fluxes in post-breakup aurora. *J. Geophys. Res.* **76**, 5259–5267, 1971
- Miller, J.R., Whalen, B.A.: Characteristics of auroral proton precipitation observed from sounding rockets. *J. Geophys. Res.* **81**, 147–154, 1976
- Urban, A.: Measurements of low energy auroral ions. *Planet. Space Sci.* **29**, 1353–1365, 1981
- Whalen, B.A., Miller, J.R., McDiarmid, I.B.: Evidence for a solar wind origin of auroral ions from low-energy ion measurements. *J. Geophys. Res.* **76**, 2406–2418, 1971
- Whalen, B.A., McDiarmid, I.B.: Further low-energy auroral ion composition measurements. *J. Geophys. Res.* **77**, 1306–1310, 1972
- Whalen, B.A., Bernstein, W., Daly, P.W.: Low altitude acceleration of ionospheric ions. *Geophys. Res. Lett.* **5**, 55–58, 1978
- Wilhelm, K.: Study of magnetospheric substorm events. *ESA-SP* **152**, 269–277, 1980
- Wilhelm, K., Stüdemann, W.: Evidence of field-aligned electrostatic acceleration of auroral electrons. *J. Geophys.*, in press 1982

Received May 25, 1981; Revised version November 16, 1981 and January 22, 1982

Accepted January 28, 1982

Joint Two-Dimensional Observations of Ground Magnetic and Ionospheric Electric Fields Associated with Auroral Currents

5. Current System Associated with Eastward Drifting Omega Bands

Dieter André^{1*} and Wolfgang Baumjohann²

¹ Max-Planck-Institut für Aeronomie, Postfach 20, D-3411 Katlenburg-Lindau 3, Federal Republic of Germany

² Institut für Geophysik der Universität Münster, Corrensstr. 24, D-4400 Münster, Federal Republic of Germany

Abstract. Magnetograms from the Scandinavian Magnetometer Array (SMA) and ionospheric electric field measurements from the Scandinavian Twin Auroral Radar Experiment (STARE) have been evaluated during the passage of several omega bands (or eastward travelling surges) over Northern Scandinavia around 0400 MLT on 16 February 1977. The eastward motion of the omega bands was a pure $E \times B$ drift and was associated with Ps6 disturbances of the east-west component particularly of magnetic and electric fields. The two-dimensional distributions of electric and magnetic disturbance fields strongly support the hypothesis of a three-dimensional current system which is embedded in a rather homogeneous westward Hall current and where north-south aligned localized regions of field-aligned currents, flowing upward close to the omega band's wave-crest and downward east and west of it, are travelling eastward with the omega band. These field-aligned currents are associated with eastward and westward electric field disturbances and southward and northward Hall currents between them. It remains unclear whether these Hall currents close within the ionosphere, i.e. forming counterclockwise and clockwise ellipsoidal Hall current vortices, or if they diverge from the ionosphere as field-aligned currents at the southern and northern boundary of the region that is disturbed by the omega band.

Key words: Scandinavian Magnetometer Array – STARE – Magnetic fields – Electric fields – Auroral zone – Omega bands – Eastward travelling surges – Ps6

Introduction

In his review paper on auroral forms Akasofu (1974) defines omega bands as a wavy structure of the poleward auroral boundary with the shape of a series of Ω open to the pole, appearing most often in the local morning or midnight sector and generally drifting eastward (therefore the term “eastward travelling surge” is also used). We have started from this definition to determine a period when omega bands occurred and have then evaluated the STARE and magnetometer data for this time. This means that our classification is different from that of earlier papers, which describe mainly a certain class of magnetic variations, and may thereby also include events that are not connected to omega bands.

A comparison of our magnetic variations show a certain similarity to the Ps6 described by Saito (1974, 1978) and with

eastward travelling disturbance described by Kawasaki and Rostoker (1979) and Gustafsson et al. (1981). Saito infers that Ps6 are caused by substorms and generated by a wavelike motion of the electrojet travelling eastward in the morning sector. Kawasaki and Rostoker (1979) have modelled their disturbances by a system of travelling field-aligned currents closing in the north-south direction. Gustafsson et al. (1981) reproduced their equivalent current pattern by Hall current vortices around an east-west aligned pair of line-type field-aligned currents.

Since, in contrast to these earlier measurements, we have two-dimensional measurements of ground magnetic and ionospheric electric fields, we are – as in earlier papers of this series (Baumjohann et al. 1980; 1981; Baumjohann and Kamide 1981; Inhester et al. 1981) – able not only to verify that our measurements belong to the same class of events, but can decide which model is the real one.

Instrumentation

The sites of the SMA-magnetometers (Scandinavian Magnetometer Array) used for this study are shown in Fig. 1. Most of the instruments are located along six roughly parallel, north-south profiles. A detailed description of this magnetometer network has been given by Küppers et al. (1979) and, for profile 5, by Maurer and Theile (1978). The coordinate system indicated in Fig. 1 has been introduced by Küppers et al. (1979) and has been named the Kiruna system. It is a Cartesian system obtained by a stereographic projection of the globe onto a tangential plane centered at Kiruna, Sweden (67.8° N, 20.4° E). The y_{KI} axis of the system has been chosen as the tangent to the projection of the line ϕ_c (KIR)=64.8° with ϕ_c denoting the revised corrected geomagnetic latitudes as given by Gustafsson (1970). The x_{KI} axis points approximately 12° west of geographic north at KIR.

Again referring to Fig. 1, the area enclosed by the dashed trapezoid is the common region seen by the two STARE radars (Greenwald et al. 1978). The STARE radars are sensitive to electrostatic plasma waves in the auroral E region. These waves, often called irregularities, are produced by the combined effects of the two stream and gradient drift plasma instabilities (see Greenwald, 1974). Greenwald (1979) has summarized evidence that in the auroral E layer the net drift velocity of the irregularities is nearly a pure $E \times B$ drift. Hence, the transverse ionospheric electric field vector is orthogonal and proportional to the measured drift velocity vector. Strong experimental evidence for this relationship has been found by Ecklund et al. (1977), Cahill et al. (1978) and Zanetti et al. (1980).

* Present address: Siemens AG, Hofmannstr. 51, D-8000 München 70, Federal Republic of Germany

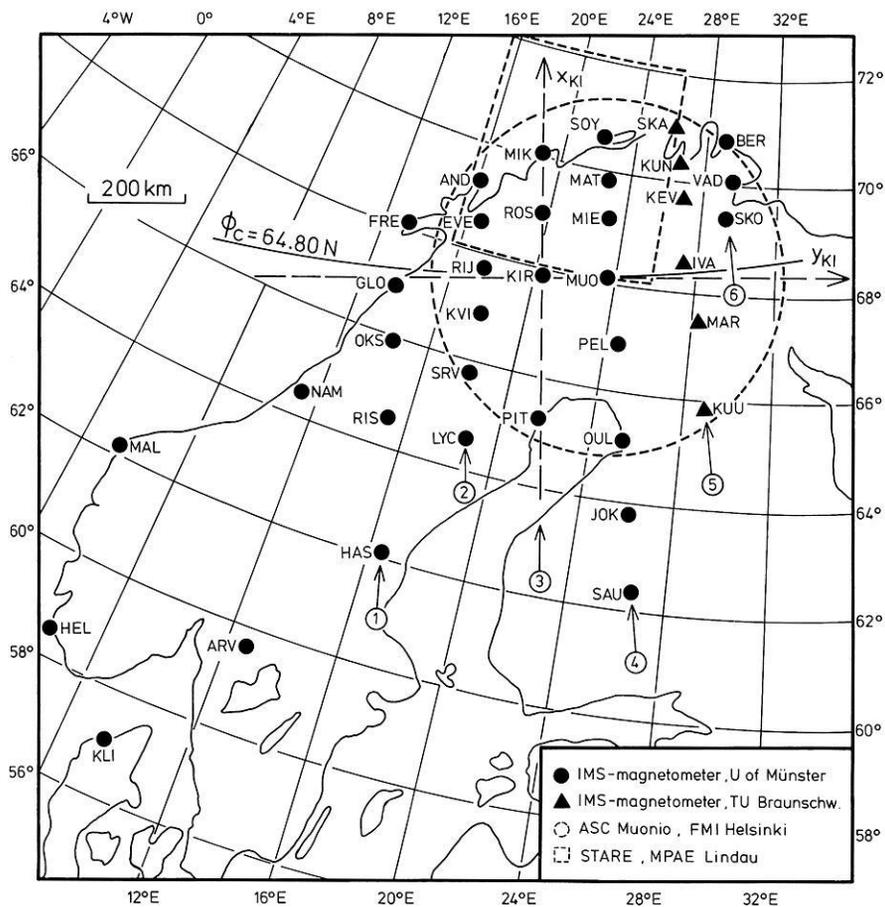


Fig. 1. Locations of the magnetic stations used in this study. The dashed trapezoid represents the common observation area of both STARE radars. The observational coverage of the Muonio all-sky camera is indicated by the broken circle (limited by an elevation angle of 15° and a height of 100 km). The axes define the Kiruna system (see text). The ringed numbers are the profile numbers of the six latitudinal magnetometer chains

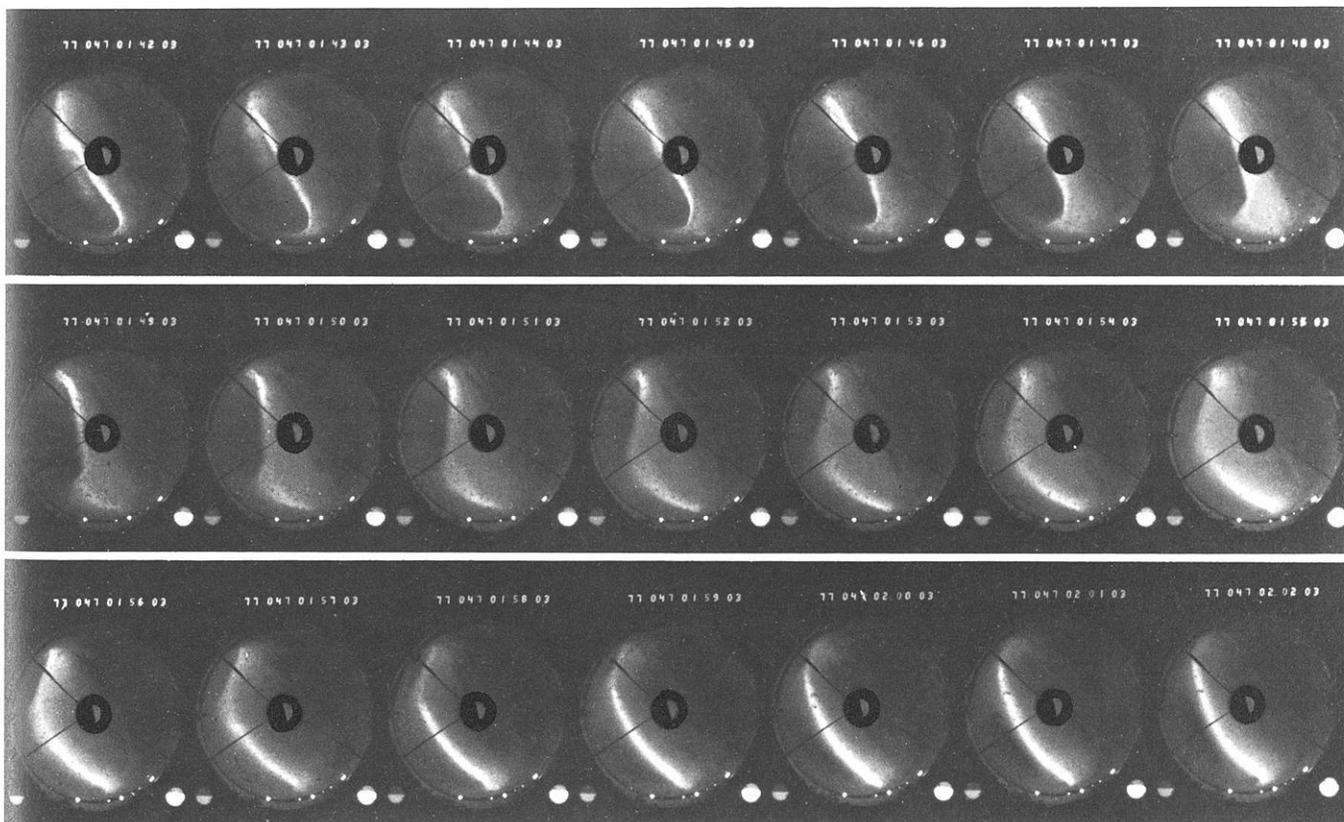


Fig. 2. Photographs taken by the all-sky camera at Muonio every minute between 0142 UT (upper left corner) and 0202 UT (lower right corner). Geographic north is to the left and geographic east to the top of the figure

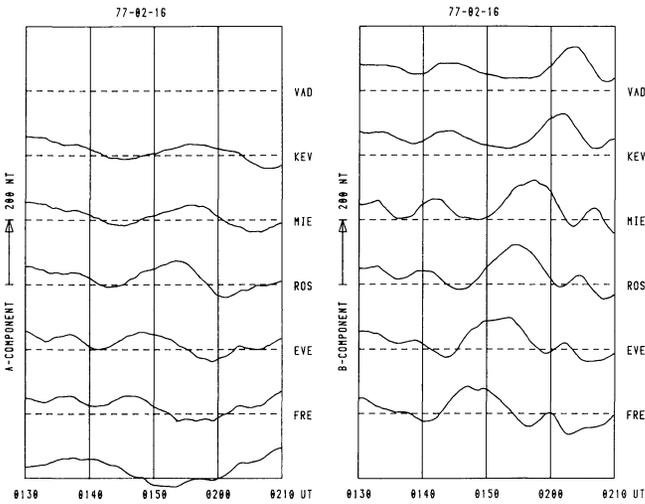


Fig. 3. Magnetograms of A and B components on a longitudinal profile (VAD-FRE, x_{KI} : +100 km, y_{KI} : +400 to -300 km). The A and B components are aligned along the x_{KI} and y_{KI} axes, respectively (see Fig. 1)

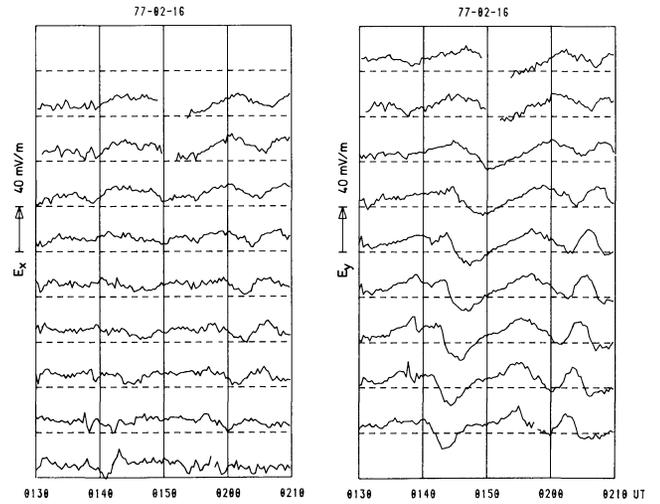


Fig. 4. Horizontal ionospheric electric field components (E_x and E_y are parallel to x_{KI} and y_{KI} , respectively) along a longitudinal profile at $x_{KI} = +250$ km. The corresponding y_{KI} coordinates are (from top): +250, +200, +150, +100, +50, 0, -50, -100, -150 km. The components displayed are averages of the STARE observations within a radius of 25 km around the grid point (exponentially weighted by their distance from the center). Gaps in the traces occur where the electric field magnitude was below the threshold level of 15–25 mV m^{-1}

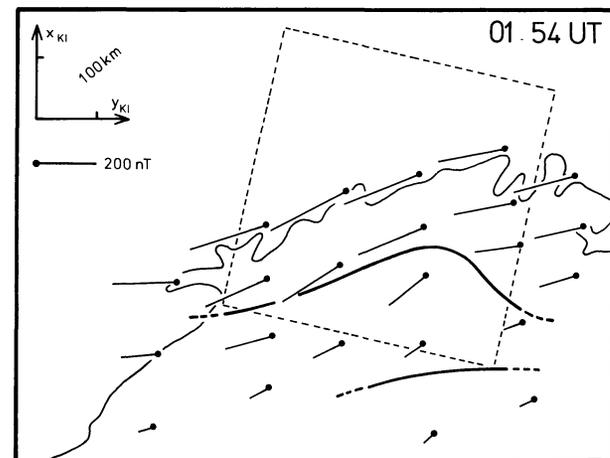
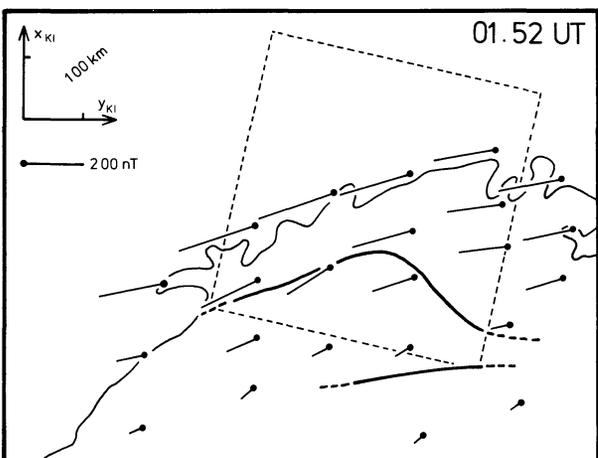
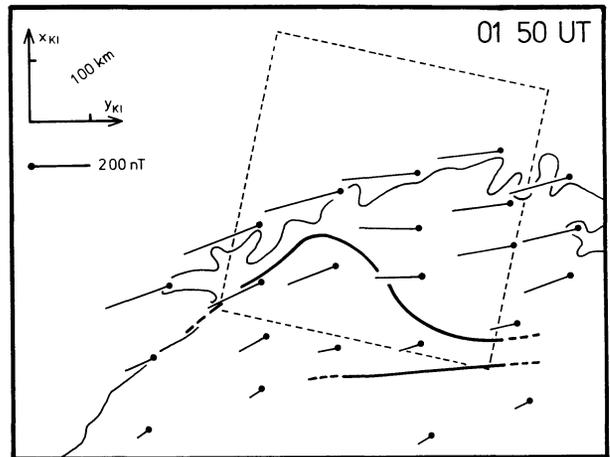
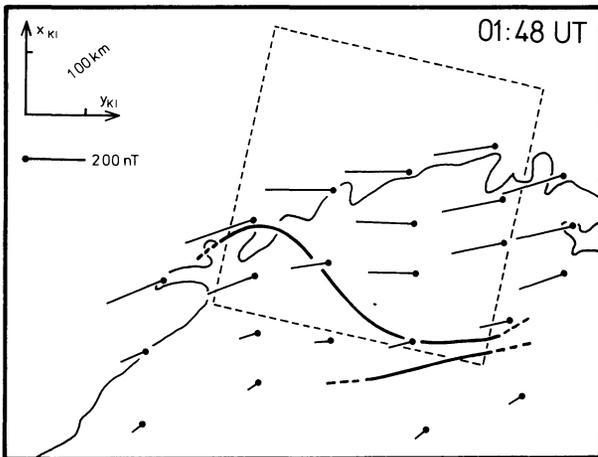


Fig. 5. Spatial distribution of equivalent current vectors on the ground and the lower boundary of the omega band (dashed where extrapolated) every 2 min between 0148 and 0154 UT. The equivalent current vectors originate where the corresponding magnetic disturbance vectors were observed. The *dashed square* gives the location of the STARE frame for easier comparison with Fig. 6

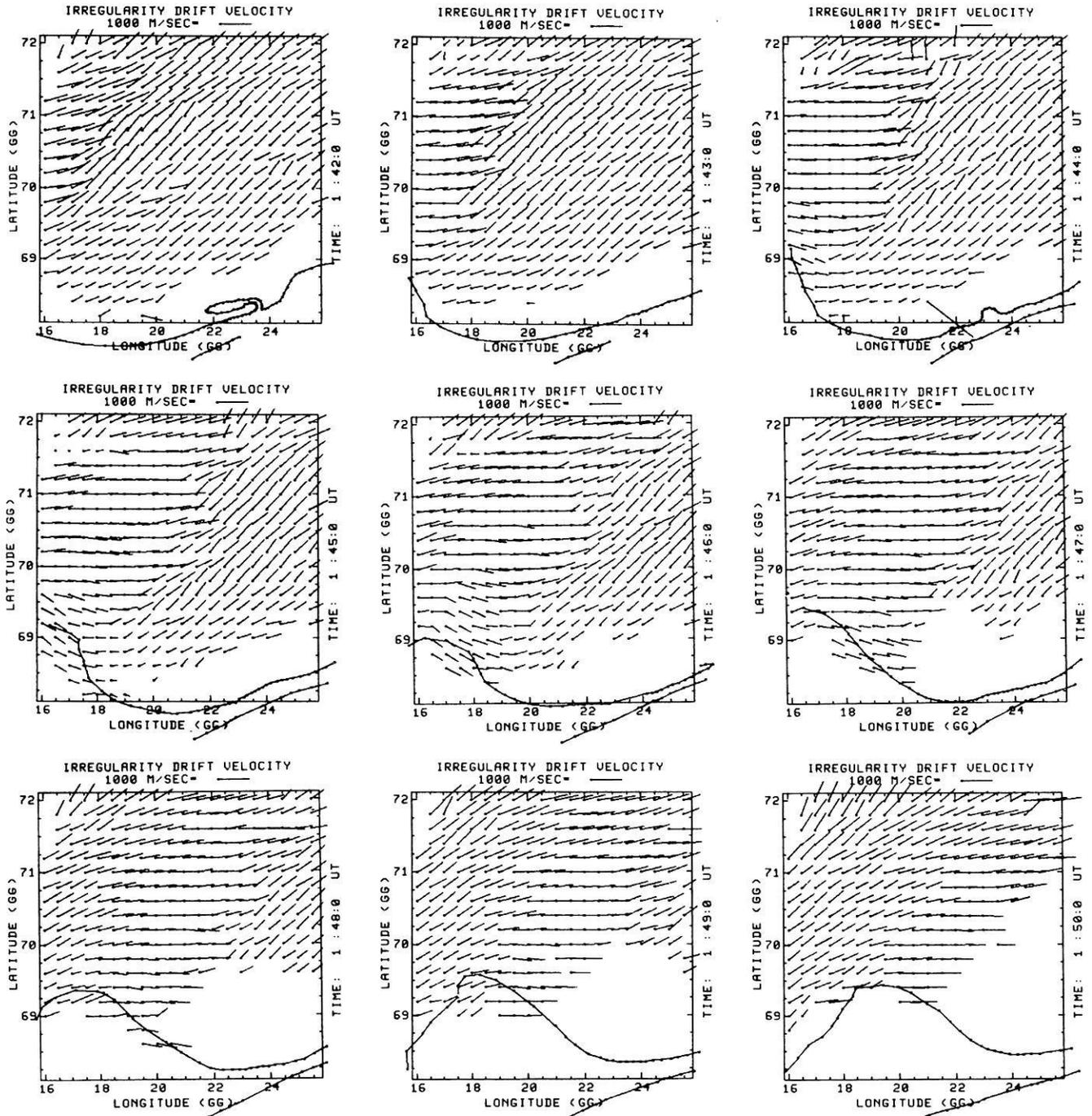


Fig. 6. Spatial distribution of STARE irregularity drift vectors and the lower boundary of the omega band every minute between 0142 and 0150 UT

The lower aurora boundary in the all-sky pictures of the camera at Muonio (Hyppönen et al. 1974) has been digitized, rectified and mapped into geographical coordinates and into the Kiruna system under the assumption that the lower boundary of the aurora is located at a height of 100 km (Boyd et al. 1971). The resulting error in the location if the true height is 10 km higher or lower than assumed may be estimated to about 20 km for a zenith distance of 70° (Opgenoorth et al. 1980).

