PLASMA AND THE UNIVERSE

Dedicated to Professor Hannes Alfvén on the Occasion of His 80th Birthday, 30 May 1988

Guest Editors

CARL-GUNNE FÄLTHAMMAR
Royal Institute of Technology, Stockholm, Sweden

GUSTAF ARRHENIUS
Scripps Institution of Oceanography, University of California, La Jola, California, U.S.A.

BIBHAS R. DE
Chevron Oil Field Research Company, La Habra, California, U.S.A.

NICOLAI HERLOFSON
Royal Institute of Technology, Stockholm, Sweden

D. ASOKA MENDIS,
University of California at San Diego, La Jolla, California, U.S.A.

and

ZDENĚK KOPAL
Department of Astronomy, The University, Manchester, U.K.

Reprinted from Astrophysics and Space Science
Vol. 144, Nos. 1/2 (1988)

Kluwer Academic Publishers
Dordrecht / Boston / Lancaster

Library of Congress Cataloging in Publication Data

CIP

ISBN-13: 978-94-010-7858-0 e-ISBN-13: 978-94-009-3021-6
DOI: 10.1007/978-94-009-3021-6
CIP-data will appear on separate card.

Published by Kluwer Academic Publishers,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Kluwer Academic Publishers incorporates
the publishing programmes of
D. Reidel, Martinus Nijhoff, Dr W. Junk and MTP Press.

Sold and distributed in the U.S.A. and Canada
by Kluwer Academic Publishers,
101 Philip Drive, Norwell, MA 02061, U.S.A.

In all other countries, sold and distributed
by Kluwer Academic Publishers Group,
P.O. Box 322, 3300 AH Dordrecht, The Netherlands.

ASTROPHYSICS AND SPACE SCIENCE / *Vol. 144 Nos. 1–2 May 1988*

SPECIAL ISSUE

PLASMA AND THE UNIVERSE

Dedicated to Professor Hannes Alfvén on the Occasion of His 80th Birthday, 30 May 1988

Guest Editors

CARL-GUNNE FÄLTHAMMAR, GUSTAF ARRHENIUS, BIBHAS R. DE, NICOLAI HERLOFSON, D. ASOKA MENDIS, and ZDENĚK KOPAL

CARL-GUNNE FÄLTHAMMAR, GUSTAF ARRHENIUS, NICOLAI, BIBHAS R. DE,- HERLOFSON, D. ASOKA MENDIS, and ZDENĚK KOPAL / Special Issue Dedicated to Professor Hannes Alfvén on the Occasion of His 80th Birthday, 30 May 1988 1

LENNART LINDBERG / Observations of Propagating Double Layers in a High Current Discharge 3

N. BRENNING and I. AXNÄS / Critical Ionization Velocity Interaction: Some Unsolved Problems 15

B. LEHNERT / Velocity Limitation of a Neutral Dust Cloud Colliding with a Magnetized Plasma 31

MICHAEL A. RAADU and J. JUUL RASMUSSEN / Dynamical Aspects of Electrostatic Double Layers 43

PER CARLQVIST / Cosmic Electric Currents and the Generalized Bennett Relation 73

WALTER J. HEIKKILA / Elementary Ideas Behind Plasma Physics 85

BIBHAS R. DE / The Dielectric Analogue of Magnetohydrodynamics 99

CARL-GUNNE FÄLTHAMMAR / Magnetosphere-Ionosphere Interactions – Near-Earth Manifestations of the Plasma Universe 105

L. P. BLOCK / Acceleration of Auroral Particles by Magnetic-Field Aligned Electric Fields 135

BENGT HULTQVIST and RICKARD LUNDIN / Parallel Electric Fields Accelerating Ions and Electrons in the Same Direction 149

T. A. POTEMRA / Birkeland Currents in the Earth's Magnetosphere 155

P. E. SANDHOLT and E. EGELAND / Auroral and Magnetic Variations in the Polar Cusp and Cleft – Signatures of Magnetopause Boundary-Layer Dynamics 171

C. E. McILWAIN / Plasma Acceleration, Injection, and Loss: Observational Aspects 201

M. HORANYI, H. L. F. HOUPIS, and D. A. MENDIS / Charged Dust in the Earth's Magnetosphere 215

SHELLEY H. ROGERS and ELDEN C. WHIPPLE / Generalized Adiabatic Theory Applied to the Magnetotail Current Sheet 231

R.-L. XU / Particle Dynamics and Currents in the Magnetotail 257

L. J. LANZEROTTI and C. G. MACLENNAN / Hydromagnetic Waves Associated with Possible Flux Transfer Events 279

J. N. TOKIS / Free Convection and Mass Transfer Effects on the Magnetohydrodynamic Flows Near a Moving Plate in a Rotating Medium 291

S.-I. AKASOFU / An Electric-Current Description of Solar Flares 303

CORNELIS DE JAGER / Solar Flares through Electric Current 311

J. O. STENFLO / Global Wave Patterns in the Sun's Magnetic Field 321

D. LAL / Temporal Variations of Low-Energy Cosmic-Ray Protons on Decade to and Million Year Time-Scales: Implications to Their Origin 337

W.-H. IP / Electrodynamics of the EUV/X-ray Bright Point and Filamentary Flux Loop Complexes 347

Y. ÖHMAN / Peculiar Solar Flares and Plages of Possible Interest for the Study of Electric Quadrupole Lines of Na I and Mg II 353

GUSTAF ARRHENIUS and SUSANNE ARRHENIUS / Band Structure of the Solar System: An Objective Test of the Grouping of Planets and Satellites 357