Data Presentation and Analysis

Omega bands were drifting through the all-sky camera field of view from 0130 UT to 0230 UT on 16 February 1977 with

an approximate speed of 40 km/min and a repetition period of 15 to 20 min. In Fig. 2 all-sky camera pictures taken during the passage of the most pronounced omega band are shown.

The whole structure looks like a wave moving eastward (after transformation into geographical or Kiruna coordinates this is even more pronounced). Omega bands occurring before and after this time have the same general appearance, but are less pronounced, in the magnetic and electric field variations.

The temporal development of the horizontal components of ground magnetic and ionospheric electric fields observed along the (longitudinal) y_{KI} axis are displayed in Figs. 3 and 4, respectively (A and E_x are parallel to x_{KI} , B and E_y are parallel to y_{KI}). Both magnetic and electric variations are clearly within

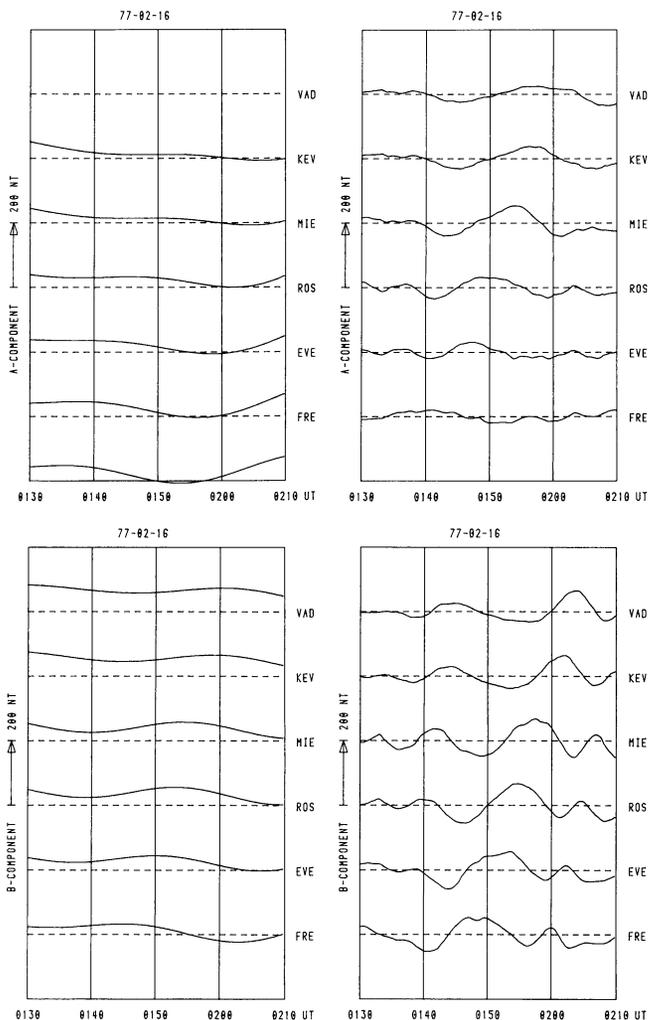


Fig. 7. Filtered magnetograms of *A* and *B* components displayed in Fig. 3. The low frequency traces (periods greater than 30 min) are given in the left-hand column while the higher frequency traces (periods less than 30 min) are given in the right-hand column

the Ps6 range (15–20 min) and were probably initiated by substorm activity west of Scandinavia before 0100 UT (see Baumjohann and Kamide 1981). As is usual for Ps6, the magnetic fluctuations are more pronounced in the (east-west) *B* component (Saito 1972; 1974; 1978; Kawasaki and Rostoker 1979; Gustafsson et al. 1981). The electric field also varies considerably more in the (east-west) *E_y* component, in agreement with the measurements of Gustafsson et al. (1981). It can also be seen clearly that the extrema are observed earlier at the western stations and that they are travelling eastwards with the same velocity as the omega bands, i.e. about 40 km/min.

The relation between the omega bands and the disturbances in magnetic and electric fields can be seen even more clearly in Figs. 5 and 6, where we show the two-dimensional distribution of equivalent current vectors and STARE drift velocities together with the auroral boundary during the passage of the most pronounced omega band over Scandinavia. The STARE data show a rather sharply bounded region of eastward drift embedded in the otherwise northeast drift. The region is about 150 km wide in longitude and extends 250 km to the northeast. It is moving with the omega band. The equivalent current vectors display a southwest orientation in the region of northeast irregularity drifts and westward equivalent current near the omega

band wave-crest where STARE observed eastward drifts. This behaviour of irregularity drifts and current vectors is equivalent to the wavy motion of the polar electrojet that has been proposed by Saito (1978) as an explanation for Ps6 magnetic variations.

In order to analyze more accurately the disturbances of the background electric field and electrojet caused by the omega band, we have followed Gustafsson et al. (1981) and separated the Ps6 fluctuations from the more slowly varying contributions by filtering (with the -6dB cutoff point at a period of 30 min) magnetograms and electrograms measured by the SMA and STARE (some of these are shown in Figs. 3 and 4). The results of this filtering, i.e. the high frequency signal associated with the omega band and the low frequency signal associated with background electric field and electrojet, are shown in Figs. 7 and 8 for those traces displayed in Figs. 3 and 4. Since the STARE electrograms showed some very high frequency scatter as well as some data gaps due to subthreshold electric fields (Cahill et al. 1978) smooth cubic spline functions were fitted to the electrograms before filtering (these are also displayed in Fig. 8).

We will now concentrate on the most pronounced omega band that crossed northern Scandinavia between 0140 and 0200 UT (Figs. 3 and 4). This omega band travelled to the east with a velocity component $v_y = 670 \text{ ms}^{-1}$ along the y_{KI} axis and had also a small northward velocity component $v_x = 85 \text{ ms}^{-1}$ parallel to the x_{KI} axis. Figures 2, 3 and 4 show that the omega band maintains its shape during the traversal of northern Scandinavia. In addition, the irregularity drift and equivalent current vector fields obviously remain constant relative to the wave crest. This strongly suggests a stationary pattern for the omega band and the magnetic and electric perturbations related to it which drifts with the omega band. In order to display this behaviour clearly the (two-dimensional) total, the low frequency background, and the higher frequency disturbance vector field distributions measured between 0135 and 0205 UT were each superimposed with the moving omega band as a natural reference frame and averaged in $50 \times 50 \text{ km}^2$ cells. The resultant vector distributions are shown in Fig. 9.

The magnetic and electric vector fields display quite clearly that the filtering in the time domain has actually resulted in a spatial filtering where we have separated the total measured vector fields into a both temporally and spatially rather homogeneous background field which is very probably due to magnetospheric convection and a temporally and spatially inhomogeneous disturbance field that is associated with the omega band.

Additional Results and Discussion

In previous studies of eastward travelling disturbances mainly ground-based magnetic data were analysed and thus the different authors were not able to definitely discriminate between the two (following) different current systems that are both able to reproduce the higher frequency part of the magnetic variations on the ground:

- 1) Circular Hall current loops around line-type field-aligned currents in the centers of the vortices with no conductivity gradients in the ionosphere (Baumjohann 1979; Gustafsson et al. 1981).
- 2) A southward Cowling current flowing in a region of higher conductivity closed by field-aligned currents in the north and south (Baumjohann et al. 1977; Kawasaki and Rostoker 1979; Rostoker and Barichello 1980).

Our two-dimensional observations of ground-based magnetic and ionospheric electric fields allow us to infer the real nature

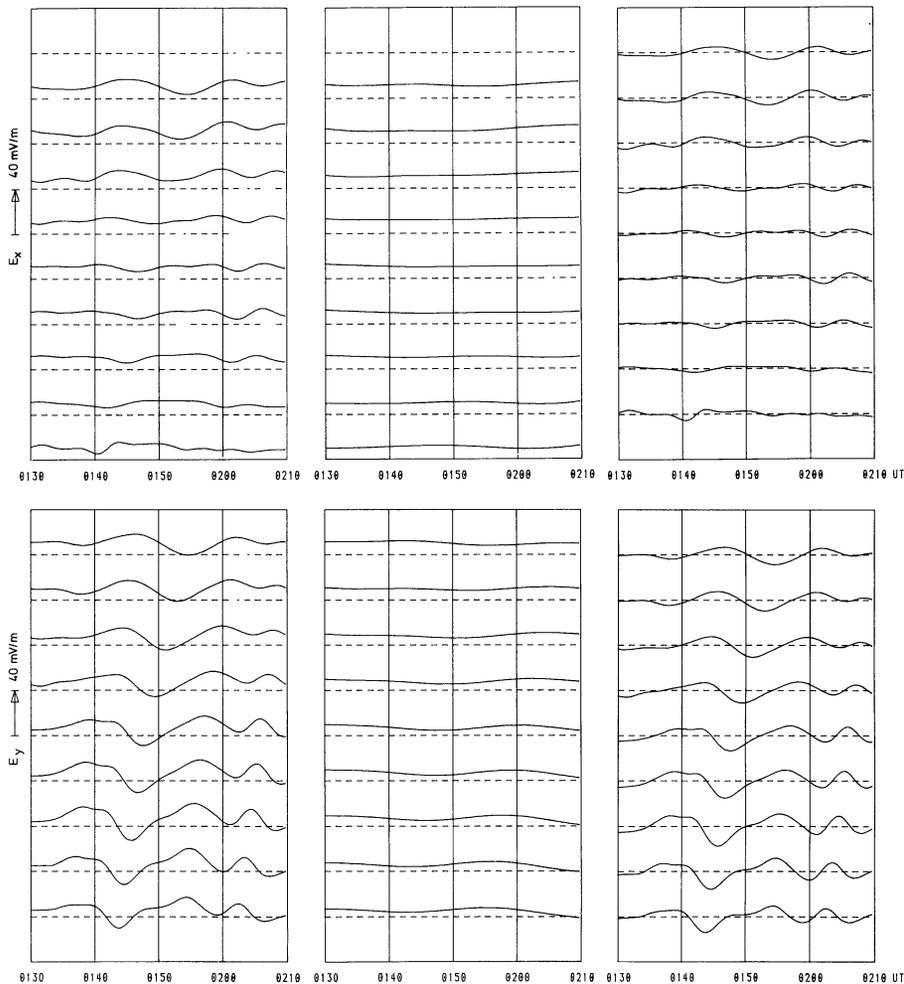


Fig. 8. Filtered electrograms of E_x (upper row) and E_y (lower row) components. The traces in the left-hand column have been obtained by fitting smooth cubic spline functions to the original data displayed in Fig. 4 in order to reduce very high frequency scatter and to fill data gaps due to subthreshold fields before filtering. Low frequency and higher frequency traces (periods greater/less than 30 min) are given in the middle and right-hand columns, respectively

of both the slowly varying and homogeneous background electrojet and the rather inhomogeneous currents associated with the omega band.

The low frequency background electric field has amplitudes of about -30 and 5 mVm^{-1} for the E_x and E_y components, respectively. These values are approximately the same as those found by Baumjohann and Kamide (1981) 1 h later on the same day after the omega bands have passed over Scandinavia and are also in good agreement with other electric field observations in this local time sector (Mozer and Lucht 1974; Madsen et al. 1976; Holzworth et al. 1977). The slowly varying background horizontal magnetic field and equivalent current were nearly exactly parallel and orthogonal, respectively, to the direction of the convection electric field, i.e. 12° eastward and northward from the negative x_{KI} and y_{KI} axes, respectively. This leads to the conclusion that the background westward electrojet is a Hall current flowing along the auroral oval that is typically also inclined by about 10° – 15° against lines of constant magnetic latitude, i.e. the x_{KI} axis (Davis 1962; Gustafsson 1967; 1969; Meng et al. 1977; Lassen and Danielsen 1978; see Baumjohann and Kamide (1981) for a more thorough discussion of this topic).

The drift velocity vector of the omega band points about 10° eastward from the negative y_{KI} axis, i.e. perpendicular to the background electric field. Moreover, by comparing the drift velocity of the omega band $(v_x, v_y) = (85 \text{ ms}^{-1}, 670 \text{ ms}^{-1})$ with the horizontal background electric field components given above, it becomes obvious that the eastward motion of the omega band along the auroral oval is a pure $E \times B$ drift. The slightly too

high velocity of the omega band as compared to the convection electric field measured by STARE in the earth-fixed frame may easily be attributed to the dynamo action of neutral winds in the ionosphere (Brekke et al. 1973; Comfort et al. 1976; Rino et al. 1977).

The electric field distribution associated with the omega band can be explained by a north-south aligned region of upward field-aligned currents and thus negatively charged field lines just north of the wave crest and of downward field-aligned currents and thus positively charged field lines east and west of the omega band (the electric charge dissipating as Pedersen currents in the ionosphere from the foot of the field lines is assumably compensated by field-aligned current from the magnetosphere; see Fukushima 1974). The location of the upward field-aligned current region corroborates the results of other studies on Ps6 and omega bands, which describe observations of strong riometer absorption and thus energetic electron precipitation that is supposed to carry upward field-aligned current (Kamide and Rostoker 1977; Theile and Wilhelm 1980) west of the southward equivalent current region (Gustafsson et al. 1981) and just east of the region of purely eastward total irregularity drift vectors (Oksman et al., personal communication 1981).

A comparison of the higher frequency electric and equivalent current vector fields shows that eastward and westward electric field vectors and southward and northward equivalent current vectors are nearly perpendicular to each other. This favours the Hall current vortex model against the southward Cowling current model. However, it remains unclear whether the Hall

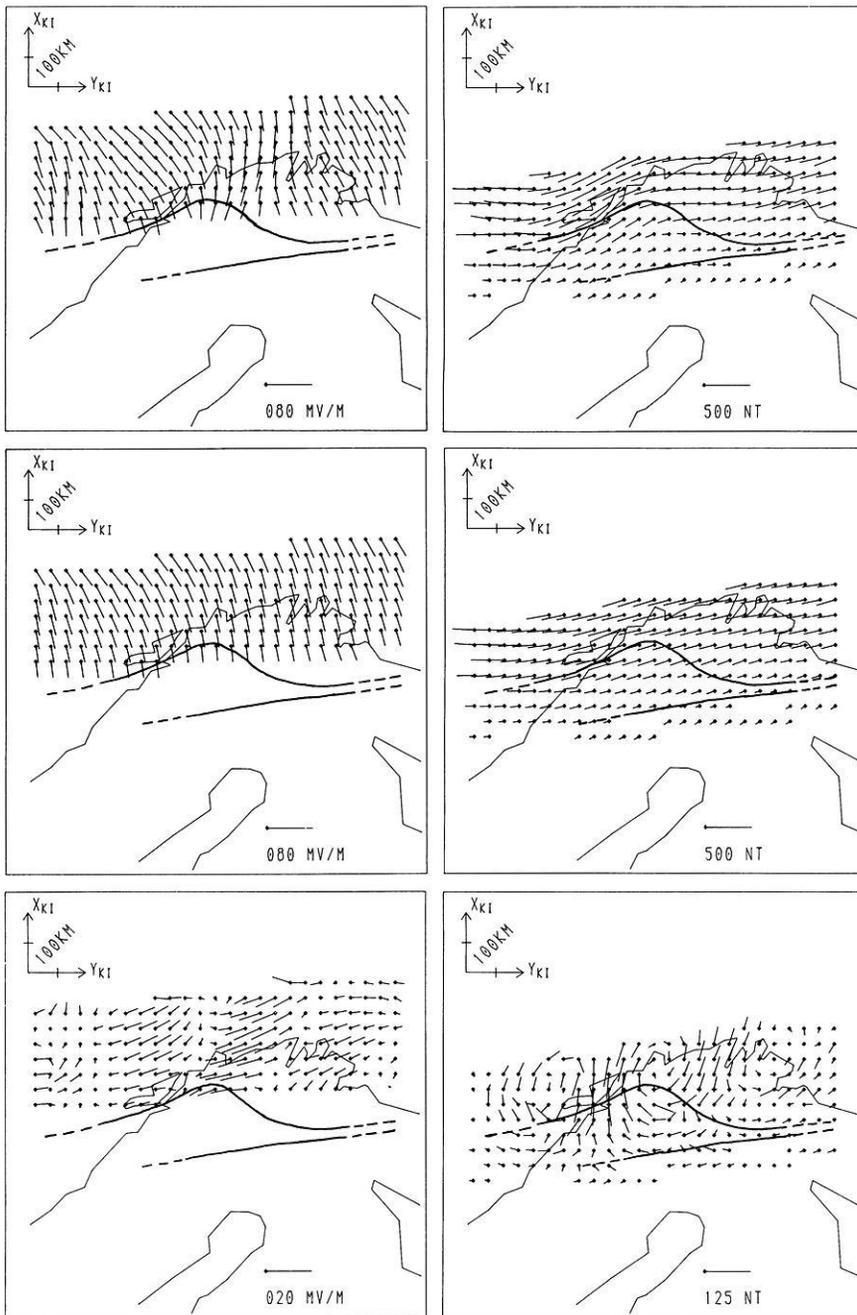


Fig. 9. Spatial distribution of electric field vectors (left-hand column) and equivalent current vectors on the ground (right-hand column) measured between 0135 and 0205 UT, superimposed and averaged in $50 \times 50 \text{ km}^2$ cells under the assumption of a stationary pattern drifting with the omega band, whose average lower boundary has also been drawn in. The upper panel gives the unfiltered vector fields while middle and lower panels give lower and higher frequency components respectively. The outline of the Scandinavian coast is given relative to its location at 0150 UT

currents close within the ionosphere via east-west Hall currents or diverge from the ionosphere as field-aligned currents at the southern and northern boundary of the region that is disturbed by the omega band, since the electric field below the omega band's wave-crest was obviously too low to be detected by STARE and thus we are not able to decide if the eastward and westward equivalent currents in this region are Hall currents or if they are caused by the magnetic effect of field-aligned currents.

From the equivalent current vector distribution one can see that the Hall conductivity near and just behind the wave crest is somewhat enhanced, since the southward current vectors are comparatively bigger than the northward current vectors to the east. (note that this east-west aligned conductivity gradient is parallel to the electric field and to the gradient of the electric potential and thus, as in the case of homogeneous conductivity, only Hall currents are observed on the ground). The Hall conduc-

tivity enhancement is apparently caused by the precipitation of energetic electrons that carry the upward field-aligned current.

Acknowledgements. The magnetic observations were performed in cooperation with the Aarhus University, the Royal Institute of Technology at Stockholm, the Finnish Meteorological Institute at Helsinki, the University of Bergen, the Geophysical Observatory at Sodankylä, the Kiruna Geophysical Institute, the University of Oulu and the University at Tromsø. The STARE radars are operated with ELAB and the Norwegian Technical University in Trondheim and the Finnish Meteorological Institute in Helsinki. We thank these institutions and all members of the magnetometer group at the University of Münster and the STARE group at the MPAE for their support. We would also like to thank H. Maurer and T.U. Braunschweig for supplying us with magnetic data of the Braunschweig chain and R.J. Pellinen, FMI Helsinki, for making available the all-sky camera data. The magnetometer array observations and the work of W. Baumjohann were supported by grants from the Deutsche Forschungsgemeinschaft.

References

- Akasofu, S.-I.: A study of auroral displays photographed from DMSP-2 satellite and from the Alaska meridian chain of stations. *Space Sci. Rev.* **16**, 617–725, 1974
- Baumjohann, W.: Spatially inhomogeneous current configurations as seen by the Scandinavian Magnetometer Array. In: Proceedings of the International Workshop on Selected Topics of Magnetospheric Physics, Japanese IMS Committee, eds.: pp. 35–40. Tokyo 1979
- Baumjohann, W., Kamide, K.: Joint two-dimensional observations of ground magnetic and ionospheric electric fields associated with auroral zone currents. 2. Three-dimensional current flow in the morning sector during substorm recovery. *J. Geomagn. Geoelectr.* **33**, 297–318, 1981
- Baumjohann, W., Gustafsson, G., Küppers, F.: Large-amplitude rapid magnetic variations during a substorm (abstract only). *EOS Trans. AGU* **58**, 718, 1977
- Baumjohann, W., Untiedt, J., Greenwald, R.A.: Joint two-dimensional observations of ground magnetic and ionospheric electric fields associated with auroral zone currents. 1. Three-dimensional current flows associated with a substorm-intensified eastward electrojet. *J. Geophys. Res.* **85**, 1963–1978, 1980
- Baumjohann, W., Pellinen, R.J., Opgenoorth, H.J., Nielsen, E.: Joint two-dimensional observations of ground magnetic and ionospheric electric fields associated with auroral zone currents. 4. Current systems associated with local auroral breakups. *Planet. Space Sci.* **29**, 431–447, 1981
- Boyd, J.S., Belon, A.E., Romick, G.J.: Latitude and time variations in precipitated electron energy inferred from measurements of auroral height. *J. Geophys. Res.* **76**, 7694–7700, 1971
- Brekke, A., Doupnik, J.R., Banks, P.M.: A preliminary study of the neutral wind in the auroral E region. *J. Geophys. Res.* **78**, 8235–8250, 1973
- Cahill, L.J. Jr., Greenwald, R.A., Nielsen, E.: Auroral radar and rocket double-probe observations of the electric field across the Harang-discontinuity. *Geophys. Res. Lett.* **5**, 687–690, 1978
- Comfort, R.H., Wu, S.T., Swenson, G.R.: An analysis of auroral E region neutral winds based on incoherent scatter radar observations at Chatanika. *Planet. Space Sci.* **24**, 541–560, 1976
- Davis, T.N.: The morphology of the auroral displays of 1957–1958. 2. Detail analysis of Alaska data and analysis of high latitude data. *J. Geophys. Res.* **67**, 75–110, 1962
- Ecklund, W.L., Balsley, B.B., Carter, D.A.: A preliminary comparison of F region plasma drifts and E region irregularity drifts in the auroral zone. *J. Geophys. Res.* **82**, 195–197, 1977
- Fukushima, N.: Equivalent current pattern for ground geomagnetic effect when the ionospheric conductivity is discontinuous at the foot of a field-aligned current sheet. *Rep. Ionos. Space Res. Jap.* **28**, 147–151, 1974
- Greenwald, R.A.: Diffuse radar aurora and the gradient drift instability. *J. Geophys. Res.* **79**, 4807–4810, 1974
- Greenwald, R.A.: Studies of currents and electric fields in the auroral zone ionosphere using radar auroral backscatter. In: Dynamics of the magnetosphere, S.-I. Akasofu, ed.: pp. 213–248. Dordrecht: D. Reidel 1979
- Greenwald, R.A., Weiss, W., Nielsen, E., Thomson, N.R.: STARE: a new radar auroral backscatter experiment in northern Scandinavia. *Radio Sci.* **13**, 1021–1039, 1978
- Gustafsson, G.: On the orientation of auroral arcs. *Planet. Space Sci.* **15**, 277–294, 1967
- Gustafsson, G.: Auroral orientation curves and the auroral oval. *Tellus* **21**, 852–860, 1969
- Gustafsson, G.: A revised corrected geomagnetic coordinate system. *Ark. Geofys.* **5**, 595–617, 1970
- Gustafsson, G., Baumjohann, W., Iversen, I.: Multi-method observations and modelling of the three-dimensional currents associated with a very strong Ps6 event. *J. Geophys.* **49**, 138–145, 1981
- Holzworth, R.H., Berthelier, J.-J., Cullers, D.K., Fahleson, U.V., Fälthammar, C.-G., Hudson, M.K., Jalonen, L., Kelley, M.C., Kellogg, P.J., Tanskanen, P., Temerin, M., Mozer, F.S.: The large-scale ionospheric electric field: its variation with magnetic activity and relation to terrestrial kilometric radiation. *J. Geophys. Res.* **82**, 2735–2742, 1977
- Hyppönen, M., Pellinen, R., Sucksdorff, C., Tornainen, R.: Digital all-sky camera. *Techn. Rept. No. 9*, Finnish Meteorol. Inst., 1974
- Inhvester, B., Baumjohann, W., Greenwald, R.A., Nielsen, E.: Joint two-dimensional observations of ground magnetic and ionospheric electric fields associated with auroral zone currents. 3. Three-dimensional currents associated with a westward travelling surge. *J. Geophys. Res.* **49**, 155–162, 1981
- Kamide, Y., Rostoker, G.: The spatial relationship of field-aligned currents and auroral electrojets to the distribution of nightside auroras. *J. Geophys. Res.* **82**, 5589–5608, 1977
- Kawasaki, K., Rostoker, G.: Perturbation magnetic fields and current systems associated with eastward drifting auroral structures. *J. Geophys. Res.* **84**, 1464–1480, 1979
- Küppers, F., Untiedt, J., Baumjohann, W., Lange, K., Jones, A.G.: A two-dimensional magnetometer array for ground-based observations of auroral zone electric currents during the International Magnetospheric Study (IMS). *J. Geophys. Res.* **46**, 429–450, 1979
- Lassen, K., Danielsen, C.: Quiet time pattern of auroral arcs for different directions of the interplanetary magnetic field in the Y-Z-plane. *J. Geophys. Res.* **83**, 5277–5284, 1978
- Madsen, M.M., Iversen, I.B., D'Angelo, N.: Measurements of high-latitude ionospheric electric fields by means of balloon-borne sensors. *J. Geophys. Res.* **81**, 3821–3824, 1976
- Maurer, H., Theile, B.: Parameters of the auroral electrojet from magnetic variations along a meridian. *J. Geophys. Res.* **44**, 415–426, 1978
- Meng, C.-I., Holzworth, R.H., Akasofu, S.-I.: Auroral circle-delineating the poleward boundary of the quiet auroral belt. *J. Geophys. Res.* **82**, 164–172, 1977
- Mozer, F.S., Lucht, P.: The average auroral zone electric field. *J. Geophys. Res.* **79**, 1001–1006, 1974
- Opgenoorth, H.J., Pellinen, R.J., Maurer, H., Küppers, F., Heikkilä, W.J., Kaila, K.U., Tanskanen, P.: Ground-based observations of an onset of localized field-aligned currents during auroral breakup around magnetic midnight. *J. Geophys. Res.* **48**, 101–115, 1980
- Rino, C.L., Brekke, A., Baron, M.J.: High resolution auroral zone E region neutral wind and current measurements by incoherent scatter radar. *J. Geophys. Res.* **82**, 2295–2304, 1977
- Rostoker, G., Barichello, J.C.: Seasonal and diurnal variation of Ps6 magnetic disturbances. *J. Geophys. Res.* **85**, 161–163, 1980
- Saito, T.: Some topics for the study of the mechanism of magnetospheric substorm by means of rocket observation in the auroral zone. *Antarctic Rec.* **43**, 65–69, 1972
- Saito, T.: Examination of the models for the substorm-associated magnetic pulsation Ps6. *Sci. Rep. Tohoku Univ. Ser. 5 Geophys.* **22**, 35–59, 1974
- Saito, T.: Long-period irregular magnetic pulsation, Pi3. *Space Sci. Rev.* **21**, 427–467, 1978
- Theile, B., Wilhelm, K.: Field-aligned currents above an auroral arc. *Planet. Space Sci.* **28**, 351–355, 1980
- Zanetti, L.J. Jr., Arnoldy, R.L., Cahill, L.J. Jr., Behm, D.A., Greenwald, R.A.: Comparative rocket observations of ionospheric electric fields in the auroral oval. *Space Sci. Instr.* **5**, 183–196, 1980