MICHEL AZAR and WILLIAM B. THOMPSON / The Cosmogonic Shadow Effect 373

MURRAY DRYER, ZDENKA KOPAL SMITH, and SHI TSAN WU / The Role of Magnetohydrodynamics in Heliospheric Space Plasma Physics Research 407

A. A. GALEEV and R. Z. SAGDEEV / Alfvén Waves in a Space Plasma and Its Role in the Solar Wind Interaction with Comets 427

K. SZEGÖ / Modeling a Cometary Nucleus 439

ANTHONY L. PERATT and A. J. DESSLER / Filamentation of Volcanic Plumes on the Jovian Satellite Io 451

S. M. KRIMIGIS and D. VENKATESAN / *In Situ* Acceleration and Gradients of Charged Particles in the Outer Solar System Observed by the Voyager Spacecraft 463

E. MÖBIUS, B. KLECKER, D. HOVESTADT, and M. SCHOLER / Interaction of Interstellar Pick-Up Ions with the Solar Wind 487

K. MARTI and H. E. SUESS / The Even-Odd Systematics in *R*-Process Nuclide Abundances 507

IMRE RUFF, GEORGE MARX, and DAVID S. DEARBORN / The Solubility Problem of Heavy Elements in Internal Stellar Plasma Revisited 519

K. PAPADOPOULOS / Electron Heating in Superhigh Mach Number Shocks 535

K. D. COLE / 'Dead' Pulsars: Cosmic-Ray Sources 549

ZDENĚK KOPAL / The Effects of Mutual Irradiation on the Observed Radial Velocity of the Components of Close Binary Systems 557

MICHEL AZAR and W. B. THOMPSON / Magnetic Confinement of Cosmic Clouds 587

T. W. HARTQUIST and J. E. DYSON / Mass Pick-Up in Astronomical Flows 615

SATIO HAYAKAWA / Infrared Emission from Interstellar Plasma 629

V. A. AMBARTSUMIAN / On the Fluctuations of the Surface Brightness of Galaxies 635

B. E. LAURENT, B. BONNEVIER, and P. CARLQVIST / On the Dynamics of the Metagalaxy 639

JUAN G. RODERER / Tearing Down Disciplinary Barriers 659

HANNES ALFVÉN

SPECIAL ISSUE DEDICATED TO PROFESSOR HANNES ALFVÉN ON THE OCCASION OF HIS 80TH BIRTHDAY, 30 MAY 1988

Guest Editors

CARL-GUNNE FÄLTHAMMAR, GUSTAF ARRHENIUS, BIBHAS R. DE, NICOLAI HERLOFSON, D. ASOKA MENDIS, and ZDENĚK KOPAL

HANNES ALFVÉN AT EIGHTY

Ten years ago, *Astrophysics and Space Science* dedicated its May issue to Professor Hannes Alfvén on the occasion of his seventieth birthday. Time has now come to honour him on his eightieth birthday by publishing this special issue of *Astrophysics and Space Science*.

The preface of the May 1978 issue outlined his life and work until then. In the decade that has passed since that time, the outstanding significance of his pioneering work has become even more evident. The waves that he discovered long ago are of fundamental importance to the whole field of plasma physics, in the laboratory as well as in space. Likewise, the concept of gyrocentre drift and the perturbation theory based on it are powerful tools for studying charged particle motion, and are being widely used and developed to ever higher sophistication. The critical velocity phenomenon in the interaction of a plasma and a neutral gas, which was proposed by Alfvén with amazing intuition – subsequently observed experimentally and only much later explained by theoreticians – has attracted sharply increased interest in recent years, following its observation in space.

The latest decade has not only verified the importance of Alfvén's past work, it has also been a decade in which he has continued a very active scientific life with new contributions to science especially in the field of cosmogony. For example, he has made use of the latest data from the Voyager spacecraft to test, with remarkable success, his theoretical predictions about the detailed structure of the Saturnian rings.

Hannes Alfvén is, in the true sense of the words, a pioneer and an explorer who has opened new vistas in science. His early contributions earned him the Nobel Prize for Physics in 1970. These contributions date back to the 1940's and were at that time highly controversial and largely ignored by the scientific community. Much later the progress of experiments and observations have proved his ideas to be right.

However, Hannes Alfvén never rests on his laurels. In recent years his mind has turned to the 'Plasma Universe', ranging in space from the Earth's own neighbourhood throughout the dephts of the Cosmos, and in time to the events in the past that led to the formation of the solar system. Some of his concepts about the Plasma Universe are

Astrophysics and Space Science **144** (1988) 1–2.

already receiving observational support. Others are still controversial, but, given the record of his previous contributions, one may well find them to be validated by future observations.

Not only has he contributed to science by his own work, but Hannes Alfvén has also inspired a generation of young scientists worldwide. Everybody, including his critics, agrees that he is a fascinating and inspiring speaker. Especially in the last few years, his invited talks at major international conferences have been exceptionally well received.

Most remarkably, at eighty Hannes Alfvén remains vigorously active both in his profession and in human affairs. We and his many other friends worldwide hope that for many years to come we shall have the privilege of his collaboration, his inspiration, and his contributions to science.

GUSTAF ARRHENIUS
BIBHAS R. DE
CARL-GUNNE FÄLTHAMMAR
NICOLAI HERLOFSON
ZDENEK KOPAL
D. ASOKA MENDIS

OBSERVATIONS OF PROPAGATING DOUBLE LAYERS IN A HIGH CURRENT DISCHARGE*

LENNART LINDBERG

Department of Plasma Physics, The Royal Institute of Technology, Stockholm, Sweden

(Received 30 September, 1987)

Abstract. Observations of current disruptions and strong electric fields along the magnetic field in a high-density (2×10^{19} m^{-3}), highly-ionized, moving, and expanding plasma column are reported. The electric field is interpreted in terms of propagating, strong electric double layers (3–5 kV).

An initial plasma column is formed in an axial magnetic field (0.1 T) by means of a conical theta-pinch plasma gun. When an axial current (max 5 kA, 3–5 kV) is drawn through the column spontaneous disruptions and double-layer formation occur within a few microseconds.

Floating, secondary emitting Langmuir probes are used. They often indicate very high positive potential peaks (1–2 kV above the anode potential during a few μs) in the plasma on the positive side of the double layer. Short, intense bursts of HF radiation are detected at the disruptions.

1. Introduction

Current limitation in plasmas due to formation of electric double layers is a fundamental phenomenon of great importance in various branches of plasma physics, e.g., space physics and solar physics (Alfvén, 1981, 1986; Block, 1978; Carlqvist, 1986) as well as in technical applications.

Most laboratory investigations of double layers (DL) have been made in stationary or quasi-stationary plasmas, using sophisticated data averaging technique to allow comparison with theories. See Palmadesso and Papadopoulos (1979), Michelsen and Rasmussen (1982), Schrittwieser and Eder (1984). Only a few experiments with high-density, pulsed plasmas have been reported. Lutsenko *et al.* (1975) observed indications of DLs by means of external, capacitively coupled probes. Torvén and Babič (1975), in early experiments on current disruptions in low pressure arcs ($n_e < 5 \times 10^{17}$ m^{-3}), observed sharply localized potential jumps by means of capacitive probes. Inuzuka *et al.* (1985) have studied formation of strong DLs in an axial pinch. Takeda and Yamagiwa (1985) and Yamagiwa and Takeda (1987) claim to have observed short-lived DLs with extremely high voltage and reversed polarity.

2. Experimental Device

We have used an electrodeless plasma gun (conical theta-pinch) to create an initial, highly ionized plasma column in an axial magnetic field (0.1 T) (Figure 1). The current in the theta-pinch coil oscillates as a damped 300 kHz sinusoid, beginning at $t = 0$. The plasma is mainly ejected during the second half period. Thereafter, another condenser

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

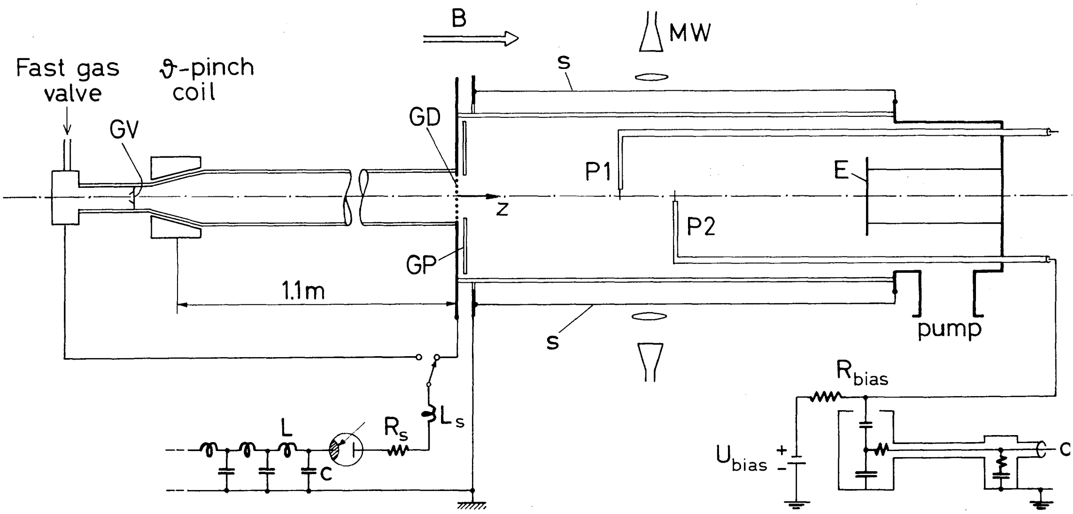


Fig. 1. Cross-section of apparatus. Theta-pinch plasma gun produces an argon plasma column in an axial magnetic field, 0.1 T. 20–50 μs later an axial current is drawn through the column by a capacitor bank of transmission-line type, between GD (or gas valve GV) and the grounded end plate E (at z = 740 mm). Base vacuum < 0.2 mPa. Double layers are observed by floating probes P_1, P_2 connected to capacitive voltage dividers. A microwave transmission link is placed between the probes. The axial z-coordinate is measured from GD, in the flow direction.

bank, of transmission-line type, is switched in to set up an axial current through the plasma column (C = 7 μF, L = 0.4 μH, $R_s \approx 1\ \Omega$, $U_b = \pm 3$ to 5 kV). If short-circuited, the bank would give a constant current $U_b/(R_s + \sqrt{L/C})$ for about 70 μs, which then decays without reversal. We have chosen this current value well above what corresponds to the thermal electron current density in the plasma column.

The end electrode E (Figure 1) is always connected to ground through four symmetrically arranged metal strips s with low inductance. In the case illustrated in Figure 1 the condenser bank is charged negative, and either a coarse grid GD, or the end surface of the fast gas valve GV, in the plasma gun is used as cathode. L_s represents the stray inductance or any additional inductance in the circuit. In this configuration E is anode and the current flows antiparallel to the motion and expansion of the plasma. Experiments are also made with positive polarity (parallel current), with E as cathode, GV as anode. Whenever GV is used as electrode the grid GD is removed. A glass plate GP with 65 mm aperture prevents radial spreading of the discharge.