Received September 7, 1981; Revised version November 9, 1981
Accepted November 13, 1981

Thermospheric Winds During the Energy Budget Campaign: Ground-Based Fabry-Perot Observations Supported by Dynamical Simulations with a Three-Dimensional, Time-Dependent Thermospheric Model

D. Rees, P.A. Rounce, P. Charleton, T.J. Fuller-Rowell, I. McWhirter, and K. Smith

Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom

Abstract. Two very stable and sensitive Fabry-Perot interferometers were operated continuously throughout the period of the Energy Budget Campaign, at ESRANGE, Kiruna, to monitor the time-dependent variations of upper thermospheric (200–300 km) and lower thermospheric (90–120 km) winds, using the 630.0 nm and 557.7 nm forbidden lines of OI, respectively. Both instruments used vacuum-sealed etalons of 13 cm diameter with cemented spacers of ‘Zerodur’, providing a velocity reference stable to 10 m s^{-1} . Imaging photon detectors (based on a proximity-focused microchannel plate intensifier and resistive anodes) eliminated the use of pressure or piezoelectric scanning of the etalons and provided a sensitivity increase of a factor of ten over previous ground-based instruments. The high time-resolution data obtained during moderate auroral conditions ($\geq 500 \text{ R}$, allowing 5 min per vector measurement) allows the rapid response of the thermosphere to geomagnetic substorms to be followed in detail. The continuous data obtained from both instruments is being used in conjunction with the University College London global, three-dimensional, time-dependent model of the thermosphere from the region of the mesopause upwards to understand the time-dependent energy and momentum sources of the thermosphere. A comparison of model and empirical data shows excellent agreement when low energy particle sources, concentrated in the auroral oval, are introduced to augment solar UV and EUV heating and polar energy and momentum sources associated with the magnetospheric electric field. During substorms the model predicts the generation of long-duration vortices in the lower thermosphere, but this cannot yet be confirmed by available experimental data. These vortices may have been observed during the latter part of the Energy Budget Campaign when simultaneous observations from Kiruna and Spitzbergen were possible. A joint analysis of these data sets and of the green line 557.7 nm data will be presented in a future paper.

Key words: Thermospheric winds – Auroral heating of upper atmosphere – Fabry-Perot interferometers

Introduction

Ground-based Fabry-Perot interferometers have been used for many years to observe the forbidden oxygen emissions of the upper atmosphere due to either airglow or auroral excitation mechanisms (Armstrong 1956; Chamberlain 1961). Of the many physical and chemical aeronomic quantities which may be stud-

ied on the basis of such observations, the thermal and dynamical structure of the upper mesosphere (OI 557.7 nm) and thermosphere (OI 630.0 nm) are of particular interest, and have been investigated by continuously-improving instrumental and data reduction techniques, particularly during the past twenty years (Armstrong 1969; Hays and Roble 1971 a, b; Hernandez and Roble 1976 a, b; Biondi and Fiebelman 1968; Nagy et al. 1974; Hays et al. 1969; 1979).

Very recently it has been possible to couple together major advances in Fabry-Perot etalon fabrication techniques, the development of imaging photon detectors (IPD) and the availability of low-cost, but powerful microcomputers, to build ‘observatory-class’ instruments which have the stability and sensitivity to measure temperature and vector winds with a time resolution of less than five minutes under mid-latitude airglow conditions. Under moderate auroral conditions (1 kR), a time resolution of less than 1 min can be obtained for a wind vector error of 10 m s^{-1} .

To support the Energy Budget Campaign, two such Fabry-Perot interferometers (FPI) were run continuously from 28 October to 9/10 December 1980, at ESRANGE, Kiruna, Sweden. One of these instruments was used to observe the OI 630.0 nm line, monitoring F-region winds between about 200 and 300 km, while the second observed the OI 557.7 nm line to monitor winds in the lower thermosphere from about 100 to 120 km. Both instruments provided continuous data throughout the night of the C salvo of the Energy Budget Campaign on 10/11 November, and in the one hour period of most intensive auroral and rocket launching activity of the A2 salvo of 1 December 1980. In addition, the data of many other nights of varied geomagnetic activity will be used to provide background information for thermospheric dynamics.

The University College London (UCL) three-dimensional, time-dependent model will be used to complement the ground-based and rocket-borne wind data, with their respective limitations to one height and to a short time period.

Adjustment of the ‘open’ parameters of the model, particularly the magnitude and extent of the polar electric field, auroral electron density and polar and auroral particle precipitation, will be necessary to fit the observed response to particular geomagnetic events. The data which is available from the two ground-based FPIs provides extended time coverage of the winds at 240 km and 100–120 km altitude, complementing the detailed height profiles from the rocket trails and falling spheres. Hence a representation of the complete time-dependent and three-dimensional time structure of the wind values for the N. Scandinavian and European regions can be obtained for each of the

salvos (*C*, *B*, *A1* and *A2*), with one obvious limitation: for the *B* and *A1* salvos, since there were no wind measurements from the ground-based FPIs (owing to cloud cover at ESRANGE), the validity of the model results will be less than for the *C* and *A2* salvos where good ground-based data was available. However, since good ground-based data was obtained on other nights of activity comparable to those of the *B* and *A1* salvos, it is expected that a general predication of thermospheric dynamics can be obtained even for the *A1* and *B* salvos.

Instrumentation

The major advantages of the IPD in a Fabry-Perot interferometer (compared with a pinhole/photomultiplier detector) are:

1. The sensitivity improvement gained by integrating the entire Fabry-Perot image at all times. This improvement is, in practice, rather larger than the ratio of the number of steps per interferogram (20 to 32) divided by the ratio of the quantum efficiencies of a gallium arsenide photomultiplier photocathode to the S25 photocathodes we have been able to use within our present IPDs. Experience with operational photomultiplier and IPD systems would put the IPD advantage at a factor of about 10.
2. An extremely stable, non-scannable etalon can be fabricated and sealed in an evacuated chamber which is itself thermally controlled to about $\pm 0.2^\circ\text{C}$. The use of such a stable cavity considerably simplifies the data reduction procedure since frequent wavelength calibrations are not required. This simplifies the operating procedure of the interferometer, enhancing the already significant sensitivity gain due to the IPD itself.

Figure 1 shows, in schematic form, the optical and mechanical configuration of each of the two interferometers used during the Energy Budget Campaign. The specifications are summarized in Table 1. A computer-controlled scanning mirror system feeds light from a selected region of the sky directly into the 13 cm aperture etalon which is in a sealed and evacuated chamber. The entire first interference ring formed by this etalon is imaged via a Cassegrain telescope onto the photocathode of an IPD (Rees et al. 1980c; 1981) after passing through a 1 nm bandpass filter centered on 630.0 or 557.7 nm.

Table 1. Specification of 'red line' and 'green line' interferometers

Aperture	13.0 cm (coated full aperture except small area near 3×15 mm diameter spacers)
Etalon spacing	10 mm (Zerodur – cemented)
Reflection/Reflective finesse	86% (630 nm) \sim 20
Free spectral range	20 pm (at 600 nm)
Prefilter	1 nm (at 630.0 nm) 1 nm (at 557.7 nm) 60% peak transmission
Detector	Image Photon Detector 18 mm photocathode S20/S25 photocathode
Detected quantum efficiency	at 630.0 nm 4% at 557.7 nm 8%
Max. useful random event rate without pile-up	10 kHz, limited by electronics and A.D. converter
Spectral bins per free spectral range	\sim 50
Effective finesse of complete system	8
Velocity stability of complete instrument:	
Long term drift	$< 1 \text{ m s}^{-1} \text{ day}^{-1}$
Thermal drift	$< 40 \text{ m s}^{-1} (\text{ }^\circ\text{C})^{-1}$
Residual pressure in etalon cavity	40 torr (for tuning Fabry-Perot ring pattern of the IPD)
Thermionic emission rate at operational temperature of 10°C	$\sim 60 \text{ s}^{-1}$ from entire photocathode

All of these optical elements are individually mounted onto four 22 mm diameter steel rods, which are themselves mounted onto an optical bench (Fig. 2). In use, the optical bench is mounted vertically, and the scanning mirror system is mounted

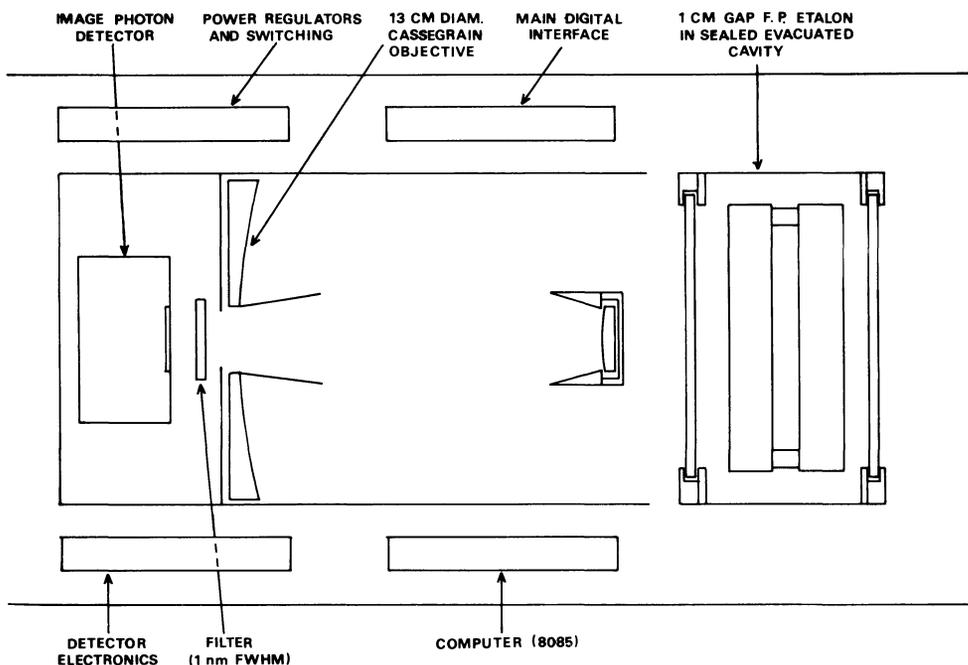


Fig. 1. Schematic configuration of the UCL ground-based Fabry-Perot interferometers used in the Energy Budget Campaign

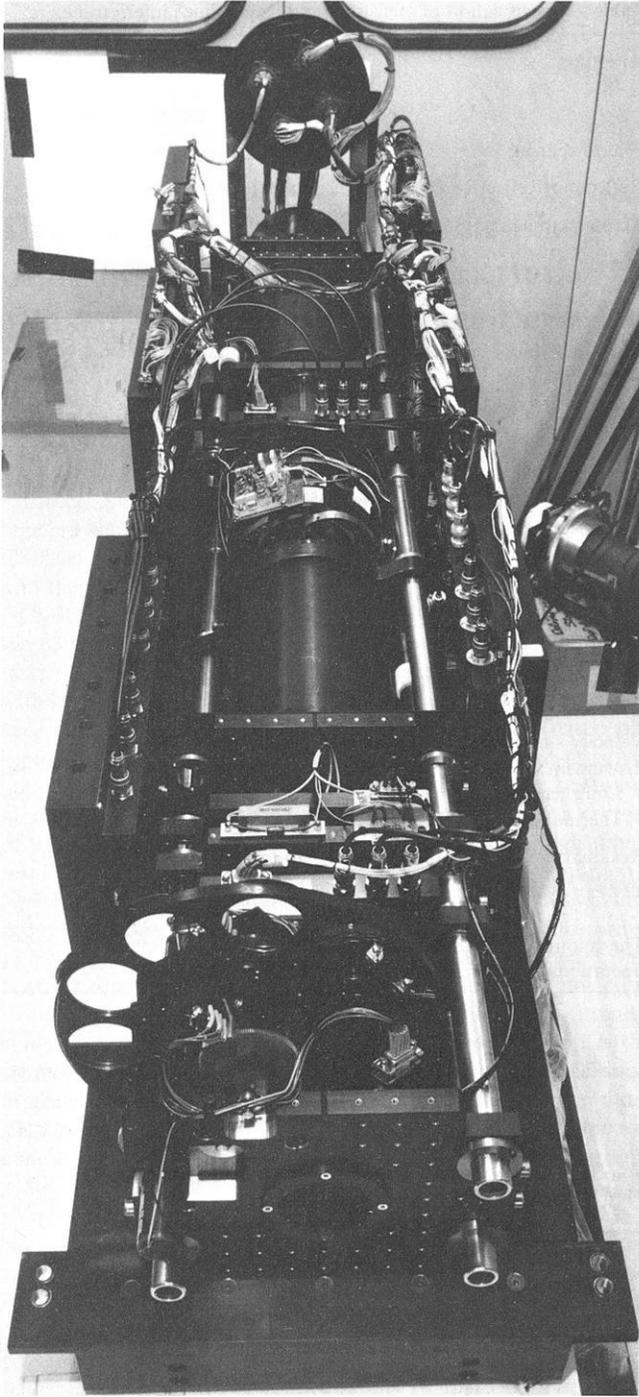


Fig. 2. Photograph of the UCL balloon-borne Fabry-Perot interferometer showing method of construction of the double optical bench arrangement

on top of the optical bench. A light shield is mounted immediately around the main optical components, and several baffle plates are used, both internally and externally, to preclude scattered light reaching the IPD.

The main instrument electronics (Fig. 1) is mounted on the optical bench in five boxes immediately around the sides of the optical assembly, and the entire assembly is completely covered by another light shield. All mechanical components are matt black anodised except for the four steel rods which are chrome-plated to ensure ease of assembly and adjustment. The entire optical assembly is very rigid and is not noticeably affected

by thermal changes within the range of about $15^{\circ}\text{C} \pm 5^{\circ}\text{C}$, except for etalon changes which are avoided by thermal control of the entire sealed etalon mount to $\pm 0.2^{\circ}\text{C}$.

The optical assembly is extremely stable in use – no detectable changes in the image were discerned over 6 weeks observing during the Energy Budget Campaign. One of the systems was demounted as a complete unit at the end of this campaign and transported by car to United Kingdom and has since been operated continuously as a mid-latitude observatory instrument without requiring any readjustment of etalons, optics or electronics.

The etalon construction is based on techniques developed for the NASA Dynamics Explorer Fabry-Perot interferometer (FPI) – a collaboration between the University of Michigan (Professor P.B. Hays) and University College London. The etalon plates are cemented together using three identical spacers made of Zerodur (a ceramic material of extremely low thermal expansion coefficient, made by Schott). The cementing process, which is carried out after the etalon has been tested as an optically-contacted device, used a UV setting cyano-acrylic ‘super-glue’ or a similar material such as ‘Norland 61’, which leaves considerable flexibility to the operator in obtaining a parallelism of better than $\lambda/40$ in the finally assembled etalon. Exhaustive tests have shown that an effective coefficient of expansion of less than $5 \times 10^{-8} (\text{C}^{-1})$ can be achieved by batch selection of the Zerodur. Such an etalon is also extremely stable in both short and long term in respect of drifts and thermal behaviour. The cemented bonds will easily withstand stresses applied to the etalon which (at levels $> 300\text{ N}$) permanently strain the etalon plates. In practice the complete etalon is mounted within a cantilevered unit which allows the fine tuning of the parallelism of the etalon without imparting thermal changes of the mount to the etalon. This mount is also designed to be sealed and evacuated without transmitting the strain of the evacuated outer section to the inner mounting which actually supports the etalon. Finally, this mount provides essential mechanical support to the etalon to withstand shocks and vibration encountered during transportation.

Signal Processing and Data Analysis System

Figure 3 illustrates the general electronic signal processing configuration used to provide on-line interactive colour graphics display of the raw data, and several levels of image processing and analysis. The raw IPD data is accumulated photon by photon in the 8K (kilobyte) RAM (random access memory) of the 8085 microcomputer of the FPI itself. Features such as integration time and view direction are commandable either interactively or automatically. A complete image is transmitted to the main microcomputer – OSI C3-OEM (56K RAM) – at the end of each frame, where a wide variety of image analysis procedures can be called up, again, either interactively or automatically. These include scaling the image for colour-graphic representation (Fig. 4), ‘reduction to radius’ – which is the corrected observed spectrum reduced to wavelength by integrating annular rings of uniform area outward from the geometric centre of the Fabry-Perot ring pattern – with display of the reduced spectrum and many other analytical aids summarized in Table 2.

The mass storage device is a dual 8” diskette (single side, single density) which allows storage of about 70 complete images when running automatically. It is not usually necessary to store the entire image, however, due to the stability of the system, and a more economical procedure of only storing the reduced spectrum plus the associated analysis is normally used. In this mode the data and results of six complete nights of observation,

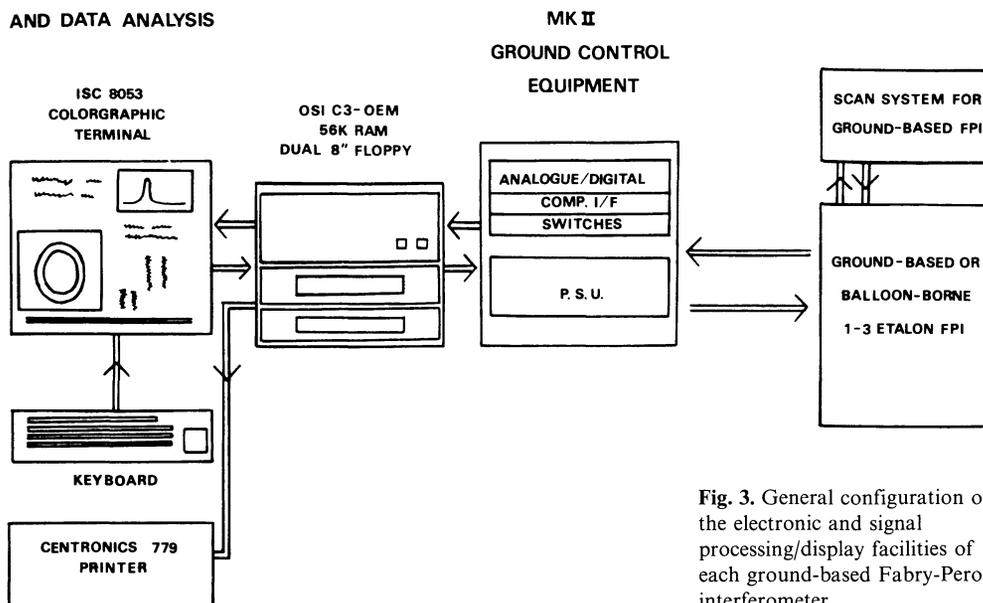
**INSTRUMENT COMMAND
AND DATA ANALYSIS**


Fig. 3. General configuration of the electronic and signal processing/display facilities of each ground-based Fabry-Perot interferometer

Table 2. Summary of interactive and automatic image analysis and display function of the OSI and ISC facility

Function	Parameters	Operation
Image	Background (nn)/scale (mm)	Performs linear conversion of photon image to 8-colour graphics display with variable background subtraction (nn) and intensity level scaling (mm)
Clock	“HH MM SS” on “DD MM YY”	Sets 100 year, real time, clock
Mirror “Direction”	N, E, S, W	Manual direction control of mirror
Mirror rotate after ‘M’ “Direction”	N, E, S, W	Automatic direction control of mirror
Display	H, V, ‘M’	Performs and displays slice horizontally or vertically through image; automatically scales residual in top right hand corner of video display
Subtract		Generalized function which can be used to correct fully for thermionic emission and non-uniform sensitivity of detector and display or to reduce functions
Store 1, 2		Calibration data used as basis for ‘Subtract’
Reduce		For Fabry-Perot operations: reduces and displays the total X, Y , photon image data set as an integrated spectrum vs wavelength; corrects for geometrical properties of X, Y vs R, θ co-ordinates
LSQ		Least square quadratic fit over specified ‘limits’ for quick-look analysis of wavelength or velocity
Analyse		Equal areas fit to data for wavelength or velocity. Useful after generalized correction of photometer data by ‘Subtract’
Time	Seconds	Controls length of integration period from 1 s up to 65000 s
Message	“Any comment”	Note pad
List	All, plot, radius	Used for printer listing of image, reduced data etc.
Disk	On/off drive, track	Specifies disk record (automatic log) of data in interactive or automatic modes
Printer	On/off	Printer command
Countsave	On/off	Stores full image, or reduced spectrum on disk at end of each frame

at two minutes integration time per image, can be stored within the two available diskettes (2×275 K).