Experimental parameters are: the kind of gas (H_2 or Ar), the amount of gas admitted to the gun, and the time at which the axial current is initiated. We choose this time short, 20–50 μs to be sure the plasma is free from contact with walls. This, however, has the consequence that the plasma is still in motion, expanding from the gun, during the axial discharge.

The plasma gun has previously been described in several papers, e.g., Danielsson and Brenning (1975), Lindberg (1978), Brenning *et al.* (1981). According to previous measurements the plasma is sent out in a narrow time interval at time t_{em} (≈ 2 μs) in the second half cycle of the theta-pinch current, and the ions expand approximately as free particles. At a distance s from the gun, the plasma (ion) velocity is approximately described by a hyperbola

$$v_i \approx s/(t - t_{em}), \tag{1}$$

(for $t > t_{em} + 2$ μs); and the density increases monotonically during the time of interest. Raadu (1979) made a theoretical study of the expansion process and the space and time evolution of the distribution functions.

Records of four signals are stored in two Biomation 8100 waveform recorders with sampling interval 0.2 μs. To achieve maximum transient response anti-aliasing filters are not used, but the sampling instants of the four channels have been carefully synchronized by means of individual signal delay cables, so that all samples are taken synchronously with an accuracy within 10 ns. Hence, we are sure that any difference, e.g., between two probe voltage signals is true.

Probe voltage changes are measured by means of high impedance, purely capacitive voltage dividers, 4 : 40 000 pF (Craggs and Meek, 1954). As one method to achieve an operating point not too far from the plasma potential, a bias circuit U_{bias}, R_{bias} (Figure 1), has been used in some of the experiments (Lindberg and Kristoferson, 1970). Probe characteristics are further discussed in the Appendix.

3. Experimental Observations

Transmission of 8 mm microwaves indicates that the plasma density approaches or even exceeds the cut-off density (2×10^{19} m^{-3}) at the time when the axial discharge is switched on. According to previous probe measurements the electron temperature is 5–10 eV. Hence, the Debye length is ≈ 0.01 mm and the random electron current density env/4 ≈ 0.3 to 1×10^6 A m^{-2}, corresponding to 1 to 3 kA in the plasma column. When the condenser bank can deliver a current well above this value current disruptions and DL formation occur at every shot when the current reaches a critical value.

Figure 2(a) shows discharge current, microwave transmission and potential variations on two floating probes for a typical shot with negative U_b using the grid GD as cathode – i.e., with the axial current antiparallel to the plasma stream. The first disruption is well reproducible and deep. The current is reduced to 10–20% of the initial value. Later follows a sequence of stochastic, smaller disruptions. The microwave transmission approaches slowly cut-off before the axial discharge, but is restored unexpectedly fast after the disruption. The biased probes are early driven down to near zero volts by the first arriving, very thin plasma (2×10^{16} m^{-3}). The probe signals can be interpreted as follows: a DL, which is formed at the disruption upstream of both probes, arrives at 38 μs to the first probe, moves further downstream and passes the second probe. It is then followed by three more DLs. To demonstrate the existence of a DL between the probes the difference of the probe voltages is calculated and plotted at the bottom of

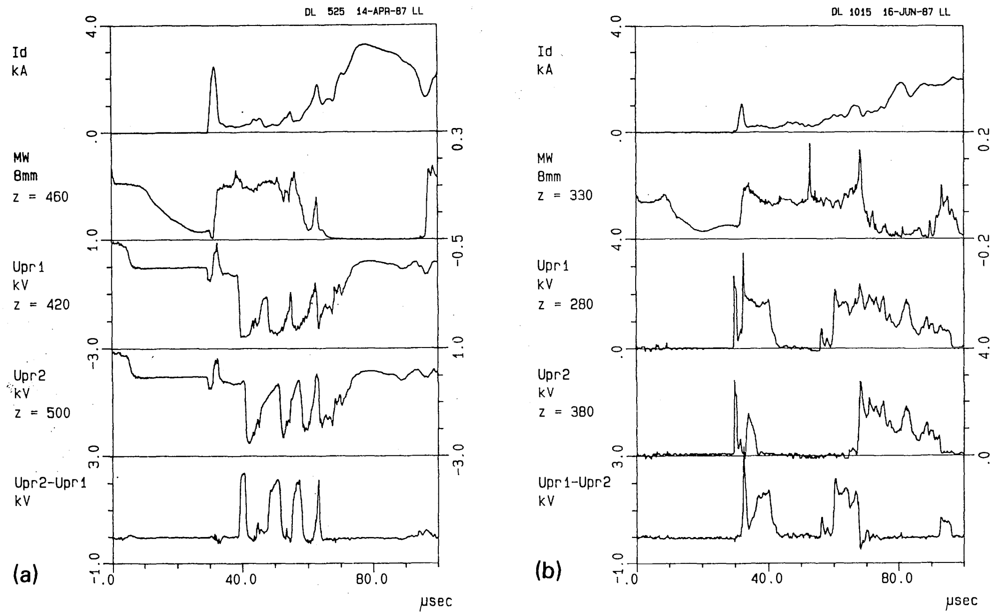


Fig. 2. Discharge current, 8 mm microwave transmission, and voltages of two probes as functions of time, 0–100 μs after the theta pinch discharge. Axial discharge initiated after 30 μs. (a) $U_0 = -3$ kV, GD cathode, current anti-parallel to plasma flow. (b) $U_b = +3$ kV, GV anode, current parallel to plasma flow. *Bottom*: Calculated probe voltage differences show presence of DL between probes. In (a) four DLs propagate downstream approximately with the plasma flow velocity, in (b) a DL propagates first upstream, then downstream.

Figure 2(a). From the time of flight between the probes the velocity of propagation is found to be approximately equal to, or somewhat lower than the local expansion velocity (Equation (1)).

Positive probe voltage pulses, ≈ 1 kV, appear in Figure 2(a) at the disruption. They were first interpreted as a consequence of the applied bias voltage, which would make the probe voltage rise to $+1$ kV in case the plasma density went down so much that the saturation electron current would not suffice to keep its voltage at the plasma potential (< 2 mA, $n_e < 2 \times 10^{16}$ m^{-3}). This is very unlikely, and this interpretation had to be given up definitely when the positive pulses were observed also when no bias was applied.

Experiments with positive polarity, using GV as anode are illustrated in Figure 2(b). The probes indicate now a DL formed downstream of both probes (right in Figure 1) which first moves upstream, passing both probes, then downstream, again passing both probes. Often the current disruption is not as deep as with negative polarity, and the current recovers more rapidly so that the voltage across the DL drops as $(U_b - R_s I_d)$.

To illustrate the DL propagation more convincingly, Figure 3 shows measurements with four probes placed on the centre line. (The two middle ones are rather close together.) The three calculated differences are shown below. For negative polarity,

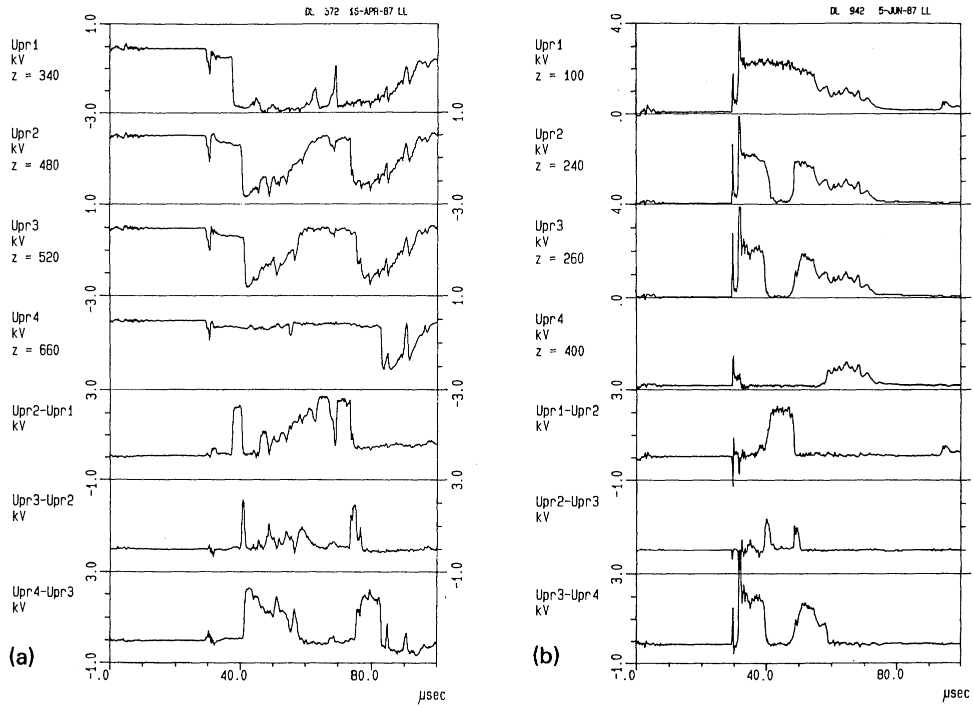


Fig. 3. Four probes on the centre line indicate propagation of DLs. (a) Anti-parallel current. At 37 μs a DL passes probe 1 and, thereafter, probes 2 and 3. At 70 μs another DL makes a similar passage of all four probes. In both cases the propagation velocity is approximately equal to the plasma flow velocity, like in Figure 2(a). (b) Parallel current. A DL forms downstream of the probes and propagates upstream, passing the probes, then downstream like in Figure 2(b).

Figure 3(a) shows two DLs propagating downstream, starting at 37 μs and at 70 μs, respectively, in both cases approximately with the local plasma velocity. The interval between these events is difficult to interpret. For positive polarity Figure 3(b) shows the upstream-downstream propagation as in Figure 2(b).

The thickness of the DL is of the order 2 cm or maybe much less, judging from the propagation velocity and the very rapid probe voltage steps often observed (in one sampling interval), in particular when hot thermionically emitting probes are used (see, e.g., Figure 4(b)).

Two more examples of positive probe pulses are shown in Figures 4(a) and 4(b). No probe bias is used. In Figure 4(a) GV is used as cathode and supplied with -3 kV. In this case the positive pulse appears at the second, small disruption on the second probe when the double layer is between the probes. In this case both probes 1 and 2 are heated to thermionic emission (saturation current > 10 mA), which improves the rise time of the probe. The calculated voltage difference is 4 kV, i.e., 25% more than the voltage applied to the condenser bank, and an even higher percentage with respect to the actual voltage across the discharge if the drop across R_s is taken into account.