A 110 character per second (parallel) printer is used to log all operations (this information is also stored on disk) such as real time clock, frame number, integration time, view direction

and any information which has been stored manually (auroral/weather conditions etc.). Normally, the results obtained from the automatic analysis algorithms are also printed out, with the reduced spectrum, as well as being stored on disk. Subroutines are used for subtracting thermionic emission from the im-

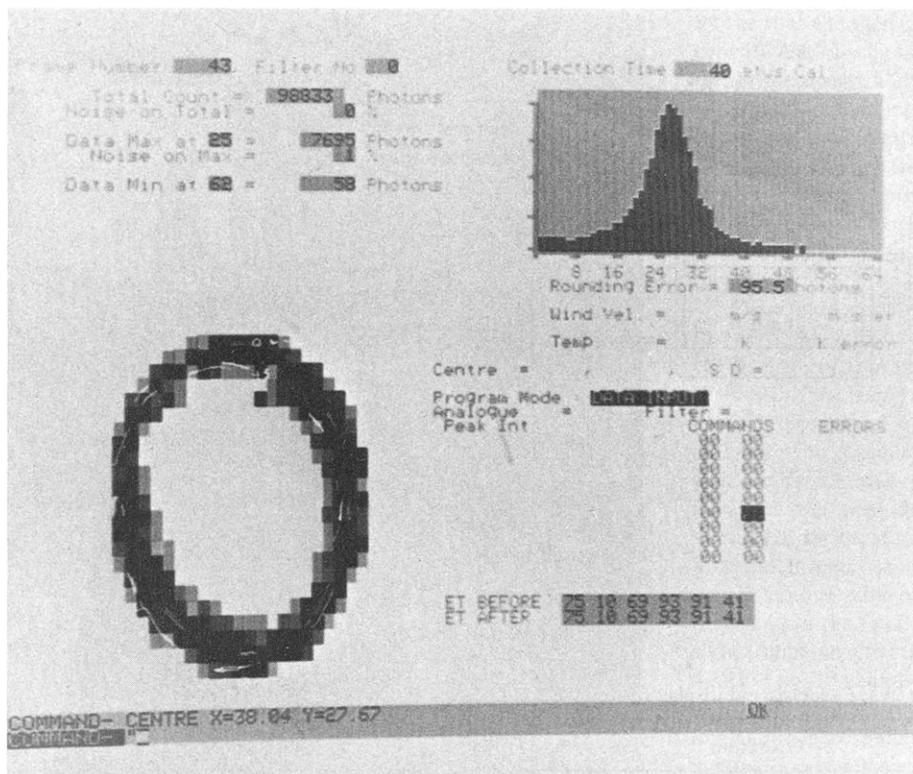


Fig. 4. Photograph of graphic terminal output

age, which may be an important factor when long integration periods are used at times of weak airglow/auroral emissions (i.e. 100 R or less).

The UCL Three-Dimensional, Time-Dependent Model

A three-dimensional, time-dependent thermospheric model has been developed at UCL to follow the complex reaction of the thermosphere to a wide range of both global and local energy and momentum sources (Rees et al. 1980a; 1980b; Fuller-Rowell and Rees 1980; 1981).

The model is entirely self-consistent in treating the neutral gas equations of motion except that, at present, only a single constituent is considered, whose mean molecular weight is a variable, dependent only on pressure. The energy sources considered are solar EUV, using most recent sources of solar flux and EUV heating efficiency of Hinteregger (1979) and Torr et al. (1980) respectively, and particulate heating associated with the polar cap region – of importance during geomagnetically quiet periods – and Joule and particulate heating associated with the auroral oval and which is strongly enhanced during geomagnetic substorms. Momentum sources are the polar cap and auroral oval magnetospheric electric field sources (Heppner 1977), which are related to geomagnetic activity, and a low- to mid-latitude 'dynamo' electric field (Richmond et al. 1980) related to seasonal and solar activity. A global electron density model due to Ching and Chiu (1973) and Chiu (1975) is adapted in an ad hoc manner to respond to geomagnetic auroral activity. The electron density/ionospheric model is not yet fully self-consistent, due to the difficulty of solving the extra equations and the lack of adequate global data with which to test predictions. Basically, the three-dimensional electron density structure is modified to produce realistic auroral electrojets consistent with the polar and auroral electric field structure.

Figure 5 shows a steady-state winter distribution of wind

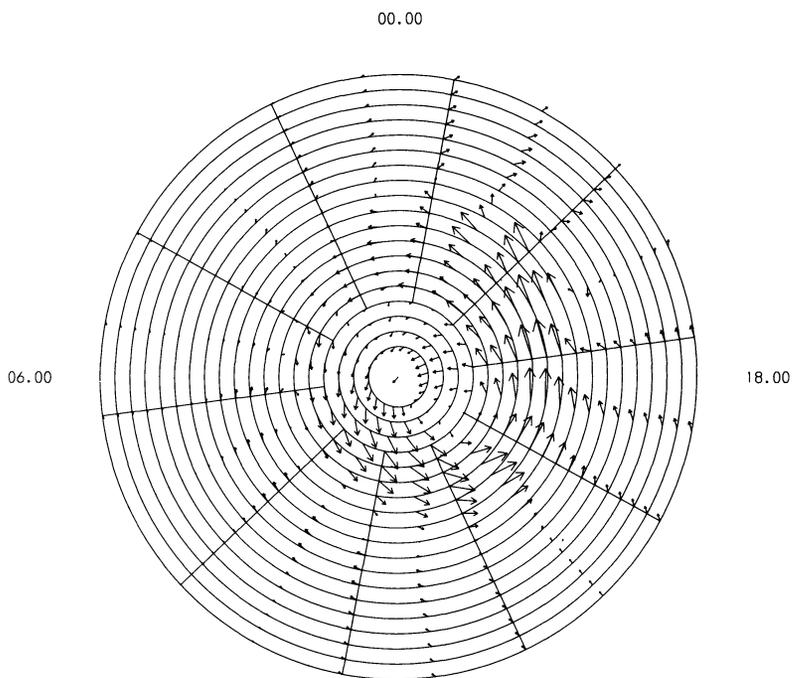
at 120 and 240 km altitude, while Fig. 6 shows the corresponding winds 80 min after the onset of a substorm which would create a 500 γ disturbance under the electrojet in the midnight region of the auroral oval.

The model also predicts the temperature and density distributions at all altitudes from 80 to about 500 km and, in the near future, we expect to be able to extend the calculations to cope with two constituents (O and N₂). In the context of the Energy Budget Campaign, the model will be used to complement and extend the wind data obtained from ground-based and rocket techniques and to calculate transport and dissipation rates of energy, momentum and minor species.

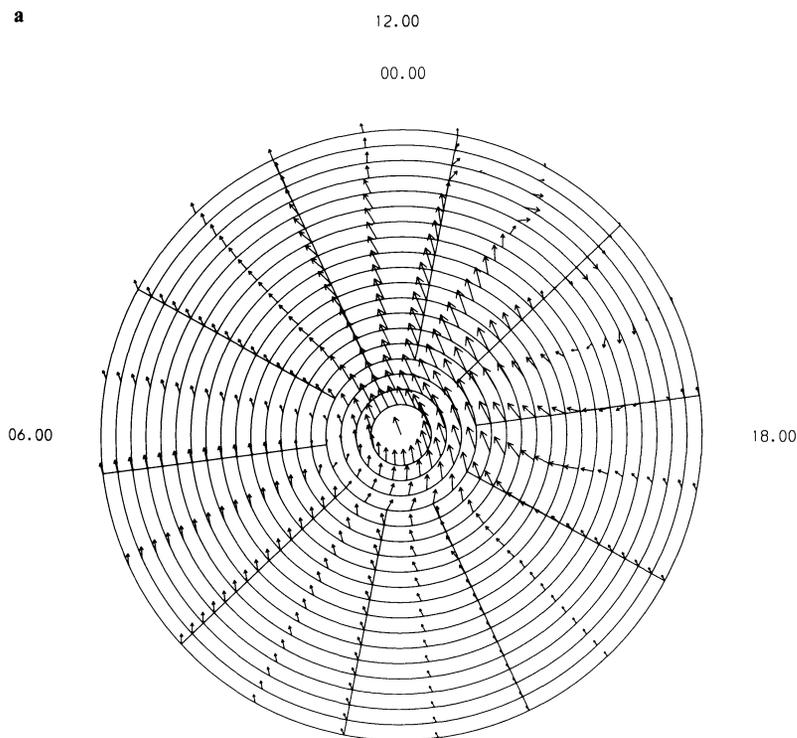
Typical Observations

One or both Fabry-Perot instruments were in operation every night from 30/31 October to 8/9 December inclusive, providing some 350 diskettes of data (100 megabytes total). While the quick-look analysis of this data is carried out and logged in real time, providing a good estimate of the magnitude and direction of the wind vectors, the final analysis of this large amount of data will not be complete for several months.

Figures 7, 8 and 9 show the wind data obtained during parts of the nights of 10/11, 19/20 and 29/30 November, respectively. These nights had substantial auroral and geomagnetic activity so that the instruments were able to make measurements in each of four directions, N, E, S and W, within a period of 6 min. Due to the high time resolution, and the significance of each individual measurement, each of the four measurements is plotted individually as two separate measurements of both the zonal and the meridional winds, looking east and west, and north and south, respectively. Assuming that the mean emission altitude is 240 km, then, at zenith distance, the horizontal separation of these data points is of the order of 800 km, some 7°.5 of latitude, or 1½ h of local time. Also shown in Figs. 7, 8



a



b

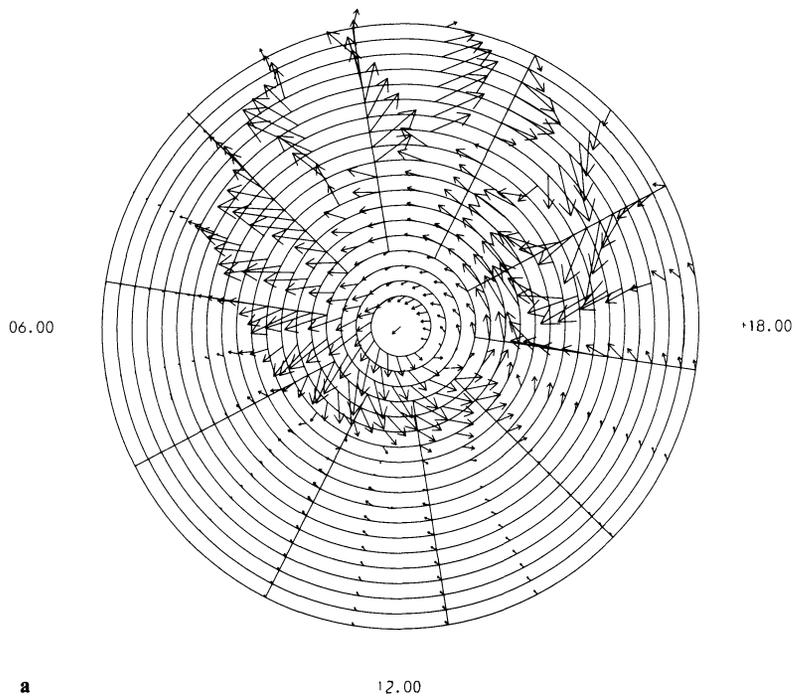
Fig. 5a and b. Wind velocities for the northern hemisphere poleward of 50° N, steady-state conditions at 0030 UT, roughly corresponding to the onset of the substorm of salvos A1 and A2. The Figs. show geographic co-ordinates at 2° latitude and 1.2 h local time spacings, and at altitudes **a** 120 km ($40 \text{ m/s} \equiv 2^\circ \text{ lat.}$); and **b** 240 km ($160 \text{ m/s} \equiv 2^\circ \text{ lat.}$)

and 9 are the steady-state meridional and zonal winds generated by the UCL model for winter conditions, and the relevant geomagnetic and solar activity conditions for the geographic and geomagnetic location of Kiruna.

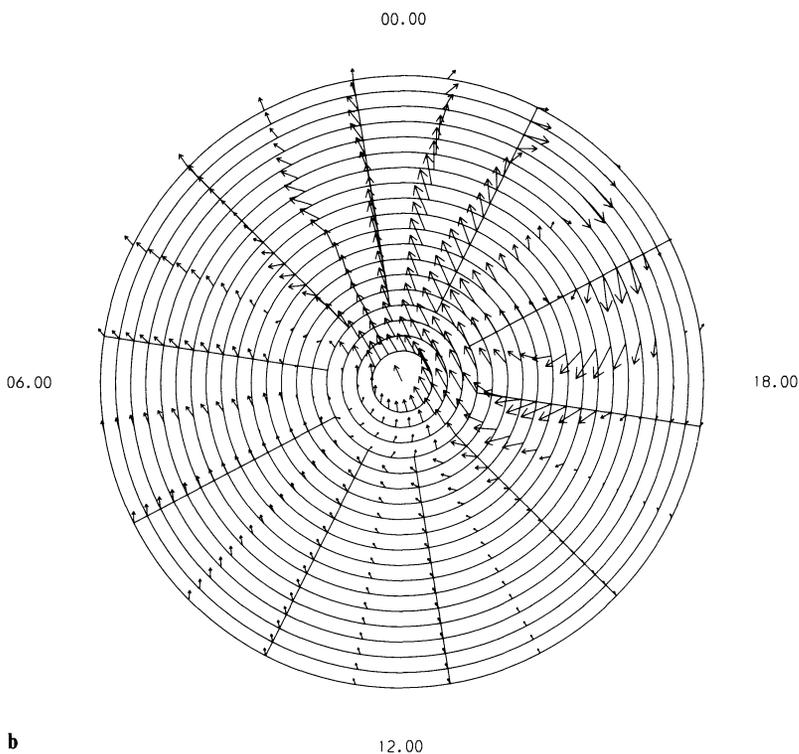
Examination of Figs. 5 and 6 shows that large changes of the horizontal wind are expected over horizontal distances of this magnitude, and Figs. 7, 8 and 9 show that these large differences do exist. Figures 7, 8 and 9 also show that the horizontal gradients of the meridional winds (measured in the north and

the south) and the zonal winds (measured in the east and the west) increased during disturbed periods as would be expected (Figs. 5 and 6). A comparison of the steady-state model and observed winds shows that ground tracks of the observed winds follow the predicted behaviour very well, with the exception of the short-timed perturbations which are associated with the effects of individual or sequential geomagnetic substorms.

One factor which we can estimate from the three-dimensional, time-dependent model is the possible contribution of large verti-



a



b

Fig. 6a and b. As for Fig. 5, but 80 min after the onset of the modelled substorm, approximating to the substorms of salvos A1 and A2, for altitudes **a** 120 km ($40 \text{ m/s} \equiv 2^\circ \text{ lat.}$) and **b** 240 km ($160 \text{ m/s} \equiv 2^\circ \text{ lat.}$)

cal convective F-region winds. At times such as 0200 UT on 30 November 1980, our three-dimensional, time-dependent model predicts that mean and large-scale upward winds near 300 km in excess of 10 m s^{-1} may have been present. Due to the likely large horizontal gradients of the vertical wind and its rapid temporal changes in response to geomagnetic activity, it is not practical to attempt to correct for the vertical wind by means of a single measurement at the local zenith.

The maximum contribution δ of vertical winds to the data

of Figs. 7, 8 and 9 is likely to be less than

$$V_{z\text{MAX}} \cos \phi$$

$$\text{where } V_{z\text{MAX}} < 50 \text{ m s}^{-1}$$

and the viewing angle to the zenith, ϕ , is 60° , so that

$$\delta \approx 25 \text{ m s}^{-1}$$

It is thus of negligible significance in terms of the zonal and meridional wind components at the onset of the 500 γ substorm near 0300 LT on 30 November (Fig. 9a and 9b).

Between 0230 and 0300 LT, as the substorm builds up, with

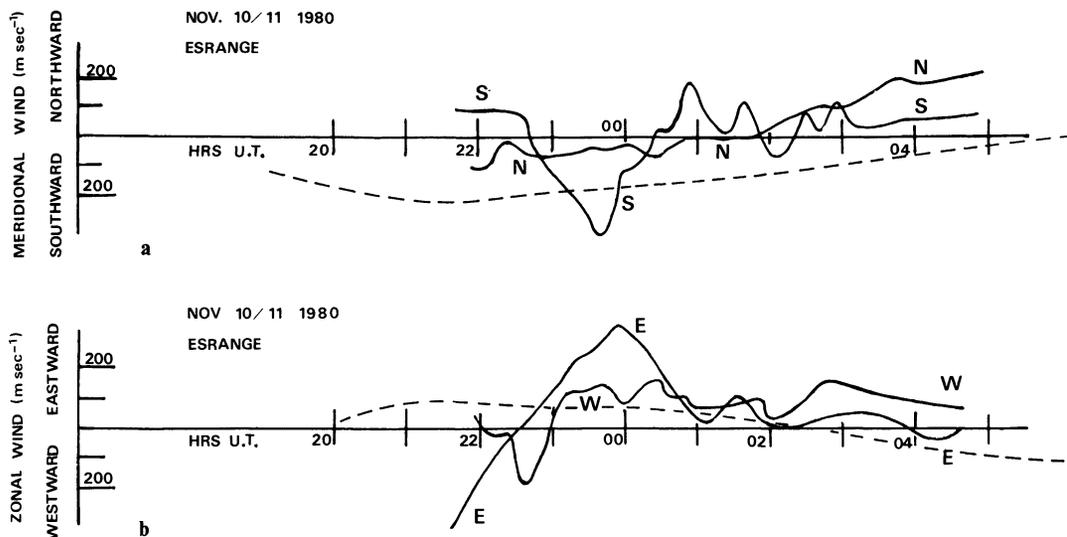


Fig. 7 **a** and **b**. **a** Meridional and **b** zonal measurements about 400 km north (N), east (E), south (S) and west (W) respectively of Kiruna, as obtained directly by scanning the Fabry-Perot at 60° zenith distance. Time resolution between individual measurements in a single direction is 6 min. The dashed line shows the theoretical steady-state wind predicted by the UCL three-dimensional, time-dependent model. Data for 10/11 November 1980

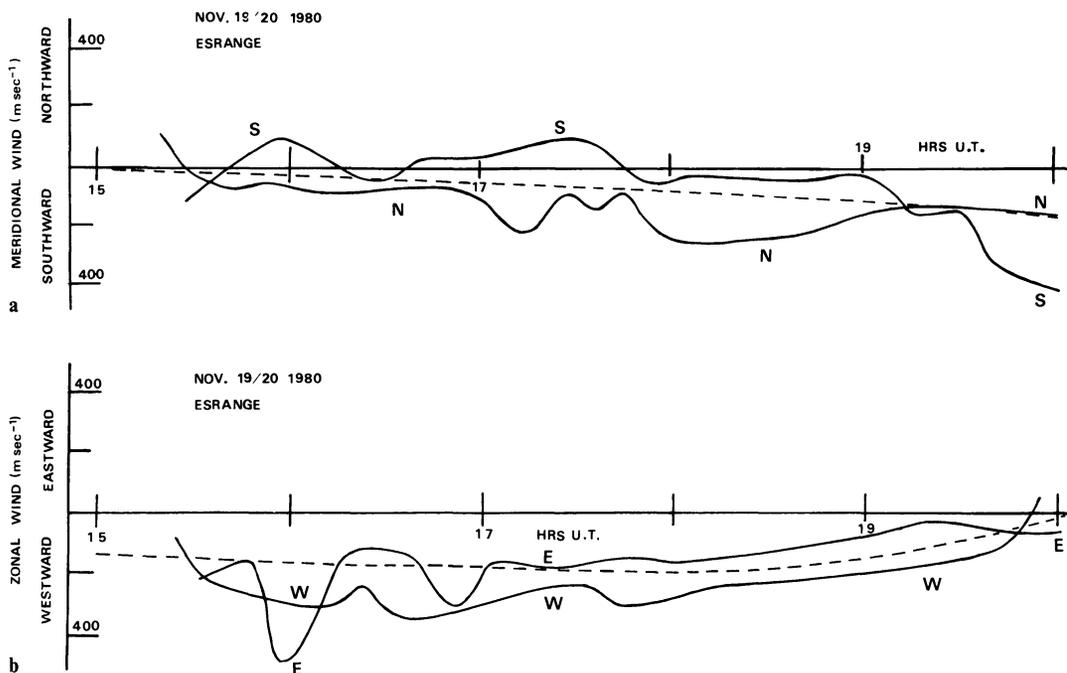


Fig. 8 **a** and **b**. As Fig. 7, for 19/20 November 1980

most of the auroral activity (and therefore heating) to the south of Kiruna, the zonal wind to the east of Kiruna increased from 200 to 330 m s^{-1} . At the same time, while the meridional winds to the south, and the zonal winds to the east changed only slightly, the meridional wind measured to the north decreased from being southward at 300 m s^{-1} , reversed sense, and reached a maximum poleward value of nearly 100 m s^{-1} . As the auroral and magnetic activity decayed after 0300 LT, the zonal wind to the east decreased again, while the meridional wind north of Kiruna again became southward reaching a maximum value of 300 m s^{-1} at 0330 UT.

This one hour period presented a fascinating picture of the

dynamical response of the thermosphere to intense and localized auroral heating, where the high time-resolution of the new instruments unambiguously followed in detail the consequence of the substorm in a way which has not been possible with earlier instruments.