Figure 4(b) shows another example, when the condenser bank voltage is increased

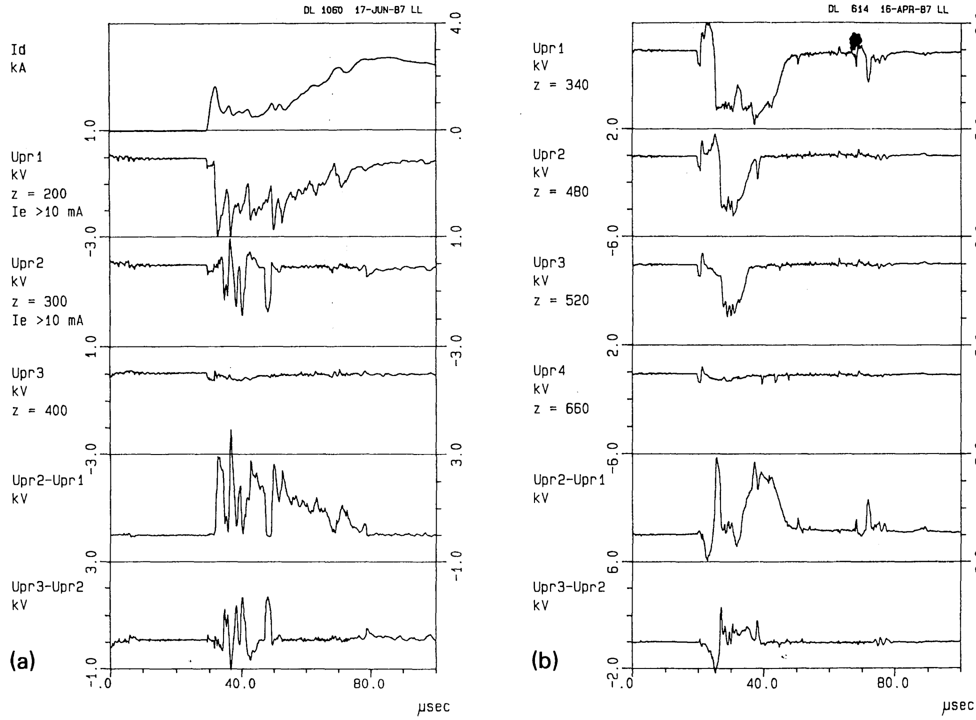


Fig. 4. Discharges with negative polarity, current anti-parallel to flow (a) $U_b = -3\ kV$; (*b*) $U_b = -5$ kV. Unexpectedly high positive probe pulses are observed, in (a) on probe 2, in (b) on probes 1 and 2, with a time difference. The pulses occur on the positive side close to a DL, and increase the voltage across the DL above the applied voltage (see also Figure 2(a)).

to -5 kV and GD is used as cathode. The DL obviously forms upstream of probe 1 which goes 2 kV positive for several μs. The second probe has its positive peak, 1.5 kV, a little later, when the DL has passed probe 1 and is located between probes 1 and 2. While probe 1 is at maximum there is a reversed voltage difference between probes 1 and 2.

Positive probe pulses occur most frequently when the axial current is ignited early (at $t < 30$ μs), i.e., when the plasma has low density and high velocity. However, there is no doubt the plasma has reached the anode E, because it has already carried a high current before the disruption.

To investigate the transverse structure and coherence of the DLs three floating probes placed in a plane at the same z and at equal distances from the axis (20 or 30 mm) have been used. The signals are similar on a slow time-scale but quite different on a fast scale. To a certain extent this may depend on different properties of the probes, but we think it is mainly an effect of a filamentary structure of the discharge. One process contributing to a filamentary character is the localized emission of the electrons from cathode spots; the electrons are then tightly bound by the magnetic field. Mechanisms in the plasma itself may also lead to filamentation.

In the microwave transmission records, e.g., Figure 2(b) short spikes often appear at the first disruption and later, in a statistical manner. When the transmitter is turned off, spikes are still observed, obviously due to short transient bursts of hf-emission.

4. Discussion

Several of the preliminary results presented in this paper are difficult to understand and would require much more careful investigations to be clarified. The very rapid restoration of the 8 mm microwave transmission after a disruption seems confusing since the inertia of the argon ions should prevent any rapid change of plasma density. Filamentation or pinching might make the plasma partly transparent to the microwaves. Local DL-formation in filaments or local pinching could be the cause of the small, secondary disruptions seen in the current oscillograms (Figure 2). Pinching of the whole current or local pinching of current filaments might occur, since our current is not far from what Bennett's (1934) relation requires. Interferometric measurement of the plasma density distribution in space and time and high-speed photography might help to clarify these problems as well as to relate the observed propagation of the DLs with the distribution functions (Raadu, 1979).

The very high positive potential pulses observed close to the DL, Figures 2(a), 4(a), and 4(b) are remarkable in view of the fact that only negative voltage is applied to the plasma. Obviously the voltage across the DL is 'amplified' some 40%. A probable explanation could be based on the redistribution of electron energies caused by the beam-plasma interaction that takes place in the 'anode'-plasma close to a DL. In a previous investigation on a quasi-stationary DL (Torvén and Lindberg, 1980) the interaction of the beam of electrons accelerated in the DL with the thermal 'anode'-plasma was studied. A local hf-field was found some 100 Debye-lengths from the DL with frequency spectrum around the plasma frequency ($\approx$ 500 MHz). Measurements with direction-sensitive probes and a retarding field analyzer indicated that a great deal of the beam electrons got their energy drastically changed already in a single passage through the hf-field (Lindberg, 1982a, b). The hf-field must consequently be quite strong; yet very little hf was radiated (a consequence of the very different wavelengths, 5–10 mm of the Langmuir waves in the plasma, 0.6 m of a free electromagnetic wave).

In the present experiments the plasma frequency is tens of GHz and the electron beam has 3 to 5 keV energy. When this is thermalized a tail of very energetic electrons is produced, tending to leave the interaction region which then should acquire a positive potential. The beam-plasma interaction is a stationary process as long as the DL carries current and, hence, should generate a region of positive potential adjacent to the DL, with tendency to expand towards the anode. A necessary condition is then that microarcing between plasma and anode does not occur (Robson and Thonemann, 1959; Mioduszewski *et al.*, 1980; Ecker, 1983), at least not immediately. Inspection of the stainless steel anode after some time of operation, during which we are sure the current never was reversed, showed clear marks of cathode spots – i.e., that microarcing had occurred (probably with some delay and possibly not at every shot). This proves the existence of a very high electron temperature and a plasma potential much higher than the anode potential.

In a stationary DL-experiment, under quite different conditions, Maciel and Allen (1985a, b) or Allen (1985) observed DLs with different potential structures, some of them with a positive hump. They also found that ions from that region, hitting the wall, had quite high energies, and proposed this could be an effect of HF-oscillations in the anode plasma. Beam-plasma interaction in the anode plasma is likely the cause of the positive potential pulses observed in our experiment as well as the energetic ions observed by Maciel and Allen (Allen, 1987).

5. Conclusions

We have demonstrated the occurrence of high voltage (3–5 kV) potential drops along the magnetic field in a highly ionized, high density (2×10^{19} m^{-3}) plasma column. The likely interpretation is the formation of electric double layers. It is obvious that electric field measurements parallel to the magnetic field are important also in high density plasmas.

The current disruptions are reproducible and often very deep; for certain parameters the current is reduced to 10% of the value before the disruption.

New features are the observation of positive local potential maxima (magnitude 40% of the DL voltage) close to the DL, and sudden radiation of hf-bursts. Like most plasmas in laboratory and in space our plasma shows a detailed fine structure both in space and time.

Acknowledgements

The author wishes to thank Drs S. Torvén, P. Carlqvist, and M. Raadu for stimulating discussions and encouragement, and Mr J. Wistedt and B. Johansson for very skilful construction of equipment and technical assistance. The work was supported by the Swedish Natural Science Research Council.

Appendix

For potential measurements we have to rely on probes, and it is necessary to discuss how the floating potential of a probe is related to the plasma potential. Figure 5 shows characteristics of a cylindrical probe in principle. The floating potentials ($i = 0$) are marked with circles.

(a) Is the classical Langmuir probe characteristic in a plasma at rest with Maxwellian distributions. In our plasma the electron saturation current i_s to a probe with $d = 0.1$, $l = 5$ mm is $i_s = env_e \pi dl/4 \approx 1$ A. The floating potential is typically $5\,kT_e/e$ below the plasma potential U_{pl} for argon (Chen, 1965).

(b) In a streaming plasma the ion current is enhanced due to the directed motion of the ions, and the floating potential moves closer to the plasma potential.

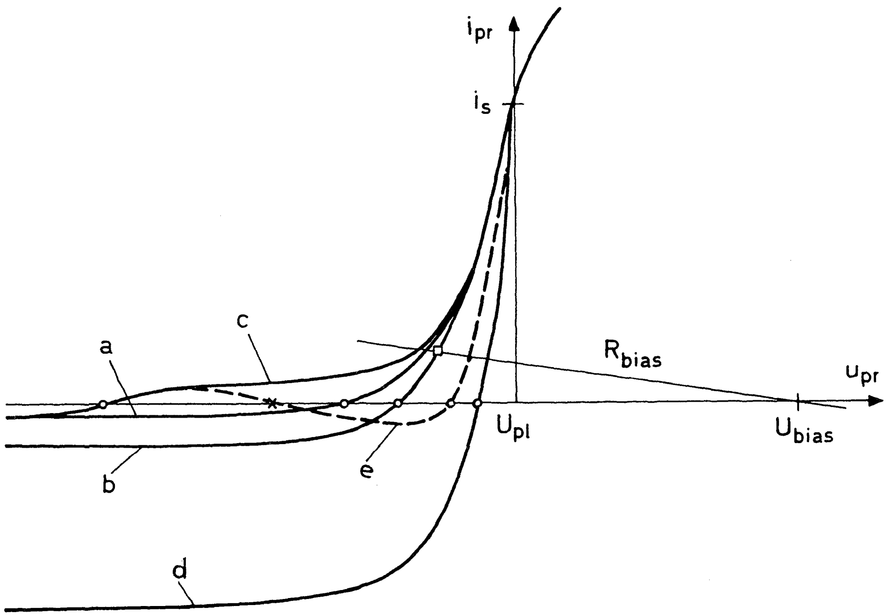


Fig. 5. Probe characteristics under different conditions.

(c) If the plasma contains a beam of energetic electrons, which is the case in the 'anode plasma' close to a DL, the floating potential will become low, approximately the same is in the cathode plasma. In our case we estimate the beam current as a fraction of the total current I_d in the plasma column (area $A \approx 30$ cm^2): $i_b \approx I_d dl/A \approx 0.2$ A if $I_d = 1$ kA. At a deep disruption I_d may go down an order of magnitude but on the other hand, due to filamentation, the local current density may become much higher.

(d) A hot, thermionically emitting probe can, with sufficient emission, overcome an electron beam and attain a floating potential close to the plasma potential. In a dense plasma like ours it is, however, questionable if sufficient emission can be achieved.

(e) A probe hit by a beam of energetic electrons will emit more secondary electrons than it collects and, hence, operate as an emitting probe. The secondary emission coefficient δ exceeds 1 for most materials when the primary energy is above a few hundred eV (Bruining, 1954; v. Engel, 1955; Wang *et al.*, 1986). The secondary emission increases in proportion to the primary current. If this is very intense a cathode spot may form (microarcing). In any case the probe will take a floating potential not far below the plasma potential.

(f) A suitably biased probe (Figure 1) may have its operating point on the steep part of the probe characteristic (square in Figure 5) and can be used to distinguish between a beam of charged particles and a plasma (Lindberg and Kristoferson, 1970).