Apart from the rapid and significant changes of zonal and meridional winds which correlate well with discrete substorm activity, general trends are visible in both Fig. 8 and 9. In the early evening, when the meridional winds are usually less than 100 m s^{-1} , there is a strong westward wind driven by the ion drag of the westward convecting ions (northward electric field). By about 1900 to 2000 LT, an equatorward wind becomes obvi-

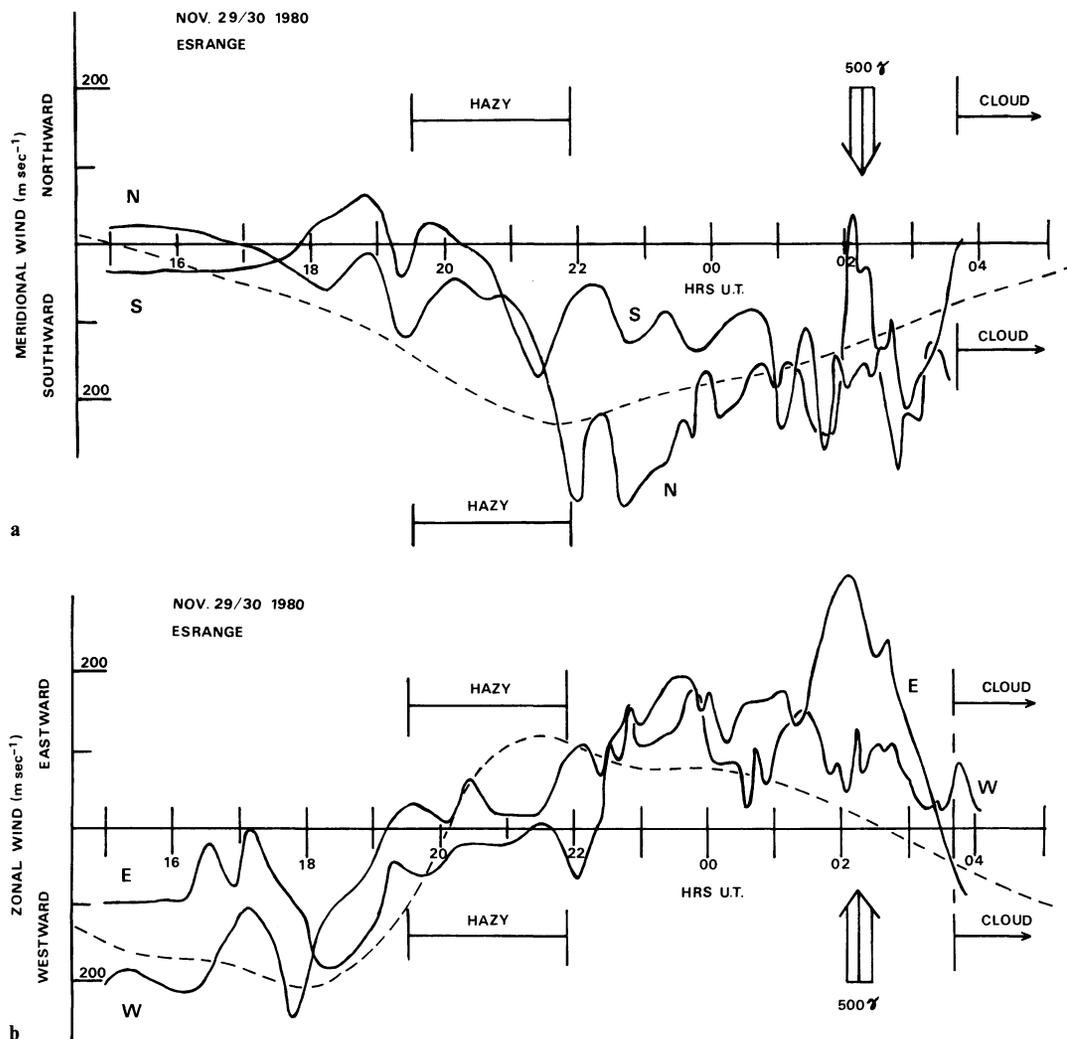


Fig. 9a and b. As Fig. 7, for 29/30 November 1980

ous, given even a modest level of geomagnetic activity, while the zonal wind reverses to become eastward between 2000 and 2100 LT.

Except for these general trends, which correlate in magnitude with the onset of geomagnetic activity, the detailed response of the winds measured respectively N and S, and E and W, of Kiruna, is not well correlated on time scales of up to ~1 h, reflecting the detailed temporal and spatial structure of the auroral momentum and energetic source.

Acknowledgements. The operation of the two Fabry-Perot interferometers at ESRANGE was greatly facilitated by the assistance and facilities made available, in particular, by Mr. Lars Larsson and Mr. Bengt Dahlstadt, and the co-ordinating activities of Professor D. Offermann and the DFVLR management.

We should also like to acknowledge a number of valuable technical and scientific discussions with Professor P.B. Hays and Dr. R.W. Smith in respect of the design and operation of Fabry-Perot interferometers at stations within the auroral oval.

This work was supported by Grants from the United Kingdom Science Research Council.

References

- Armstrong, E.B.: In *The Airglow and the Aurorae*, E.B. Armstrong and A. Dalgarno, eds.: pp 366–373. New York: Pergamon, 1956
- Armstrong, E.B.: Doppler shifts in the wavelength of the IO 6300 line in the night airglow. *Planet. Space Sci.* **17**, 957–974, 1969
- Biondi, M.A., Fiebelman, W.A.: Twilight and nightglow spectral line shapes of the oxygen $\lambda 6300$ and $\lambda 5577$ radiation. *Planet. Space Sci.* **16**, 431–443, 1968
- Chamberlain, J.W.: *Physics of the Aurorae and Airglow*. New York: Academic Press 1961
- Ching, B.K., Chiu, Y.T.: A phenomenological model of global ionospheric electron density in the E, F₁ and F₂ regions. *J. Atmos. Terr. Phys.* **35**, 1615–1630, 1973
- Chiu, Y.T.: An improved phenomenological model of ionospheric density. *J. Atmos. Terr. Phys.* **37**, 1563–1570, 1975
- Fuller-Rowell, T.J., Rees, D.: A three-dimensional, time-dependent, global model of the thermosphere. *J. Atmos. Sci.* **37**, 2545–2567, 1980
- Fuller-Rowell, T.J., Rees, D.: A three-dimensional, time-dependent simulation of the global response of the thermosphere to a geomagnetic substorm. *J. Atmos. Terr. Phys.* **43**, 701–721, 1981
- Hays, P.B., Meriwether, J.W., Roble, R.G.: Night-time thermospheric winds at high latitudes. *J. Geophys. Res.* **84**, 1905–1913, 1979
- Hays, P.B., Nagy, A.F., Roble, R.G.: Interferometric measurements of the 6300 Å Doppler temperature during a magnetic storm. *J. Geophys. Res.* **74**, 4162–4168, 1969
- Hays, P.B., Roble, R.G.: A technique for recovering Doppler line profiles from Fabry-Perot interferometer fringes of very low intensity. *Applied Optics* **10**, 193–200, 1971 a
- Hays, P.B., Roble, R.G.: Direct observations of thermospheric winds during geomagnetic storms. *J. Geophys. Res.* **76**, 5316–5321, 1971 b

- Heppner, J.P.: Empirical models of high latitude electric fields. *J. Geophys. Res.* **82**, 1115–1125, 1977
- Hernandez G., Roble, R.G.: Direct measurements of nighttime thermospheric winds and temperatures. 1. Seasonal variations during geomagnetic quiet periods. *J. Geophys. Res.* **81**, 2065–2074, 1976a
- Hernandez G., Roble, R.G.: Direct measurements of nighttime thermospheric winds and temperatures. 2. Geomagnetic storms. *J. Geophys. Res.* **81**, 5173–5181, 1976b
- Hinteregger, H.E.: Development of Solar Cycle 21 observed in EUV spectrum and atmospheric absorption. *J. Geophys. Res.* **84**, 1933–1938, 1979
- Nagy, A.F., Cicerone, R.J., Hays, P.B., McWatters, K.D., Meriwether, J.W., Belon, A.E., Rino, C.L.: Simultaneous measurement of ion and neutral motions by radar and optical techniques. *Radio Sci.* **9**, 315–321, 1974
- Rees, D., Fuller-Rowell, T.J., Rounce, P.A.: Asymmetric global circulation systems following geomagnetic substorms. *Proc. Vth ESA-PAC Symposium on European Rocket and Balloon Programmes and Related Research*, Bournemouth, 1980. ESA SP-152 June 1980, 81–88, 1980a
- Rees, D., Fuller-Rowell, T.J., Smith, R.W.: Measurements of high latitude thermospheric winds by rocket and ground-based techniques and their interpretation using a three-dimensional, time-dependent dynamical model. *Planet. Space Sci.* **28**, 919–932, 1980b
- Rees, D., McWhirter, I., Rounce, P.A., Barlow, F.E., Kellock, S.J.: Miniature Imaging Photon Detectors. *J. Phys. E: Sci. Instrum.* **13**, 763–770, 1980c
- Rees, D., McWhirter, I., Rounce, P.A., Barlow, F.E.: Miniature Imaging Photon Detectors. II. Devices with transparent photocathodes. *J. Phys. E: Sci. Instrum.* **14**, 229–233, 1981
- Richmond, A.D., Blanc, M., Emery, B.A., Wand, R.H., Fejer, B.G., Woodman, R.F., Ganguly, S., Amayenc, P., Behnke, R.A., Calderon, C., Evans, J.V.: An empirical model of quiet-day ionospheric electric fields at middle and low latitudes. *J. Geophys. Res.* **85**, 4658–4664, 1980
- Torr, M.R., Torr, D.G., Richards, P.G.: The solar ultra-violet heating efficiency of the midlatitude thermosphere. *Geophys. Res. Lett.* **7**, 373–376, 1980

Received June 26, 1981; Revised version October 26, 1981

Accepted November 18, 1981

Zonal Meteor Wind Observations at Budrio, Italy

G. Cevolani and D. Dardi

Laboratorio FISBAT, National Research Council of Italy, Bologna, Italy

Abstract. Intensive meteor radar wind observations were made at Budrio (45°N, 12°E) near Bologna, during the Energy Budget Campaign, November 1980, in the altitude region between 75 and 115 km. After describing briefly the CNR meteor radar station, zonal wind results are presented in terms of prevailing components and oscillations of different time scales: tidal (24, 12 and 8 h), gravity and planetary waves. The monthly prevailing winds differ from those established during the November 1976 and 1978 wind observations and from the CIRA 1972 model for 45°N seasonal variations. Quasi 2-day wind oscillations exhibit amplitudes lower than those obtained at the same station during the summer 1980 observations.

Key words: Meteor winds – Energy Budget Campaign – Tides – Gravity waves

Introduction

The meteor radar station at Budrio (45°N, 12°E) near Bologna, performed initial observations in 1976. Throughout the whole year February 1976–January 1977, recording runs lasting two or more days were made every month with the aim of studying the prevailing and tidal components in the lower thermospheric wind structure (Verniani et al. 1980). By that time, the interferometric system was not completely operational and we preferred to refer all wind data to the height of 95 km, instead of associating each value of the wind with the corresponding diffusive height. We thus obtained information on the seasonal variations of the prevailing wind, of the diurnal, semidiurnal and terdiurnal tides and also of long period oscillations with periods of about 2 days, all referred to this average height.

A new observational campaign was carried out at Budrio throughout the year 1978 with the aim of studying the vertical propagation of gravity, tidal and planetary waves and also of identifying other long period oscillations by means of recording runs lasting many consecutive days per month. Internal gravity waves with periods 2–8 h were observed at Budrio in the region 80–100 km during the spring and summer of 1978 and the results indicate the possible importance of these waves in propagating upwards and transporting momentum from one region to another (even a modest momentum transfer from a coherent wave system can lead to substantial change in the prevailing winds) (Cevolani and Formigini 1981). Determination of the vertical structure of wind components is made possible by the use of an interferometric system which yields accurate values for the height of each echo. The following objectives were achieved

during the 1978 wind observations at Budrio: (a) a great number of usable meteor echoes per day ($\geq 1,000$) and (b) reliable altitude definition (± 2 km).

Current work at Bologna is focussed on the zonal propagation of planetary waves and their seasonal variations. Of particular interest is the latitudinal structure of the quasi 2-day oscillations which we are investigating this year in the form of simultaneous observations together with Sheffield University, in England. The 1978, 1979 and 1980 summer data have been examined by the two different groups to gain further information on the nature of these free atmospheric modes.

Other arguments of present interest in Bologna involve the study of the atmospheric effects on the parameters of echoes. A recent paper dealing with the effects of the lower-thermospheric winds on the echo duration has been published (Cevolani and Hajduk 1979). Moreover, systematic flux measurements of individual meteor streams and a comparison of the results at different stations – Budrio (Italy), Ondřejov (CSSR) and Dushanbe (USSR) are being performed jointly. The first results of this cooperative exercise are included in another recent paper (Hajduk and Cevolani 1981).

Equipment and Data Processing

Equipment

In this section a brief description of the Bologna meteor radar is given.

Verniani et al. (1974) described the Bologna CNR meteor radar in detail and subsequently Verniani et al. (1980) presented the early results obtained during the 1976 wind observations carried out at the Budrio station. In this last paper the status of the meteor radar station before April 1978 was described. The system was then improved to measure height directly. The original meteor recording system was based on a photographic device, and meteor heights were estimated from the decay time of the echo amplitudes. However, with the pressing need for good height resolution when investigating propagating wave structure, a seven aerial interferometer system was set up.

The equipment situated at Budrio is a coherent pulse Doppler radar and essentially consists of a transmitter with a directional antenna, a receiving system with an interferometric antenna and a real-time echo processor. The radar transmitter operates at 42.7 MHz, with a peak power of about 200 kW. The pulse duration is 10 μ s and repetition frequency 140 Hz. The transmitting antenna consists of a linear array of five aeri-als, each made up of three half-wave dipoles; it is directed eastwards and its

direction of maximum radiation has an elevation of 45° . The beam width is $30^\circ \times 30^\circ$ at 3 dB points. The receiving system is composed of seven directional antennas, seven receivers and a phase-sensitive product detector. The interferometric system has been built up for determining the height of the reflection point on each meteor trail. The height is determined by the range of the reflection point (deduced from the echo decay) and by the zenith angle which, in turn, can be calculated by the phase differences between the signals received by the various antennas. The real-time processor (Schaffner 1966) first discriminates the meteor echoes from among all the received signals, then extracts the desired information from the selected echoes and records it on magnetic tape.

At the beginning of 1979, a second, alternative three-aerial interferometer directed eastwards was constructed and, at the end of 1980, another analogous interferometric system directed northwards was built up, so that in the immediate future the total wind vector may be derived. Consideration of calibration of the two interferometric systems performed in all observing directions, and meteor echo results, suggest that the uncertainties in echo height and radial velocity are approximately ± 2 km and $\sim 3 \text{ ms}^{-1}$ for individual echoes. For the following results, it is first assumed that the wind values depend only on time and altitude, and represent the component of the horizontal wind along the main axis of the aerals. It is further assumed that echoes exceeding $\pm 15^\circ$ from the mean axis are eliminated, minimizing the error introduced by the quadrature component.

Data Processing

The measured parameters include range, Doppler frequency, radial velocity, zonal velocity at 95 km, "real" zonal velocity, interferometric height of echo, decay time of echo, diffusive height of echo, and corresponding errors. Zonal velocities at 95 km are calculated by considering the mean height of 95 km in the relation $u = u_r \cos[\arccos(h/r)]$ (Verniani et al. 1980) where u is the zonal wind velocity, u_r the radial velocity, h the height and r the range. "Real" zonal velocities are calculated by using the corresponding interferometric height of echo. Many of the methods used to process meteor wind data assume periodic models and fit the models to the data in some minimum error sense. In order to ensure a reasonable fit, the raw wind samples for each meteor trail are used to produce a uniform time series. Hourly averages of the wind are computed for every hour from the values of the zonal wind and are assigned a weight. This series is frequency analysed to determine which components are appropriate to be included in the model.

In the data processing, we use four main programs. The first program calculates the echo heights and zonal wind velocities after determining the distance of the echo and the radial wind velocity. A second program sorts the echoes into height intervals and places this information into several files, then reads from these files and calculates hourly wind weighted values for each height interval. Data are generally sorted into both seven and four height intervals: the first solution gives better height resolution but the number of echoes contributing to each hourly point is lower and therefore the reliability is reduced. The third program calculates the amplitudes and phases of the different wind components by using a weighted least-squares method and the frequencies of these waves by using the conventional spectral analysis (autocovariance function or discrete Fourier transform). The same program, although referred to as a spectral analysis method, produces an amplitude periodogram. The resolution of the discrete Fourier transform is essentially equal to the recip-

rocal of the length of the total time sample. Thus, a fairly long time sample is needed to determine the shape of a given spectral component. The fourth program also performs a spectral analysis by using the "maximum entropy" method (MEM) which is mainly employed when short time series are encountered. The MEM spectra have a much higher resolution and are capable of following a spectral component with a shorter total time sample. Also, long period components can be studied with shorter and less costly observing runs. The main shortcoming of the MEM spectral estimate is the lack of quantitative method of determining the length of the prediction error filter. It is known from experience that a too short length results in a smoothed estimate, whereas an excessive length introduces spurious detail into spectrum.

The results thus obtained indicate that the zonal wind can be split up into the prevailing wind, three oscillations which can be interpreted as a diurnal, a semidiurnal and a terdiurnal tide and into other perturbations that include both short period waves (gravity waves) and long period oscillations (planetary waves) (Verniani et al. 1980).

Results

The present data refer to the 7–30 November 1980 observing period. The data of November's first week were not too successful with interference and several breakdowns, mainly late in the morning (08–13 h LT), and are not included here. Interference during October and November 1980 represented serious problems for our observations at Budrio, but nevertheless, about 19,000 usable echoes have been collected during the whole November period. Data were also recorded during several periods in 1980, 12–26 March, 4–18 June, 16 July–4 August, 3–16 September, 16–31 October, but since the data reduction for these periods is not fully completed, only occasional reference is made to some relevant results mainly in connection with long period wind oscillations. Because of the prominence of periodic fluctuations in the data, all data time series were subjected to harmonic analysis in which the steady component represents the prevailing wind, probably including unresolved amplitudes of wind fluctuations whose periods are longer than the time series. Data on atmospheric oscillations are presented in the form of the amplitude and phase of a particular periodic component. Following the usual convention, zonal wind amplitudes are positive towards the east and phases refer to the local time at which the zonal wind amplitude is maximum (in our case, this occurs in the east direction). Since most methods of spectral estimation require uniformly spaced data, a long data time series (7–30 November) of horizontal velocities averaged over the 75–115 km height interval (average 95 km) has been used for a monthly comprehensive study. Remaining gaps (comprising 16 non-consecutive points in the time series) have been filled by linear interpolation. These gaps did not cause any significant uncertainty in the spectral estimates. In order to deduce wind profiles of main zonal wind components for periods with a consistent daily meteor flux, particular attention was drawn to some days with high meteor rate by considering also the incidence of large wind amplitudes. In fact, the daytime meteor echo rate at Budrio was generally below the monthly mean for the November month on account of multiple ionospheric-ground scatter due to intense solar activity during the Energy Budget Campaign 1980. The potential superiority of the MEM method over other spectral estimators, in particular for short data lengths is recognised by other authors (Ulrich and Bishop 1975). This method has been used successfully in the analysis of individual days' data in order to study in closer

7 - 30 NOVEMBER 1980

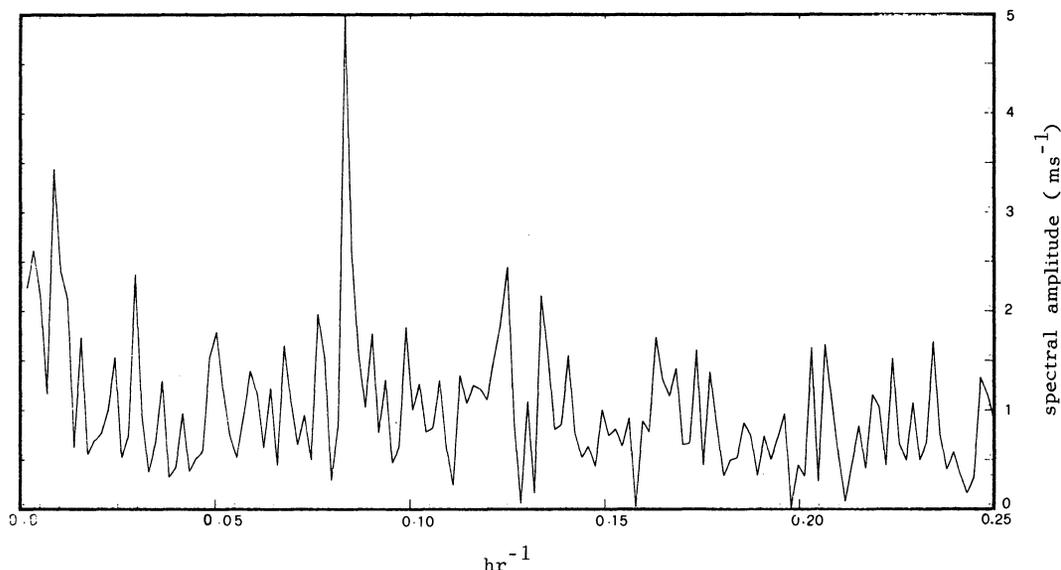


Fig. 1. Amplitude spectrum of the zonal wind at Budrio for the 7-30 November 1980 period for the 75-115 km height interval (average 95 km)

detail short-period wind components in connection with gravity wave propagation (Stening et al. 1978). Significant variations, even in brief periods, have been pointed out in the structure of the zonal winds observed at Budrio in the November 1980 period and have been attributed to the presence in the meteor zone of gravity waves, possibly with a period of a few hours, superimposed on the diurnal trend of the wind. The nature of gravity waves suggested to us that we should select days during which the amplitudes of the irregular wind component were consistent and this, taking into account also the requirement for periods with an adequate hourly meteor flux at different levels between 80 and 110 km, took place on only a few days. MEM, used for these short periods, thus allowed us to study gravity wave propagation associated with velocity-altitude profiles. MEM spectra have been obtained for data sorted into seven uniformly spaced height intervals with sampling at 5 km intervals between 80 and 110 km, whereas longer data series could not be extracted from the total November data length, in order to use the Fourier transform method reliably.

The long time series over the quoted height interval (average 95 km) has been subjected to spectral analysis by using the discrete Fourier transform: the corresponding amplitude spectrum is presented in Figure 1.

Figures 2a-d show the 1976 and 1978 monthly wind results obtained at Budrio. These results are compared with the present ones and thus allow us to study the variability of the prevailing wind and of the diurnal, semidiurnal and terdiurnal wind harmonics in different years.

Two short unsmoothed time series (23-24, 28-29 November 1980) have been subjected to the MEM method. The spectra obtained with a filter length of $\frac{2}{3}n$, where n is the data length, allow us to obtain profiles of the prevailing wind, gravity waves ($T=3.5, 4.5, 6$ h) and tides ($T=8, 12, 24$ h) (Figs. 3 and 4). A filter length of $\frac{1}{3}n$ results in a smoothed estimation obviating the resolution advantages of MEM, while a filter length of n introduces spurious detail into the spectrum. Since the number of data points (crosses in Figs. 3 and 4) is limited, a linear interpolation has been preferred to any other fitted polynomial

curve. The amplitudes and phases of the zonal velocity components for the seven height intervals between 80 and 110 km and for the two-day periods were subjected to harmonic analysis, a least squares fit being made to the data with a prevailing or mean component, 24, 12 and 8 h periodic components, plus velocity fluctuations at 3.5 and 6 h periods. Sufficient data were available to allow this analysis to be undertaken for all heights and for the two short time series.

The Fourier transform results indicate a slightly dominant 12-h oscillation, but the amplitudes of all the main components were unusually low with respect to the theoretical predictions and the previous results obtained at Budrio in different years. Several of the amplitudes are of a similar magnitude to the r.m.s. velocity uncertainty (3 ms^{-1}). MEM has also been applied to the long velocity time series and the resulting spectra confirm the periods and magnitudes revealed by the Fourier transform method.

Prevailing Wind

The amplitudes of the mean zonal flow at 95 km of altitude, as provided by using the discrete Fourier transform are low (of the order of a few ms^{-1}). The average prevailing components from the sample, including all altitudes (average 95 km), are compared with the corresponding ones for November 1976 and 1978, together with the monthly variations of the prevailing wind in the two different years and the values interpolated from the CIRA (1972) model at 95 km for 45°N latitude. As is shown in Fig. 2a, our 1976 and 1978 measurements and the CIRA model exhibit similar trends with a minimum in April, when the model lists a negative zonal wind. Though the CIRA model exhibits two minima corresponding to equinoctial months, the low amplitudes recorded in the November 1980 appear to be atypical of this month.

The average prevailing component variation with altitude for two short time series (23-24, 28-29 November 1980), has been obtained by using the MEM method and is shown in Figures 3a and 4a. The direction of the zonal flow in the two

closely-spaced periods changes slightly with altitude and very small amplitudes are generally observed. At 90 km altitude, the amplitudes are negative; in the 95–100 km region, the magnitudes have the same values in the two different periods and are in close agreement with the amplitudes extracted from the spectrum of Figure 1 relative to the 75–115 km height interval obtained for the 7–30 November 1980 period.

Tides

Regular wind observations carried out at Budrio in the 1976–1980 year period, have revealed not only systematic variations of the amplitude and phase of the semidiurnal, diurnal and terdiurnal tides throughout the year but also often show little consistency between sets of data obtained during the same month in different years or even between observations in one month. Information on the structure of tidal modes present in the wind patterns of November 1980 are derived from the profiles of Figure 3d, e, f and 4d, e, f, which relate to the 23–24 and 28–29 November 1980 periods, using the MEM method. It has become customary to try to identify which tidal mode is present in any profile by measuring the vertical wavelength of the tide, and this is possible by plotting the phase of the tide as a function of altitude. A downward phase propagation is consistent with an upward propagation of energy.

Semidiurnal Tide. In our November 1980 velocity spectrum for an average height of 95 km (Fig. 1) the semidiurnal tide exhibits a well defined peak. However, the observed amplitudes are low and differ markedly from those observed in November 1976 ($\sim 30 \text{ ms}^{-1}$) though they are closer to the November 1978 values ($\sim 7 \text{ ms}^{-1}$). In addition, the semidiurnal tide profiles obtained for the above mentioned short periods show small amplitudes in the overall 75–115 km region (Figs. 3e and 4e). Our results appear to be consistent with modes of short vertical wavelength ($\lambda \cong 30 \text{ km}$) corresponding to the S_2^1 mode (Fellous et al. 1975). However, an irregular behaviour is often found in the 7–30 November period. These irregularities can be seen over several adjacent (in space and time) estimates of amplitude and phase, so that the effect is only partly due to data unreliability. Many causes may be responsible for this irregular behaviour: (a) changes in the source (Bernard 1981); (b) the presence of more than one tidal mode leading to interference, or reflections leading to standing waves (Fellous et al. 1975); (c) coupling of energy from one mode into another or into the background flow (Spizzichino 1970). In the presence of background winds, shorter wavelength propagating modes are expected to be enhanced in the meteor region by mode coupling. The presence of superposed modes, possibly propagating in both directions can alter the apparent phase gradients (Poulter 1980). Variations in time of the semidiurnal tide over scales greater than 5–6 days as observed at Budrio and described hereafter, suggest that the most probable cause of these irregularities may be changes in the source (ozone and water vapour distribution) or propagation conditions. In order to study in closer detail the amplitude variations of the semidiurnal tide during the November 1980 period, an inverse Fourier transform has been calculated using a window centred on 12 h of proper bandwidth (Kaiser et al. 1979). The inverse transform thus indicated how the amplitude varied throughout the recording period. An inverse Fourier transform centred on the 12 h peak has been calculated for the 75–115 km height interval (time is relative to 13 hours LT, 7 November 1980) and indicates that this tide is modulated in amplitude throughout the recording period, varying over time scales of about 5 days (Fig. 5).

Diurnal Tide. The diurnal tide represents a small part of the total wind energy: the amplitudes derived from the velocity spectrum of Figure 1 are very small. A mode at 20 h period has been extracted in the spectrum, with a very low amplitude. It is likely that these amplitudes are much reduced by the averaging over the whole 7–30 November period, due to the strong phase changes which occur in this month (Fellous et al. 1975). Moreover, the non linear vertical phase variation of this tide can mean that different modes are present simultaneously during the quoted transitional period (Fellous et al. 1975). However, individual days' data presented in Figures 3f and 4f, exhibit larger amplitudes than those extracted from the spectrum of Figure 1: the consistent amplitudes observed at 100 km (13 ms^{-1} on 23–24 November and 18 ms^{-1} on 28–29 November 1980) indicate the possible importance of significant short vertical wavelength propagating modes. The individual phase profiles for the two quoted periods demonstrate rapid changes near 90–100 km which are consistent with a vertical wavelength, $\lambda = 20 \text{ km}$ (S_1^1 mode). Above these heights the phase progression is generally less, as would be expected of evanescent modes. Similar irregularities in phase progression emphasize the possible importance of non-linear interactions between the diurnal modes and other wind components (Spizzichino 1970).

Terdiurnal Tide. The amplitude spectrum of Figure 1 exhibits a peak with a period near 8 h. As for the other two tides, the amplitudes derived for the whole observing period are smaller than those recorded in the individual examples quoted (Figs. 3d and 4d) and this is probably due to the lack of coherence of this oscillation. The two profiles of Figure 3d and 4d show very similar trends below 100 km and the relative amplitudes do not exceed 10 ms^{-1} . Phase gradients seem to be consistent with wavelengths of 30–40 km. These values do not agree with the theoretical vertical wavelength of a directly excited terdiurnal mode which must be very large, but they are in close agreement with previous observational results (Glass and Spizzichino 1974). On the other hand, the amplitudes of this mode are quite similar in the years 1976 and 1978 (Fig. 2d) and exhibit the theoretical annual variation, with a reversal around the equinoxes (Glass and Fellous 1975). The recurrence of this harmonic in the amplitude spectra of 4-year observations at Budrio and the existence of a 8 h oscillation in ground level data, support the suggestion that it is a tidal harmonic.

Gravity Waves

Gravity waves with periods between 2 and 8 h were previously observed in 1978 by the meteor radar at Budrio (Cevolani and Formigini 1981). The summer 1978 amplitude spectra of the zonal wind exhibited significant peaks with periods of 3, 4, 5 and 6 h and the phase of the corresponding oscillations generally increased with altitude. These oscillations then propagated vertically with a downward phase speed as expected theoretically for gravity waves. Examples of gravity wave phase propagation observed at Budrio in the summer of 1978, are presented in Figure 6. Typical vertical wavelengths of gravity waves are also represented. In the 1978 results, the gravity wave energy per unit volume increased with height. As a possible explanation of this behaviour, Spizzichino (1970) proposed the existence of non linear interactions between gravity waves and tides. Moreover, the summer 1978 data emphasized the importance of wave motions as sources of momentum in the upper atmosphere and even a modest momentum transfer from a coherent wave system was shown to lead to substantial changes in the prevailing winds.

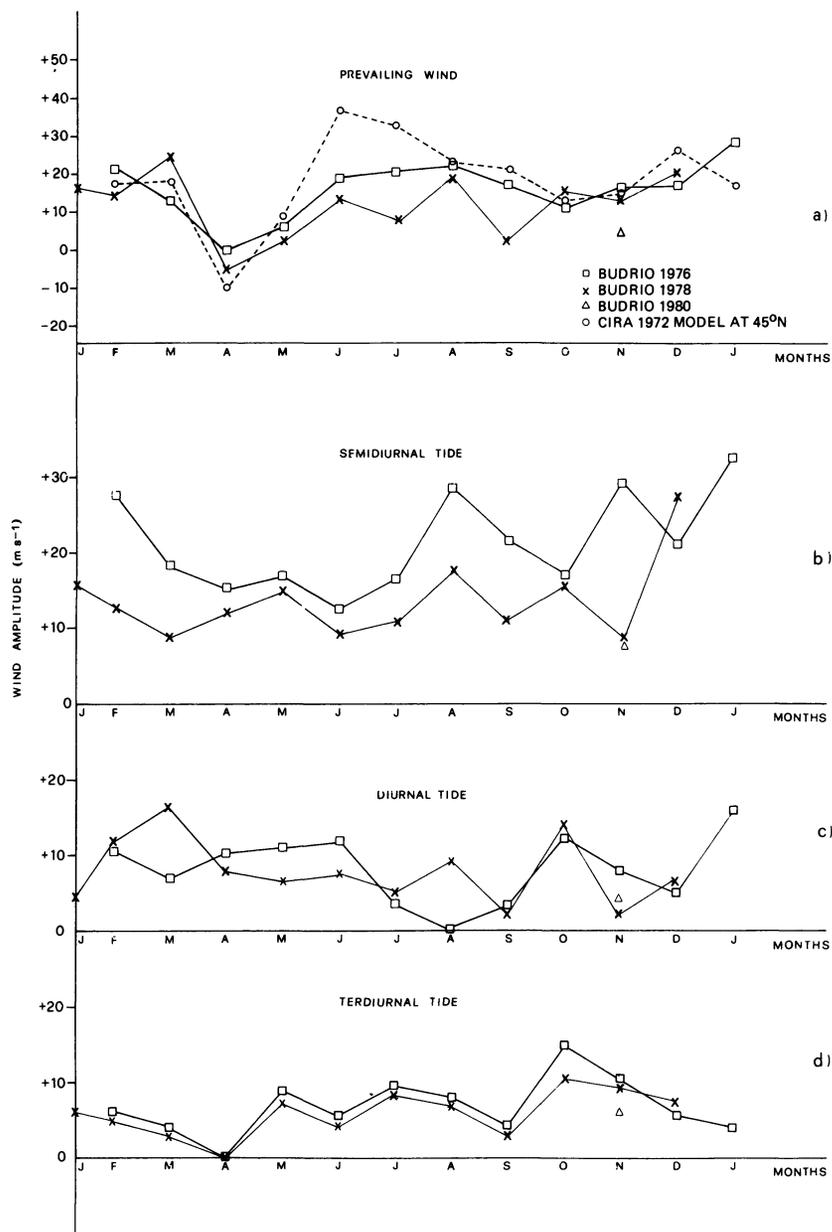


Fig. 2a-d. Seasonal variations of a the prevailing wind; b the semidiurnal tide; c the diurnal tide; d the terdiurnal tide in different years at Budrio

The present hourly rate (10–150 echoes) of the meteor radar of Budrio does not allow us to identify gravity waves with periods smaller than about 1–2 h. Individual days' data (23–24, 28–29 November 1980) have been subjected to spectral analysis using the MEM method for seven different uniformly spaced height intervals in the 75–115 km region. The profiles of Figure 3b, c and 4b, c indicate the possible importance of gravity waves propagating upwards with amplitudes larger than those of the other main wind components. At 80–85 km, amplitudes of 14–16 ms⁻¹ are present in different profiles, associated with gravity waves with periods of 3.5 and 4.5 h. These examples show that the energy densities of gravity waves can become dominant in the lower thermosphere. The phase gradients derived by performing a least squares fit to the individual days' data, allow us to obtain vertical gravity wavelengths of 20–45 km with periods between 3.5 and 6 h. Amplitudes are of the order of 10 ms⁻¹. The derived wavelengths are typical of gravity waves and indicate that these waves propagate upwards with a phase velocity of

the order of 1–2 ms⁻¹. These results appear to be in close agreement with those obtained for the 1978 wind campaign.

Long Period Oscillations

By using the discrete Fourier transform we have found several long period oscillations in the zonal wind measured at Budrio in the 7–30 November 1980 period for the 75–115 km height interval (average, 95 km, Fig. 1). As shown in this spectrum, the periods of such oscillations are usually of 2–3, 5–6 and about 10 days.

The presence of long period oscillations in the meteor zone with quasi 2-day periods, has been noticed in the Budrio recording runs since long term observations began in 1976. Quasi 2-day oscillations were detected in the quoted height interval in our November 1980 observations, with an amplitude of about 2–3 ms⁻¹. Some indication of the presence of such modes has been found in our data time series for summers of previous

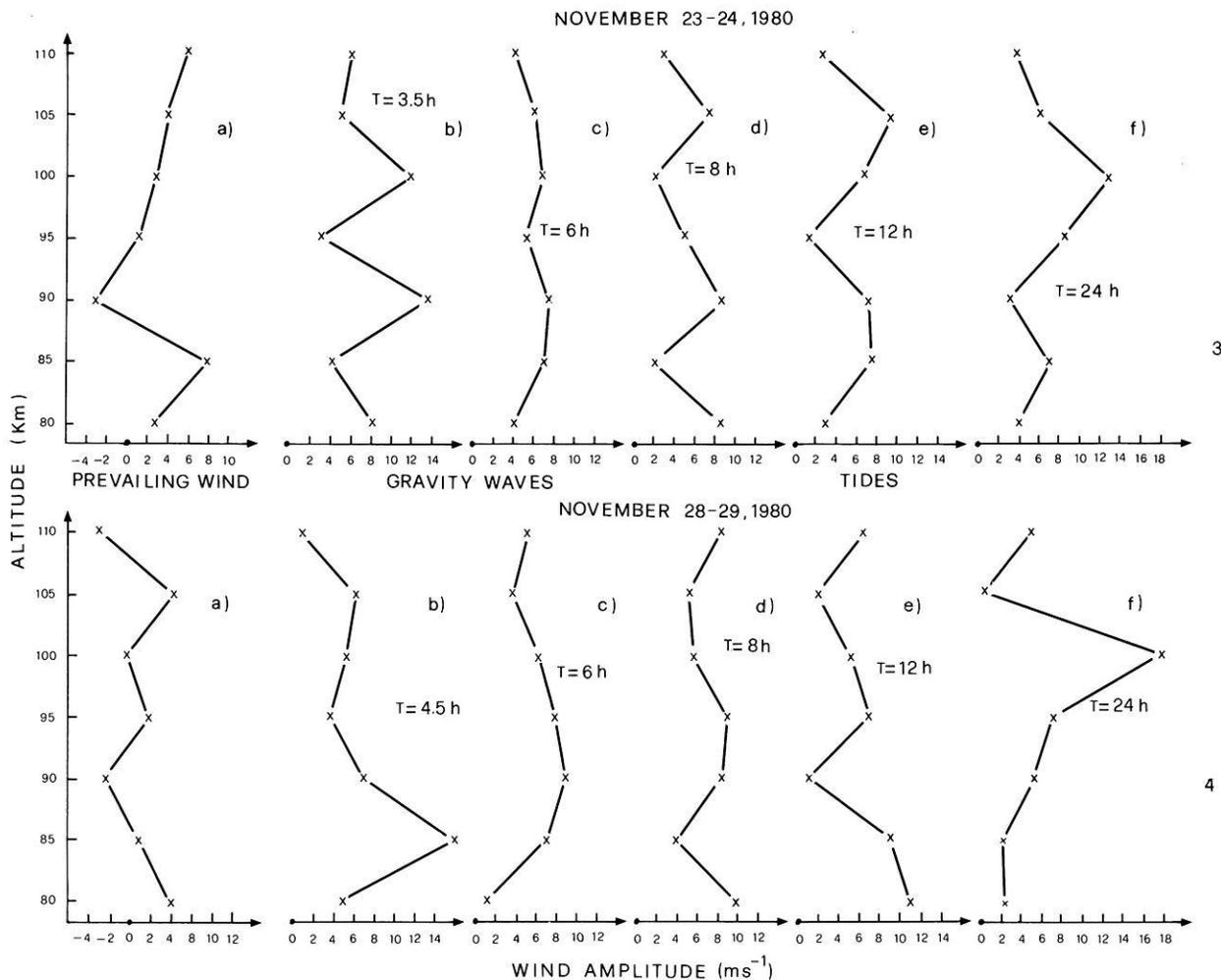


Fig. 3a-f. Altitude profiles of the prevailing wind, gravity waves and tides for the 23-24 November 1980 period, at Budrio

Fig. 4a-f. Altitude profiles of the prevailing wind, gravity waves and tides for the 28-29 November 1980 period, at Budrio

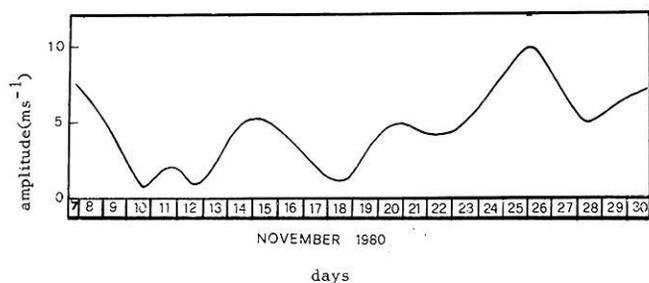


Fig. 5. Inverse Fourier transform of the 12 h peak in the Fourier transform with a bandwidth of 11.8-12.4 h for the 75-115 km height interval. Time is relative to 13.00 hours LT, 7 November 1980

years obtained at an average height of 95 km: in the spectra of 13-16 July and 24-26 August 1976 two wind oscillations of 51 and 40 h periods and of 7 and 10 ms^{-1} amplitude respectively were isolated (Verniani et al. 1980); in those of 11-14 July and 20-24 August 1978, two significant peaks of 51 and 60 h periods and of 5 and 13 ms^{-1} amplitude have been identified; moreover, in the 26 July-5 August 1979 spectrum, a very significant peak of 43 h period and of 17 ms^{-1} amplitude has recently been pointed out (well defined peaks appeared having

about the same phase in three different height intervals, 75-90 km, 91-104 km and 105-120 km) (Cevolani, in press 1982). Recently, in a cooperative exercise with the University of Sheffield, two peaks with periods of about 43 h and 60 h have been detected in the wind amplitude spectrum for the 16 July-2 August 1980 period at an average height of 95 km (Fig. 7). As can be seen, an important feature associated with these quasi 2-day oscillations is the variability from year to year of the precise period. The existence of such oscillations in the meteor wind has now been confirmed by observations at various locations across the globe and it is also known that considerable amplification of these waves takes place about a month after the summer solstices (Muller and Nelson 1978). The latitudinal structure of these is scarcely known and there is not yet significant evidence for the existence of global waves of such period at meteor altitudes. A series of observations from satellites have clearly pointed out wave patterns in the temperature fields at stratospheric and mesospheric heights in summer (Muller et al. 1979). A westward travelling wave, zonal number 3, whose period is close to 2 days has, by that time, become firmly established. The fact that planetary wave energy may propagate vertically through the stratosphere and mesosphere in the presence of strong westward winds during the summer and that the amplification of quasi 2-day modes is observed only during a comparati-

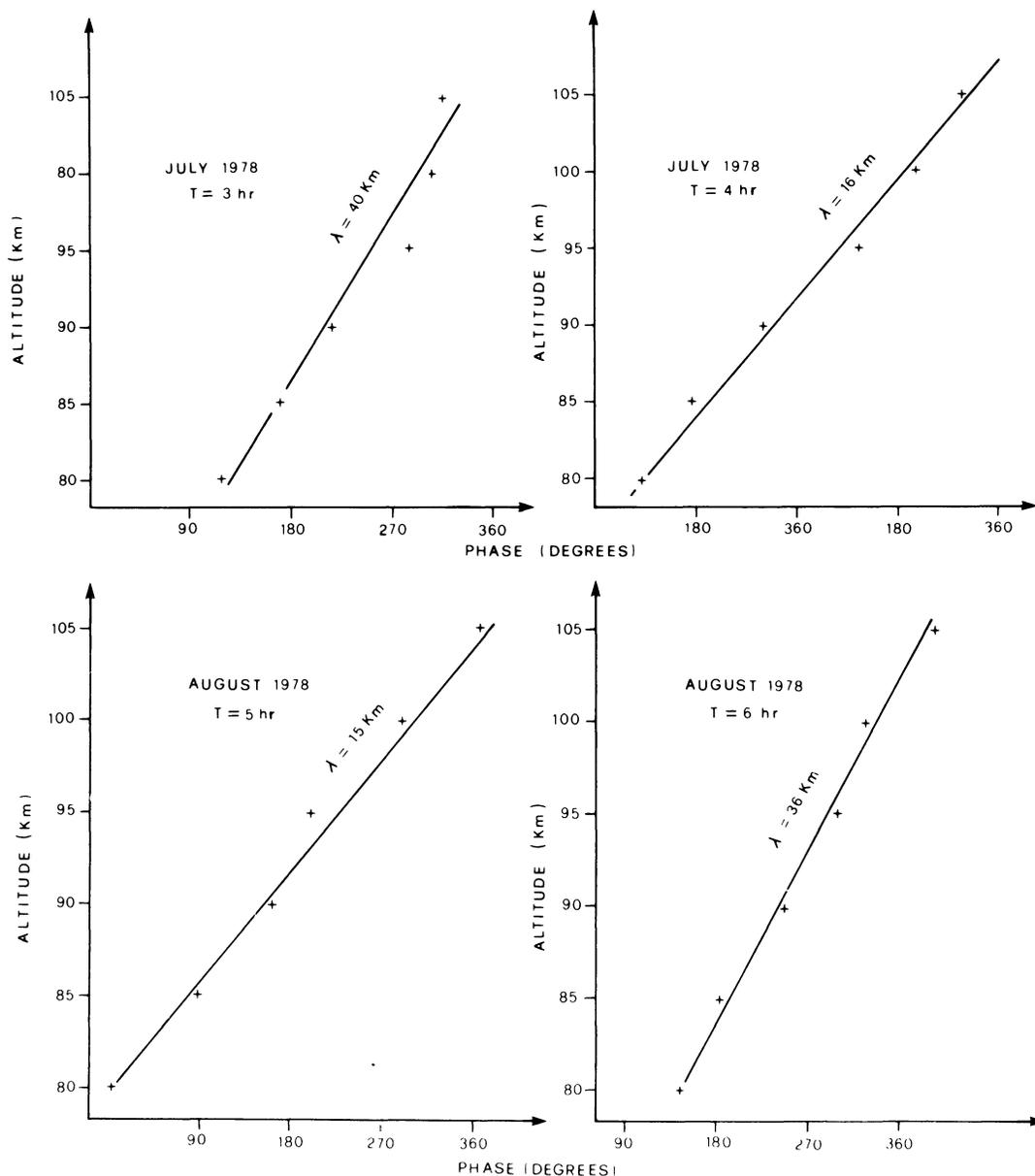


Fig. 6. Examples of phase variation versus height for some particular peaks of the small-scale wind amplitude spectra for July and August 1978

vely short period, suggests that we are dealing with atmospheric free oscillations (Muller et al. 1979). The theoretical problem of the vertical propagation of planetary waves was considered by Salby (1978) for the case of travelling waves. The author found that for free modes in a realistic atmosphere there are only two discrete vertical modes of propagation possible. The first mode is evanescent except for a thin region in the upper stratosphere. The lowest order wavenumber 3 mode of Salby's first type has a period of 2.08 days. Geisler and Dickinson (1976) showed that the response of the quoted mode is greatest at solstices and the summer mesosphere appears resonant to this free mode. The amplification of the quasi 2-day oscillations observed at Budrio about a month after the summer solstices of 1979 and 1980 confirms these theoretical predictions.

In addition to quasi 2-day oscillations, a 5-day zonal wave has been extracted in our November 1980 spectra (Fig. 1). The amplitudes of this oscillation at different heights are of the order of $3\text{--}4\text{ ms}^{-1}$. It has been suggested that this wave corresponds

to the greatest symmetric low-frequency external normal mode of the atmosphere (Geisler and Dickinson 1976). A significant peak of period near 5 days, observed at lower levels by Madden and Stokes (1975) was identified in both summer and winter in ionosonde measurements and in stratospheric temperatures by Fraser (1977). A similar disturbance has also been noticed in tropical pressure data (Burpee 1976) and in partial reflection observations (Manson et al. 1978).

As mentioned above, a wind fluctuation of period near 10 days, has been identified in the quoted spectra of Figure 1. A 10–15 days oscillation has been observed in E region ionosonde measurements and was found to be correlated with the wavenumber 1 component of stratospheric radiance during autumn and winter (Cavalieri 1976). Oscillations of comparable period were also noticed during autumn and winter in a 10-mb. analysis constructed from rocket and rawinsonde data (Finger et al. 1966) and in the coherence spectra between ionosonde and stratosphere measurements (Fraser and Thorpe 1976).

16 JULY - 2 AUGUST 1980

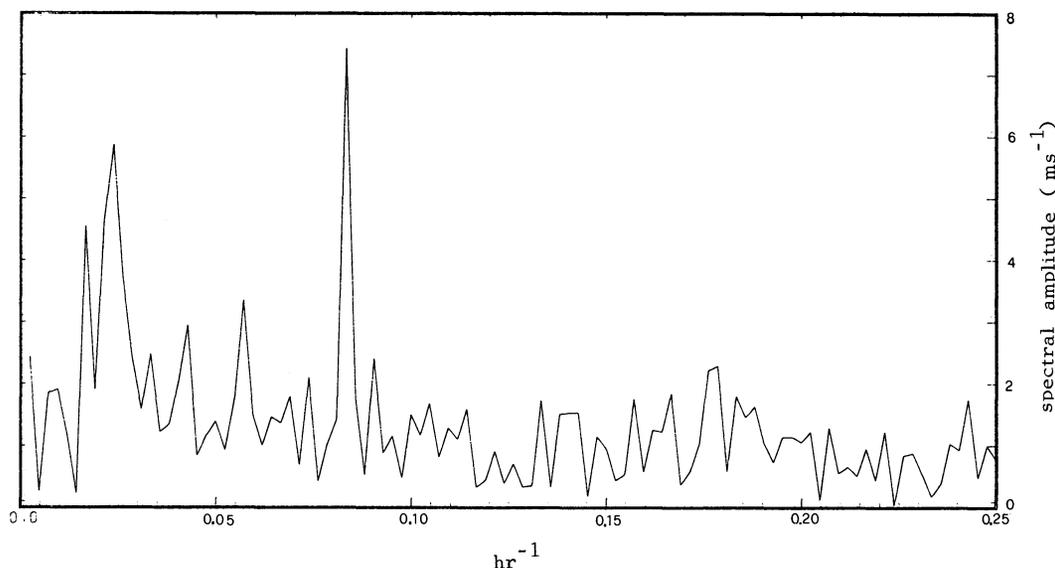


Fig. 7. Amplitude spectrum of the zonal wind at Budrio for the 16 July–2 August 1980 period for the 75–115 km height interval (average 95 km)

Conclusions

In this paper we have described the results of the November 1980 zonal wind oscillation observations carried out at Budrio during an integrated ground based, balloon and rocket campaign – the “Energy Budget Campaign 1980” – planned for the study of energy inputs and outputs in the upper atmosphere (60–180 km) during geomagnetic disturbances.

One of the main aims of this work has been to emphasize the variability of the winds in the upper mesosphere and lower thermosphere as observed in late autumn at Budrio. It has been shown that all the major wind components (prevailing winds, tides and gravity waves) exhibit amplitudes generally lower than those measured in November of previous years at Budrio. The daily variations in the amplitudes of the semidiurnal tide over time scales larger than 5–6 days suggest changes in the source whereas the daily drift in the phase of this tide could indicate that its vertical structure is consistent with short vertical wavelength propagating modes. The amplitudes of the diurnal and terdiurnal tides appear not relevant and the phases of these components exhibit non linear vertical variations. As for the summer 1978 results, the November 1980 wind observations indicate the possible importance of gravity waves in the energy budget of the quoted regions. Moreover, the wind time series examined here support the contention of earlier studies, that oscillations of periods near 2–3, 5–6 and 10–15 days occur frequently at lower ionospheric heights.

Acknowledgments. The work described above has been supported by the National Research Council of Italy. The authors are grateful to G. Bortolotti, C. Franceschi, S. Gottardi and G. Trivellone for their technical cooperation, mainly during the recording of the data discussed in this paper.

References

- Bernard, R.: Variability of the semidiurnal tide in the upper mesosphere, *J. Atmos. Terr. Phys.* **43**, 663–674, 1981
- Burpee, R.W.: Some features of global scale 4–5 day waves. *J. Atmos. Sci.* **33**, 2292–2299, 1976
- Cavaliere, D.: Travelling planetary-scale waves in the E-region. *J. Atmos. Terr. Phys.* **38**, 965–977, 1976
- Cevolani, G.: Evidence of planetary-scale waves in the lower thermospheric winds. *Il Nuovo Cimento C*, in press 1982
- Cevolani, G., Formiggini, C.: Evidence of internal gravity waves in the lower thermosphere. *Il Nuovo Cimento C* **4**, 1–25, 1981
- Cevolani, G., Hajduk, A.: Zonal winds at 95 km and radar meteor echoes. *Bull. Astron. Inst. Czechoslovakia* **30**, 297–307, 1979
- CIRA: Cospas International Reference Atmosphere, Akademie, Berlin, 1982
- Fellous, J.L., Bernard, R., Glass, M., Masseur, M., Spizzichino, A.: A study of the variations of atmospheric tides in the meteor zone. *J. Atmos. Terr. Phys.*, **37**, 1511–1524, 1975
- Finger, F.G., Woolf, H.M., Anderson, C.E.: Synoptic analysis of the 5,2 and 0.2 millibar surfaces for the IQSY period. *Monthly Weather Rev.* **94**, 651–661, 1966
- Fraser, G.J.: The 5-day wave and ionospheric absorption. *J. Atmos. Terr. Phys.* **39**, 121–124, 1977
- Fraser, G., Thorpe, M.: Experimental investigations of ionospheric/stratospheric coupling in southern midlatitudes. 1 – Spectra and cross spectra of stratospheric temperatures and the ionospheric f-min parameter. *J. Atmos. Terr. Phys.* **38**, 1003–1011, 1976
- Glass, M., Fellous, J.L.: The eight-hourly (ter-diurnal) component of atmospheric tides. *Space Research XV*, Akademie Berlin, 181–197, 1975
- Glass, M., Spizzichino, A.: Waves in the lower thermosphere: recent experimental investigations. *J. Atmos. Terr. Phys.* **36**, 1825–1839, 1974
- Geisler, J.E., Dickinson, R.E.: The five day wave on a sphere with realistic zonal winds. *J. Atmos. Sci.* **33**, 632–641, 1976
- Hajduk, A., Cevolani, G.: Simultaneous radar meteor observations at Ondrejov and Budrio. *Bull. Astron. Inst. Czechoslovakia* **32**, 304–310, 1981
- Kaiser, T.R., Kingsley, S.P., Muller, H.G., Rutter, P.: The semidiurnal tide in the lower thermosphere. XVIIth General Assembly of IUGG Canberra, December 1979
- Madden, R.A., Stokes, J.: Evidence of global-scale-5-day waves in a 73-year pressure record. *J. Atmos. Sci.* **32**, 831–836, 1975
- Manson, A.H., Gregory, J.B., Meek, C.E., Stephenson, D.G.: Winds and wave motions to 110 km at mid-latitudes. V. An analysis of data from September 1974–April 1975. *J. Atmos. Sci.* **35**, 592–599, 1978

- Muller, H.G., Kingsley, S.P., Rodgers, C.D., Prata, A.J.: Quasi 2-day oscillations in the wind at meteor altitudes and related temperature oscillations in the stratosphere. XVIIth General Assembly of IUGG Canberra, December 1979
- Muller, H.G., Nelson, L.: A travelling quasi 2-day wave in the meteor region. *J. Atmos. Terr. Phys.* **40**, 701–706, 1978
- Poulter, E.M.: Winter motions in the southern hemisphere meteor region. *J. Atmos. Terr. Phys.* **42**, 661–672, 1980
- Salby, M.L.: Ph. D. Thesis, Georgia Institute of Technology 1978
- Schaffner, M.: The circulating Page Loose System, a new solution for data processing. Radard Meteor Project Report n. 15, Harvard College Observatory and Smithsonian Astrophysical Observatory, pp. 129, 1966
- Spizzichino, A.: Etude des interactions entre les different composants du vent dans la haute atmosphere. *Ann. Geophys.* **26**, 9–34, 1970
- Stening, R.J., Meek, C.E., Manson, A.H., Stephenson, D.G.: Winds and wave motions to 110 km at midlatitudes. VI. Tidal, gravity and planetary waves, 1976. *J. Atmos. Sci.* **35**, 2194–2204, 1978
- Ulrich, T.J., Bishop, T.N.: Maximum entropy spectral analysis and autoregressive decomposition. *Rev. Geophys. Space Phys.* **13**, 183–200, 1975
- Verniani, F., Schaffner, M., Sinigaglia, G., Bortolotti, C., Dardi, A., Formiggini, C., Franceschi, C., Gottardi, S., Trivellone, G.: The Budrio station of the CNR meteor radar system for systematic measurements of upper atmospheric temperature and density. *Riv. Ital. Geofis.* **23**, 243–256, 1974
- Verniani, F., Cevolani, G., Grassi, G., Pedrini, F.: Systematic measurements of northern hemisphere winds in the meteor zone at 45°N. *J. Atmos. Terr. Phys.* **42**, 83–95, 1980

Received July 13, 1981; Revised version December 17, 1981

Accepted January 13, 1982

The Theory of Particle Acceleration in Astrophysical Objects Containing Shock Waves and Turbulent Plasma Motions*

A.M. Bykov and I.N. Toptyghin

Department of Theoretical Physics, M.I. Kalinin Polytechnical Institute, 195251, Leningrad USSR

Abstract. The theory of fast charged particle acceleration by an ensemble of large-scale plasma motions with randomly distributed shock waves (i.e. supersonic turbulence) is developed. We have considered the effect of suprathreshold particle acceleration in a compressible plasma with “frozen in” magnetic field and turbulent motions of different scales. The kinetic equation in the diffusion approach and diffusion coefficients in momentum space which describe the fast particle acceleration effect are obtained. There is a second order effect of Fermi acceleration in statistically uniform (non-expanding) systems with supersonic turbulence, although a single shock front leads to first order Fermi acceleration. This result is connected with the cancelation of the first order acceleration effect of shock fronts by particle deceleration in rarefaction waves between shocks.

The gain times for this acceleration mechanism are estimated. In some cases particle acceleration by supersonic turbulence is able to exceed the adiabatic deceleration losses. Some observational evidence and theoretical speculations relevant to the existence of supersonic turbulence in extragalactic radio sources, the interstellar medium, supernovae remnants and the solar wind are discussed briefly. We consider also some plausible applications of this acceleration mechanism.

Key words: Particle acceleration – Shock waves – Astrophysical supersonic turbulence

Introduction

There are several interesting works devoted to the problem of fast particle acceleration by astrophysical shock waves (Axford et al. 1977; Krymskiy 1977; Bell 1978; Blandford and Ostriker 1978). It has been shown that a strong plane shock wave propagating through turbulent plasma accelerates suprathermal particles and a power-law spectrum for the fast particles with a universal exponent is created for a wide range of initial parameters. The multiple interaction of fast particles with the shock front is provided by particle scattering on both turbulent plasma motions and “self-excited” magnetohydrodynamic waves. On the other hand there are many astrophysical objects where the existence of large scale

turbulent plasma motions with shock waves seems highly likely. As we shall show below, the existence of supersonic turbulence is plausible for the jet models of extended radio sources, shock waves, overlapping regions in interstellar medium, young and old supernova remnants and for stars with strong stellar winds. Our goal is now to study the acceleration of fast particles by supersonic turbulence with shock waves.

The Acceleration Mechanism

Let us consider fast charged particle interaction with an ensemble of randomly distributed shock waves and turbulent large-scale plasma motions with a “frozen-in” magnetic field. We use the following scale separation: large-scale plasma motions have scale $L \gg A$, where A is the fast particle scattering length. In our case scale L has an order of magnitude equal to the mean distance between shock fronts. We shall assume the presence of the regular magnetic field \mathbf{B}_0 and small-scale subsonic turbulence which provides fast particle scattering between shock fronts. The acceleration of fast particles by small scale subsonic turbulence has been studied in detail (see Tverskoy 1967; Toptyghin 1973), and we do not discuss it in this article. In this section we also suggest the following relations between the space scale sizes: $R_0 \gg L \gg A \gg R_L \gg \delta$ where R_0 is the system scale size, R_L are the gyro-radii of fast particles in the mean magnetic field \mathbf{B}_0 and δ is the characteristic width of the shock front. Under these assumptions the kinetics of fast particles can be described in terms of a distribution function in the diffusion approach $N(\mathbf{r}, p, t)$. The $N(\mathbf{r}, p, t)$ evolution between shock fronts is described by the transport equation in a strong magnetic field (see Bykov and Toptyghin, 1980)

$$\frac{\partial N}{\partial t} = \frac{\partial}{\partial r_\alpha} \chi_{\alpha\beta} \frac{\partial N}{\partial r_\beta} - U_\alpha \frac{\partial N}{\partial r_\alpha} + \frac{p}{3} \frac{\partial N}{\partial p} \frac{\partial U_\alpha}{\partial r_\alpha} \quad (1)$$

where $\chi_{\alpha\beta}$ is the space diffusion tensor, $\mathbf{U}(\mathbf{r}, t)$ is the large-scale plasma motion velocity field. It should be noted that, in the case of weak magnetic fields ($R_L \gg l_c$, where l_c is a small scale turbulence correlation length) the transport equation has same form (1) (Dolginov and Toptyghin 1966) but there are differences in expressions for fast particle currents. Since the transport equation (1) on a shock wave surface a priori is invalid, it is necessary to include the boundary conditions for the fast particle distribution functions into the expression. These boundary conditions for the case of the plane oblique shock wave have been obtained by means of direct micro-

* Paper presented at the Workshop on Acceleration of Particles by Shock Waves, 7–9 October 1980
Max-Planck-Institut für Aeronomie, Lindau, Federal Republic of Germany

scopic calculation by Vasiljev et al. (1978) and Bykov and Toptyghin (1980). This calculation takes into account the main processes on a shock wave surface: fast particle reflection and passage and the shock front's electric field (see for more detailed discussion the review by Toptyghin, 1980). The boundary condition calculation mentioned above is valid only under the assumption $\frac{U}{Y} \ll \alpha \ll 1$ where α is an angle between magnetic field and shock waves surface S and these conditions have the form:

$$N_1 = N_2 \quad (2)$$

$$\begin{aligned} & n_\alpha \left(\chi_{\alpha\beta}^{(2)} \frac{\partial N_2}{\partial r_\beta} - \chi_{\alpha\beta}^{(1)} \frac{\partial N_1}{\partial r_\beta} \right) \\ &= \frac{n_\alpha (U_\alpha^{(1)} - U_\alpha^{(2)})}{3} p \frac{\partial N}{\partial p} \end{aligned} \quad (3)$$

on S , where indices 1 and 2 are connected with upstream and downstream regions respectively, \mathbf{n} is the shock surface normal vector. It should be noted that the boundary conditions, Eqs. (2), (3) may be generalized for all α . In fact, these conditions have the same form for $\alpha = \frac{\pi}{2}$, in this case the magnetic field jump and electric field are absent. Thus for $\alpha = \frac{\pi}{2}$, $N_1 = N_2$ and condition (3) may be obtained from the fast particle current continuity condition. In the general case the boundary condition (3) may be obtained from the fast particle continuity law in the form of the transport equation (1), by integrating Eq. (1) over a small interval in the vicinity of the shock surface around a direction normal to it, under the assumption that Eq. 2 is valid. But the validity of (2) is proved for the limiting cases $\alpha \ll 1$ and $\alpha \rightarrow \frac{\pi}{2}$. Fast particle acceleration is described by the right-hand term of Eq. (3), the magnetic field geometry (angles $\alpha_{1,2}$) being included in the space diffusion coefficients $\chi^{(1,2)}$. We note also that in the case $\alpha = \frac{\pi}{2}$, the acceleration effect takes place due to a scattering center velocity jump on the shock surface (i.e. first order Fermi-type acceleration). In the case $\alpha \ll 1$ the shock front electric field acceleration (with account taken of multiple particle-front interaction) dominates. For arbitrary α both Fermi-type acceleration and fast particle acceleration by the shock electric field take place but their contribution is described by Eq. (3). In order to include the conditions (2) and (3) into the transport equation (1) we must consider the fast particle distribution function N as a continuous function on the shock surface, but the diffusion tensor $\chi_{\alpha\beta}$, velocity field $\mathbf{U}(\mathbf{r}, t)$ and fast particle diffusion current $\chi_{\alpha\beta} \frac{\partial N}{\partial r_\beta}$ as a discontinuous functions with finite jumps on the shock surfaces. (Hence their space argument derivatives are singular functions.) Now, we can consider fast particle acceleration by supersonic turbulence in terms of the transport equation. The distribution function N oscillations correlates with the random velocity field $\mathbf{U}(\mathbf{r}, t)$ and we must average N over the large scale motions. The method of averaging depends on the value of a dimensionless parameter $\beta = \frac{UL}{AY} = \frac{\tau_d}{\tau_c}$, where τ_d is the characteristic diffusion time $\tau_d = \frac{L^2}{AY}$, τ_c is the characteristic convective time scale $\tau_c = \frac{L}{U}$. If the fast particle scattering length Λ is high enough that $\beta \ll 1$, then we can use the quasilinear approach (Vedenov et al. 1962) for the averaging of the transport equation (1).

Let us separate out the slowly changing and the fast oscillating parts of the distribution function N :

$$N = F + \tilde{N}, \quad F = \langle N \rangle, \quad \langle \tilde{N} \rangle = 0, \quad \tilde{N} \ll F \quad (4)$$

where $\langle \rangle$ denotes averaging over the ensemble of the large scale turbulent motions. After substituting expression (4) for N into Eq. (1) and averaging it over the ensemble we obtain

$$\frac{\partial F}{\partial t} = \frac{\partial}{\partial r_\alpha} \langle \chi_{\alpha\beta} \rangle \frac{\partial F}{\partial r_\beta} - U_{0\alpha} \frac{\partial F}{\partial r_\alpha} + \frac{p}{3} \frac{\partial F}{\partial p} \frac{\partial U_{0\alpha}}{\partial r_\alpha} + \frac{1}{p^2} \frac{\partial}{\partial p} p^2 D \frac{\partial F}{\partial p} \quad (5)$$

where $\langle \chi_{\alpha\beta} \rangle$ is the average space diffusion tensor, \mathbf{U}_0 is the average velocity field and $\frac{\partial U_{0\alpha}}{\partial r_\alpha} = \left\langle \frac{\partial U_\alpha}{\partial r_\alpha} \right\rangle$ is the average divergence of the velocity field. It should be noted that in a statistically uniform nonexpanding system \mathbf{U}_0 and $\frac{\partial U_{0\alpha}}{\partial r_\alpha}$ should vanish. In other cases the corresponding terms in Eq. (5) describe the adiabatic deceleration (for expanding systems such as jets, supernova shells, stellar wind) or the first order particle acceleration (e.g. for accreting systems) and the convective transport. The effect of particle acceleration by supersonic turbulence is described by the fast particle momentum-space diffusion coefficient D :

$$D(p) = \frac{p^2}{9} \int \left\langle \frac{\partial U_\alpha}{\partial r_\alpha}(\mathbf{r}, t) \frac{\partial U_\beta}{\partial r_\beta}(\mathbf{r}', t') \right\rangle G(\mathbf{r}, \mathbf{r}', t, t') d\mathbf{r}' dt \quad (6)$$

Where $G(\mathbf{r}, \mathbf{r}', t, t')$ is the Green's function of the space diffusion equation.

For the isotropic turbulent velocity field we have the correlation function in the form of a Fourier transform:

$$\begin{aligned} & \langle U_\alpha(\mathbf{r}, t) U_\beta(\mathbf{r}', t') \rangle \\ &= \int \left[A(\kappa, \tau) \delta_{\alpha\beta} + B(\kappa, \tau) \frac{\kappa_\alpha \kappa_\beta}{\kappa^2} \right] e^{i\mathbf{k}\mathbf{x}} d\mathbf{k} \end{aligned} \quad (7)$$

where $\tau = t - t'$, $\mathbf{x} = \mathbf{r} - \mathbf{r}'$, $A(\kappa, \tau)$, $B(\kappa, \tau)$ are coefficients which depend on the spectrum of the turbulence.

We then obtain for the momentum-space diffusion coefficient:

$$\begin{aligned} & D \approx \frac{p^2}{9\chi} \int [A(\kappa, 0) + B(\kappa, 0)] d\mathbf{k} \\ & \text{or} \\ & D \approx \frac{U_{\text{eff}}^2 p^2}{9\chi} \end{aligned} \quad (8)$$

U_{eff}^2 has an order of magnitude equal to $\langle U^2 \rangle$ and $\chi \approx \frac{Y\Lambda}{3}$.

Now we consider the case when $\beta \gg 1$, which is realized for small enough fast particle scattering length Λ . Under this condition the convective term in Eq. (1) dominates and the diffusion term can be neglected. Thus Eq. (1) reduces to

$$\frac{\partial N}{\partial t} = -U_\alpha \frac{\partial N}{\partial r_\alpha} + \frac{p}{3} \frac{\partial N}{\partial p} \frac{\partial U_\alpha}{\partial r_\alpha}. \quad (9)$$

Equation (9) can be averaged directly by means of the method which is used in invariant perturbation theory in quantum electrodynamics (Abrikosov et al. 1962). We reduce the transport Eq. (9) to the integral equation:

$$\begin{aligned} & N(\mathbf{r}, p, t) = - \int_0^t d\tau \left[- \frac{\partial U_\alpha}{\partial r_\alpha}(\mathbf{r}, \tau) \frac{p}{3} \frac{\partial}{\partial p} \right. \\ & \left. + U_\alpha(\mathbf{r}, \tau) \frac{\partial}{\partial r_\alpha} \right] N(\mathbf{r}, p, \tau) + N(\mathbf{r}, p, 0). \end{aligned} \quad (10)$$

The iteration solution of (10) has the form:

$$N(\mathbf{r}, p, t) = \hat{T} \exp \left\{ - \int_0^t d\tau \left[- \frac{\partial U_\alpha}{\partial r_\alpha} \frac{p}{3} \frac{\partial}{\partial p} + U_\alpha \frac{\partial}{\partial r_\alpha} \right] \right\} N(\mathbf{r}, p, 0) \quad (11)$$

where \hat{T} is the ‘‘chronological ordering’’ operator. Since

$$\langle N(\mathbf{r}, p, 0) \rangle = N(\mathbf{r}, p, 0) = F(\mathbf{r}, p, 0)$$

we obtain

$$F(\mathbf{r}, p, t) = \left\langle \hat{T} \exp \left\{ - \int_0^t d\tau \left[- \frac{\partial U_\alpha}{\partial r_\alpha} \frac{p}{3} \frac{\partial}{\partial p} + U_\alpha \frac{\partial}{\partial r_\alpha} \right] \right\} \right\rangle F(\mathbf{r}, p, 0).$$

For the isotropic velocity field ($\langle U_\alpha(\mathbf{r}, \tau) \rangle = 0$) with Gauss distribution of amplitudes we can average the operator in expression (12) using the following relations:

$$\begin{aligned} & \left\langle \hat{T} \prod_{i=1}^n \left[- \frac{\partial U_\alpha}{\partial r_\alpha}(\mathbf{r}, \tau_i) \frac{p}{3} \frac{\partial}{\partial p} + U_\alpha(\mathbf{r}, \tau_i) \frac{\partial}{\partial r_\alpha} \right] \right\rangle \\ &= \sum_{\{p\}} \prod_{i,j=1}^n \left\langle \hat{T} \left[- \frac{\partial U_\alpha}{\partial r_\alpha}(\mathbf{r}, \tau_i) \frac{p}{3} \frac{\partial}{\partial p} + U_\alpha(\mathbf{r}, \tau_i) \frac{\partial}{\partial r_\alpha} \right] \right. \\ & \quad \left. \times \left[- \frac{\partial U_\beta}{\partial r_\beta}(\mathbf{r}, \tau_j) \frac{p}{3} \frac{\partial}{\partial p} + U_\beta(\mathbf{r}, \tau_j) \frac{\partial}{\partial r_\beta} \right] \right\rangle \end{aligned} \quad (13)$$

for $n=2\kappa$, and it drops to zero for $n=2\kappa+1$. $\sum_{\{p\}}$ is the sum over the manifold of permutations $\{p\}$ of the natural numbers $\{1, 2, \dots, n\}$

$$\left\langle U_\alpha(\mathbf{r}, \tau_i) \frac{\partial U_\alpha}{\partial r_\alpha}(\mathbf{r}, \tau_j) \right\rangle = \left\langle U_\alpha(\mathbf{r}, \tau_i) \frac{\partial U_\beta}{\partial r_\beta}(\mathbf{r}, \tau_j) \right\rangle = 0.$$

Since we have no vectors in our problem after averaging (\mathbf{B}_0 is the pseudovector).

$$\langle U_\alpha(\mathbf{r}, \tau_i) U_\beta(\mathbf{r}, \tau_j) \rangle = \phi_{\alpha\beta}(\tau_i - \tau_j) \quad (14)$$

$$\begin{aligned} & \left\langle \frac{\partial U_\alpha}{\partial r_\alpha}(\mathbf{r}, \tau_i) \frac{\partial U_\beta}{\partial r_\beta}(\mathbf{r}, \tau_j) \right\rangle \\ &= - \left\langle U_\alpha(\mathbf{r}, \tau_i) \frac{\partial^2 U_\beta}{\partial r_\alpha \partial r_\beta}(\mathbf{r}, \tau_j) \right\rangle = \Psi(\tau_i - \tau_j). \end{aligned} \quad (15)$$

Carrying out the calculations and taking into account (13), (14), (15) we obtain:

$$\begin{aligned} F(\mathbf{r}, p, t) &= \exp \left\{ \frac{1}{2} \int_0^t d\tau_i \int_0^t d\tau_j \left[\phi_{\alpha\beta}(\tau_i - \tau_j) \frac{\partial^2}{\partial r_\alpha \partial r_\beta} + \frac{1}{9} \frac{\Psi}{p^2} \frac{\partial}{\partial p} p^4 \frac{\partial}{\partial p} \right] \right\} F(\mathbf{r}, p, 0) \\ &\text{or in standard form:} \end{aligned} \quad (16)$$

$$\frac{\partial F}{\partial t} = \frac{\partial}{\partial r_\alpha} \chi_{\alpha\beta} \frac{\partial}{\partial r_\beta} F + \frac{1}{p^2} \frac{\partial}{\partial p} D p^2 \frac{\partial F}{\partial p} \quad (17)$$

where

$$\chi_{\alpha\beta} = \int_0^t \langle U_\alpha(\mathbf{r}, t) U_\beta(\mathbf{r}, t - \tau) \rangle d\tau \quad (18)$$

$$D(p) = \frac{p^2}{q} \int_0^t \left\langle \frac{\partial U_\alpha}{\partial r_\alpha}(\mathbf{r}, t) \frac{\partial U_\beta}{\partial r_\beta}(\mathbf{r}, t - \tau) \right\rangle d\tau. \quad (19)$$

The estimates of (18) and (19) under general assumptions about supersonic turbulence statistical properties give us

$$\chi \approx \langle U^2 \rangle \tau_{cor}, \quad D \approx p^2 \frac{\langle U^2 \rangle}{9L^2} \tau_{cor} \quad (20)$$

where $\tau_{cor} \approx \frac{L}{U}$ is the correlation time for large scale motions.

Finally we have $\chi \approx UL$, $D \approx \frac{U}{9L} p^2$

$$D(p) = \frac{\langle U^2 \rangle p^2}{9\chi}. \quad (21)$$

We can now summarize the results for particle acceleration by large-scale supersonic turbulence in systems with strong scattering of fast particles by small scale turbulence ($\Lambda \ll L$).

1) We have the second order Fermi acceleration effect with fast particle gain time $\tau_a \approx \frac{p^2}{D} = \frac{9\chi}{\langle U^2 \rangle}$, see Eqs. (8) and (21).

2) The acceleration effect is connected with compression of the ambient medium. If $\frac{\partial U_\alpha}{\partial r_\alpha} = 0$, it drops to zero.

3) The momentum space diffusion coefficient $D(p) = \frac{\langle U^2 \rangle p^2}{9\chi}$ is a continuous function of the dimensionless parameter $\beta = \frac{UL}{Y\Lambda}$.

4) This acceleration effect is able to exceed the adiabatic deceleration losses of fast particles for an expanding system since e.g. for $\beta \gg 1$

$$\tau_a^{-1} \approx \frac{U}{L}, \quad \tau_{ad}^{-1} \approx \frac{U}{R_0}$$

and $\tau_a^{-1} > \tau_{ad}^{-1}$ because $L < R_0$.

Next we consider fast particle acceleration by supersonic turbulence with weak scattering of fast particles by small-scale subsonic turbulence between shocks i.e. the case $\Lambda \gg L$. Gurevich and Rumyantsev (1973) were the first to consider this model. The isotropic distribution of fast particles takes place only on a scale $\gg L$; it is produced by means of interaction of fast particles with randomly distributed weak shocks ($\Delta B \ll B$). In accordance with their reasoning we can describe fast particles by an isotropic distribution function $F(\mathbf{r}, p, t)$ over the system scale R_0 . The $F(\mathbf{r}, p, t)$ evolution is described by the transport Eq. (5) We now estimate the space and momentum-space diffusion coefficients for the particle acceleration by statistically uniform (nonexpanding) systems with weak shocks and large scale motions under conditions $L \ll \Lambda \ll R_0$. We define the scattering time as $\tau_s = \frac{\Lambda}{Y}$; it is clear

that τ_s^{-1} is proportional to $\langle \mathbf{F}_m^2 \rangle$ (where $\mathbf{F}_m = \frac{l}{c} \mathbf{V} \times \Delta \mathbf{B}$ is the magnetic force) since $\langle \mathbf{F}_m \rangle = 0$. On the other hand particle acceleration takes place due to the action of the stochastic electric field with electric force

$$\mathbf{F}_e = -\frac{e}{c} \mathbf{U} \times (\mathbf{B}_0 + \Delta \mathbf{B})$$

and thus

$$\tau_a^{-1} \approx \tau_s^{-1} \frac{\langle \mathbf{F}_e^2 \rangle}{\langle \mathbf{F}_m^2 \rangle} = \frac{Y}{\Lambda} \frac{\langle U^2 \rangle}{Y^2} \frac{B_0^2}{\langle \Delta B^2 \rangle}. \quad (22)$$

Let us estimate the fast particle scattering length Λ in this case. The fast particle pitch angle θ change on the single weak shock front is $\overline{\Delta\theta} \simeq \frac{\Delta B}{B_0}$.

Since the sign of $\Delta\theta$ is stochastic, the mean change $\langle \Delta\theta \rangle$ will be equal to unity only after $N \simeq \frac{1}{\langle \Delta\theta^2 \rangle}$ interactions of the fast particle with weak shocks. If the mean distance between random distributed shock waves is L then we obtain $\Lambda \simeq NL \simeq \frac{B_0^2}{\langle \Delta B^2 \rangle} L$.

Thus in accordance with (22) we can write

$$\tau_a^{-1} \simeq \frac{\langle U^2 \rangle}{YL} = \frac{U_{sh}^2}{Y\Lambda} \quad (23)$$

where $U_{sh}^2 = \langle U^2 \rangle \frac{B_0^2}{\langle \Delta B^2 \rangle}$ is the mean square weak shock velocity. Finally we have

$$\chi \simeq YL \frac{B^2}{\langle \Delta B^2 \rangle} \quad \text{and} \quad D(p) \simeq \frac{U^2 p^2}{\chi} \quad (24)$$

i.e. the second order Fermi acceleration effect. Particles acceleration by an ensemble of weak shock fronts under the condition $\Lambda \gg L$ has been considered by Gurevich and Rumyantsev (1973, 1978). They have obtained the following estimate for the fast particle gain time

$$\tau_a^{-1} = \frac{U_{sh}}{4L} \frac{\Delta B}{B_0} \quad (25)$$

This estimate (25) differs from our estimate (23) since they have obtained the first order Fermi acceleration effect. We believe this difference is due to the incorrect allowance for rarefaction waves between shock fronts made by Gurevich and Rumyantsev. It should be noted that in all cases which we describe above the second order Fermi acceleration takes place for the fast particle acceleration by supersonic turbulence in statistically uniform, nonexpanding systems. This fact is connected with complete cancellation of the first order particle acceleration effect (which takes place on a single shock front in a turbulent medium) by adiabatic losses in rarefaction waves between shocks. The presence of such rarefaction waves is necessary for nonexpanding (or nonconverging) systems.

Discussion and Some Astrophysical Applications

In this paragraph we briefly discuss some problems which are connected with supersonic turbulence development and fast particle acceleration in astrophysical objects (see also Bykov and Toptyghin 1981).

The Extragalactic Radio Sources

There is an attractive theory for the explanation of extended radiogalaxy structure – the beam model (Blandford and Rees 1978). It was proposed that collimated supersonic flow from the active nuclei of galaxies provides, by means of interaction with large plasma clouds, the observed structure of extended radio sources. This model is a plausible explanation of the radio and optical structure of the jet in M 87 (see the optical Knots model by Blandford and Königl 1979) and the structure of the galactic stellar object with peculiar line emission – SS 433 (Begelman et al. 1980). In such systems the develop-

ment of supersonic turbulence takes place on the interaction surfaces of supersonic jets and plasma clouds (e.g. supernova shells). The presence of jet inhomogeneities (sound and entropy waves, etc.) leads naturally to a rising of shock waves by means of breaking (Rees 1978) and interaction of these waves with the cloud's bow shock continuously generates the supersonic turbulence within the cloud's surface layer. The small-scale subsonic turbulence would be generated by means of interaction of linear magnetohydrodynamic waves with the bow shock. Thus such systems may effectively re-accelerate the fast particles associated with a strong bow shock.

Cosmic Ray Acceleration in the Interstellar Medium

One of the main problems in the theory of the origin of cosmic rays is whether most cosmic rays are accelerated in localized sources or in the interstellar medium (ISM). It seems highly likely that a fast particle acceleration mechanism by large-scale plasma motions and shock waves plays an important role for both classes of models since most of the available energy storage is in large-scale plasma motions. A detailed model of cosmic rays re-acceleration by single shock waves in the extended envelopes of supernovae remnants has been considered recently by Blandford and Ostriker (1980). Models of cosmic ray acceleration in OB stars shells and supernovae remnants in dense clouds have been discussed by Völk (1981).

We now consider cosmic ray acceleration by ISM turbulence. There is much observational evidence for the existence of large-scale turbulence in the ISM. A study of interstellar motions by Larson (1981) showed that the velocity dispersion σ in molecular clouds follows an approximate power-law dependence $\sigma = 1.1(L)^{0.38}$ km/s where L is the scale size of the observed region measured in pc. Such dependence holds for $0.1 \lesssim L \lesssim 100$ pc. It is suggested that the observed motions are all part of a common hierarchy of interstellar turbulence. Large-scale fluctuations ($L \sim 100$ pc) of the interstellar magnetic field were reviewed by Heiles (1976). It seems highly likely that large-scale ISM turbulence sources are shearing flows from differential rotation of galactic gas with spiral shock waves and nonstationary phenomena such as supernovae explosions etc. The essential point is that the ISM gas energy budget is regulated by supernova shock-wave propagation through the inhomogeneous ISM. Supernovae explosions produce an interconnecting “tunnel” system in the ISM (see review by McCray and Snow 1979). Large-scale turbulence with randomly distributed shocks in the ISM tunnels, probably with coronal gas, is developed due to strong shock wave interactions with ISM inhomogeneities. In fact the shock wave interaction with randomly distributed clouds leads to secondary shock production. Small-amplitude hydromagnetic waves (e.g. generated by cosmic ray streaming instability) are amplified by strong shock waves (McKenzie and Westphal 1970). Hydromagnetic waves with wavelengths $\lambda < 10^{17}$ cm are collisionless in the ISM coronal gas phase and thus their damping is anisotropic. However, energy transfer between the different scales due to nonlinear decay is faster than collisionless wave damping for scales

$$l \gtrsim r_i \left(\frac{L}{r_i} \right)^{1/4} \left(\frac{Y_i^2}{E} \right)^{3/8}$$

where r_i and Y_i are the thermal ion gyroradius and velocity respectively, E is energy per unit mass for the main scale L of turbulence (Foote and Kulsrud 1979). Thus nonlinear decay

of the regular large scale motions leads to turbulence formation in the tunnels with coronal gas. The inertial range of such turbulence is $10^{12} \text{ cm} \lesssim l \lesssim 3 \cdot 10^{20} \text{ cm}$, velocity for the main scale of this turbulence $U(L) \sim 100 \text{ km/s}$. The speed of sound for hot coronal gas with $(n, T) \sim (3 \cdot 10^{-3} \text{ cm}^{-3}, 10^6 \text{ K})$ is $C_s \sim 100 \text{ km/s}$.

We now consider cosmic ray acceleration by large-scale turbulence in the large coherent structures of ISM-tunnels. The main features considered in the model are enumerated below:

- 1) Cosmic ray acceleration for both primary and secondary nuclei takes place in the ISM tunnels.
- 2) Primary cosmic ray nuclei are injected with energy $\varepsilon_{inj} \lesssim 10^9 \text{ eV}$ by shock waves directly from a thermal pool. Interstellar grains should be destroyed by shock waves or a betatron mechanism since heavy elements are depleted in the ISM (Eichler, 1980).
- 3) Secondary nuclei are produced by means of interaction of primary nuclei in the ISM molecular clouds. They are then injected into the acceleration regime within the tunnels.

Cosmic ray acceleration in the tunnel is described by Eq. (5), but we now consider steady state, statistically uniform turbulence with $\mathbf{U}_0 = 0$ and $\text{div } \mathbf{U}_0 = 0$. We include a source term $S(p)$ into the transport equation:

$$\frac{\partial}{\partial r_\alpha} \langle \chi_{\alpha\beta} \rangle \frac{\partial F}{\partial r_\beta} + \frac{1}{p^2} \frac{\partial}{\partial p} p^2 D \frac{\partial F}{\partial p} = S(p). \quad (26)$$

In accordance with the consideration, detailed above fast particle diffusion and acceleration are determined by the dimensionless parameter $\beta = \frac{UL}{YA}$. For relativistic particles there is an energy dependence of scattering length $A(\varepsilon) \sim \varepsilon^{2-\nu}$ where ν is the small-scale turbulence spectral index (Toptyghin 1973). Unfortunately the spectral index ν for ISM turbulence is unknown. But if the same value holds for hot ISM gas that is appropriate for molecular clouds i.e. $\nu = 1.76$ then $A(\varepsilon) \sim \varepsilon^{0.24}$.

We can estimate $\frac{L}{A(\varepsilon)}$ from the observed anisotropy of cosmic rays since for tunnels $L \lesssim R_0$ (here R_0 is a minimum scale size of a tunnel which has an elongated form). Thus we find $\frac{L}{A(\varepsilon)} > 10^4$ for $\varepsilon \approx 10^9 \text{ eV}$ and $\frac{L}{A(\varepsilon)} \approx 10^3$ for $\varepsilon \lesssim 10^{16} \text{ eV}$. Therefore we have $\beta \gtrsim 1$ for $\varepsilon < 10^{16} \text{ eV}$ and the turbulent diffusion regime is realized with large-scale diffusion coefficient $\chi \approx UL$ and $D(p) \approx \frac{p^2 U}{9L}$ within the tunnel's inner region. For primary nuclei we can introduce a source term of the form $S(p) = S_0(p) \theta \left(1 - \frac{p}{p_{inj}}\right)$ where $S_0(p)$ is an arbitrary function of momentum p which is determined by the concrete form of the injection mechanism, $p_{inj} = \frac{\varepsilon_{inj}}{c}$ with $\varepsilon_{inj} \lesssim 10^9 \text{ eV}$.

Then accelerated primary nuclei spectrum can be obtained from Eq. (26):

$$p^2 F_1(p) \sim S_0(p_{inj}) p^{-\gamma_1} \\ \gamma_1 = -\frac{1}{2} + \sqrt{\frac{9}{4} + 9 \left(\frac{L}{R_0}\right)^2}. \quad (27)$$

This power-law spectrum of primary nuclei extends over an energy range of at least $10^9 \text{ eV} < \varepsilon < 10^{16} \text{ eV}$ with spectral index $\gamma_1 = 2.85$ for $L = R_0$. For $\varepsilon \gtrsim 10^{16} \text{ eV}$, $\beta < 1$ and we have a particle

acceleration - diffusion regime with $\chi \approx \frac{YA(p)}{3}$ and $D(p) = \frac{U^2}{9\chi(p)}$ (see Eq. (8)). The declination of the cosmic ray spectrum for $\varepsilon > 10^{16} \text{ eV}$ would be determined by the small-scale turbulence spectral index ν since $A(p) \sim p^{2-\nu}$. Let us now consider the spectrum of secondary nuclei in this model. The matter traversed by primary nuclei during the acceleration time τ_a is determined by $\bar{p} c \tau_a \approx 3 \cdot 10^{-2} \text{ g/cm}^2$.

This value is too small for secondary nuclei generation in the tenuous tunnels gas. Therefore we assume that secondary nuclei production takes place in dense phases of the ISM e.g. in molecular clouds with $(n, T) \sim (10^3 \text{ cm}^{-3}, 10 \text{ K})$.

A key feature of this assumption is that a large fraction of ISM mass is in the molecular clouds and it is important for our model that active star formation probably takes place at the outer edges of molecular clouds. In order to determine the source term of transport Eq. (26) for the secondary nuclei we have to find the primary nuclei spectrum in the molecular cloud, since the secondary nuclei production cross-section is nearly energy independent above $\varepsilon > 10^9 \text{ eV}$. The primary nuclei spectrum within the scatter free molecular cloud may be found by means of an extension of the spectrum (27) with account taken of the boundary layer between the tunnel and the molecular cloud, it's likely scale size $\sim 10 \text{ pc}$. Fast particle leakage from the tunnel is determined by an ordinary diffusion contrary to turbulent diffusion within the inner tunnel region. Since the fast particle scattering length Λ is energy dependent, we have a declination of the primary nuclei spectrum in the molecular cloud and therefore the secondary nuclei source term has the form

$$S(p) = S_2 p^{-(\gamma_2 + 2)} \quad (28)$$

where

$$\gamma_2 = \gamma_1 + \Delta.$$

It should be noted that, in contrast to primary nuclei spectrum formation, the secondary nuclei source term determines the spectral index of accelerated secondary nuclei over the whole range of observed energies of the secondary cosmic rays. Therefore, for accelerated secondary nuclei we have:

$$p^2 F_2(p) \sim S_2 \cdot p^{-\gamma_2}. \quad (29)$$

Thus the observed primary to secondary energy dependence ratio can be explained in our model if

$$\Delta \approx 2(2 - \nu) \approx 0.5 \quad \text{for } 10^9 < \varepsilon < 10^{11} \text{ eV}.$$

Finally it should be noted that further development of our understanding of ISM turbulence structure and investigations of γ ray production in molecular clouds would make possible a more quantitative development of the model outlined here. It is also important to consider the injection problem in more detail.

Acknowledgement. We are indebted to Dr. V.S. Ptuskin for discussion on some problems considered in the last section of our paper.

References

- Abrikosov, A.A., Gorkov, L.P., Dzyaloshinskiy, I.E.: Quantum fields theory methods in statistical physics. Moskov: Nauka 1962
 Axford, W.I., Leer, E., Scadron, G.: Acceleration of cosmic rays at shock fronts. In: Proc. 15-th ICCR, vol. 2, p. 273, Bulgarian Academy of Sciences 1977

- Begelman, M.C., Sarasin, C.L., Hatchett, S.P., McKee, C.F., Arons, J.: Beam models for SS433. *Astrophys. J.* **238**, 722–730, 1980
- Bell, A.R.: The acceleration of cosmic rays in shock fronts. *Monthly Notices R. Astron. Soc.* **182**, 147–156, 1978
- Blandford, R.D., Königl, A.: A model for the Knots in the M87 Jet. *Astrophys. Lett.* **20**, 15–21, 1979
- Blandford, R.D., Ostriker, J.P.: Particle acceleration by astrophysical shocks. *Astrophys. J.* **221**, L29–L32, 1978
- Blandford, R.D., Ostriker, J.P.: Supernova shock acceleration of cosmic rays in the Galaxy. *Astrophys. J.* **237**, 793–808, 1980
- Blandford, R.D., Rees, M.J.: Extended and compact extragalactic radio sources: Interpretation and theory. *Physica Scripta* **17**, 265–279, 1978
- Bykov, A.M., Toptyghin, I.N.: Scattering and transport of fast particles in strong magnetic field. *Geomagn. Aeron.* **20**, 177–182, 1980
- Bykov, A.M., Toptyghin, I.N.: Cosmic ray acceleration in the sources by supersonic turbulence *Izvestia Akad. Nauk SSSR, ser. fiz.* **45**, 474–485, 1981
- Dolginov, A.Z., Toptyghin, I.N.: Multiple scattering of particles in a magnetic field with random inhomogeneities. *Soviet J. Exp. Theor. Phys.* **51**, 1771–1780, 1966
- Eichler, D.: Basic inconsistencies in models of interstellar cosmic-ray acceleration. *Astrophys. J.* **237**, 809–813, 1980
- Foote, E.A., Kulsrud, R.M.: Hydromagnetic waves in high β plasmas. *Astrophys. J.* **233**, 302–316, 1979
- Gurevich, L.E., Romyantsev, A.A.: The cosmic rays of superhigh energies. *Astrophys. Space Sci.* **22**, 79–85, 1973
- Gurevich, L.E., Romyantsev, A.A.: Kinetics of relativistic particles generation in supernovae bursts. *Pisma v Astron. J.* **4**, 505–508, 1978
- Heiles, C.: The interstellar magnetic field. *Annual Rev. Astron. Astrophys.* **14**, 1–22, 1976
- Krymskiy, G.F.: The regular mechanism of charged particles acceleration by shock front. *Doklady Akad. Nauk SSSR* **234**, 1306–1309, 1977
- Larson, R.B.: Turbulence and star formation in molecular clouds. *Monthly Notices R. Astron. Soc.* **194**, 809–826, 1981
- McCray, R., Snow, T.P.: The violent interstellar medium. *Annual Rev. Astron. Astrophys.* **17**, 213–240, 1979
- McKenzie, J.F., Westphal, K.O.: Interaction of hydromagnetic waves with hydromagnetic shocks. *Phys. Fluids* **13**, 630–640, 1970
- Rees, M.J.: The M87 jet: internal shocks in a plasma beam? *Monthly Notices R. Astron. Soc.* **184**, 61–65, 1978
- Toptyghin, I.N.: Fast particles interaction with magnetohydrodynamics turbulence. *Astrophys. Space Sci.* **20**, 329–350, 1973
- Toptyghin, I.N.: Acceleration of particles by shocks in a cosmic plasma. *Space Sci. Rev.* **26**, 157–213, 1980
- Tverskoy, B.A.: On the theory of turbulent acceleration of charged particles. *Soviet J. Exp. Theor. Phys.* **53**, 1417–1426, 1967
- Vasiljev, V.N., Toptyghin, I.N., Chirkov, A.G.: Fast particles interaction with shock front in the turbulent medium. *Geomagn. Aeron.* **18**, 415–420, 1978
- Vedenov, A.A., Velikhov, E.P., Sagdeev, R.Z.: Quasilinear theory of plasma oscillations, *Nucl. Fusion Suppl.* **2**, 465–485, 1962
- Völk, H.J.: Shock acceleration in lossy media and diffuse gamma-ray sources. *Max-Planck-institute für Kernphysik, Heidelberg, Preprint* 1981

Received January 22, 1981, Revised version November 9, 1981 and January 25, 1982

Accepted January 28, 1982