We consider also the case when a thin DL rapidly passes a probe so that the probe suddenly enters a region with a different potential. The probe has a certain capacitance C_g to ground, which has to be re-charged to change the probe voltage. If the probe suddenly enters a lower potential (negative step) the electron saturation current from the plasma, which is large, will quickly bring the probe potential down to or a little below the plasma potential. If instead the plasma potential makes a positive step, ion saturation current will continue to flow, but this is low. The formation of an electron retarding sheath goes quickly and results in a capacitive current and build-up of a surface charge on the probe which we can estimate. According to elementary plane probe theory (Chen, 1965) the potential in the sheath varies in the steady state as $V = V_0 \exp(-x/\lambda_D)$ where V_0 (< 0) is the probe potential relative to the plasma and λ_D the Debye-length. Hence, the electric field at the probe surface $E = V_0/\lambda_D$ and the final surface charge q on the probe when the sheath has formed is $q = \varepsilon_0 V_0 \pi dl/\lambda_D$ (if we for simplicity treat our probe as plane).

This charge will change the voltage on the probe by $\Delta V = q/C_g$. Assuming $d = 0.1$, $l = 5$ mm, $\lambda_D = 0.01$ mm, and $C_g = 50$ pF, we find $\Delta V/V_0 \approx 0.04$. Hence, the capacitive current to the exposed probe surface gives an almost negligible response. Capacitive coupling through the insulation around the probe lead can be estimated as through an ordinary capacitance.

Finally we consider the case with a high-frequency fluctuating plasma potential (still $f \ll f_{pl}$). A non-emitting probe would then take an average potential in the vicinity of the lowest instantaneous plasma potential. An emitting probe would alternatively draw saturation electron current from the plasma and emit saturation current, and could hardly take an average potential above the average plasma potential unless the emission saturation current is extremely high.

We conclude that in none of the cases discussed could the probe potential become higher than the plasma potential – simply because the probe then would collect saturation electron current, which would quickly bring the probe potential below the plasma potential. The positive probe pulses observed when a negative voltage is applied to the plasma should therefore also correspond to positive pulses in the plasma potential.

References

Alfvén, H.: 1981, *Cosmic Plasma*, D. Reidel Publ. Co., Dordrecht, Holland.
Alfvén, H.: 1986, *IEEE Trans. Plasma Sci.* **PS-14**, 779.
Allen, J. E.: 1985, *Plasma Phys. Contr. Fusion* **27**, 1343.
Allen, J. E.: 1987, Private discussion at ICPIG XVIII, Swansea, U.K.
Bennett, W. H.: 1934, *Phys. Rev.* **45**, 890.
Block, L. P.: 1978, *Astrophys. Space Sci.* **55**, 59.
Brenning, N., Lindberg, L., and Eriksson, A.: 1981, *Plasma Phys.* **23**, 559.
Bruining, H.: 1954, *Physics and Applications of Secondary Electron Emission*, Pergamon Press, London.
Carlqvist, P.: 1986, *IEEE Trans. Plasma Sci.* **PS-14**, 794.
Chen, F. F.: 1965, 'Electric Probes', in R. H. Huddlestone and S. L. Leonard (eds.), *Plasma Diagnostic Technique*, Academic Press, New York.
Craggs, J. D. and Meek, J. M.: 1954, *High Voltage Laboratory Technique*, Butterworth, London.
Danielsson, L. and Brenning, N.: 1975, *Phys. Fluids* **18**, 661.
Ecker, G.: 1983, *IEEE Trans. Electrical Insulation* **EI-18**, 243.
Inuzuka, H., Torii, Y., Nagatsu, M., and Tsukishima, T.: 1985, *Phys. Fluids* **28**, 703.
Lindberg, L.: 1978, *Astrophys. Space Sci.* **55**, 203.

Lindberg, L.: 1982a, 'Experimental Observation of Local Beam-Plasma Interaction Near a Double Layer', in H. Wilhelmsson and J. Weiland, *Proc. 1982 Int. Conf. Plasma Physics*, June 1982, Gothenburg, Sweden. Chalmers Univ., Technology, contributed paper, p. 37.

Lindberg, L.: 1982b, 'Experimental Studies of Electron Scattering due to Beam-Plasma Interaction near a Double Layer', in P. Michelsen and J. J. Rasmussen, *Proc. Symp. Plasma Double Layers*, Risø Nat. Lab., Denmark, Rep. R-472, p. 164.

Lindberg, L. and Kristoferson, L.: 1970, *Plasma Phys.* **12**, 831.

Lutsenko, E. I., Sereda, N. D., and Kontsevoi, L. M.: 1976, *Soviet Phys. Tech. Phys.* **20**, 498.

Maciel, H. S. and Allen, J. E.: 1985a, 'The Different Space Charge Structures of Free Double-Layers Formed in a Low Pressure Arc Discharge', in *Proc. XVII Int. Conf. Phenomena in Ionized Gases*, Budapest, p. 151.

Maciel, H. S. and Allen, J. E.: 1985b, 'Measurement of Ion Energies in a Discharge with a Double Layer', in *Proc. XVII Int. Conf. Phenomena in Ionized Gases*, Budapest, p. 155.

Michelsen, P. and Rasmussen, J. J. (eds.): 1982, in *Symp. on Plasma Double Layers*, Risø Nat. Lab., Roskilde, Denmark, June 1982, Rep. Risø R-472.

Mioduszewski, P., Clausing, R. E., and Heatherly, L.: 1980, *J. Nucl. Materials* **91**, 297.

Palmadesso, P. J. and Papadopoulos, K. (eds.): 1979, 'Wave Instabilities in Space Plasmas', in *Proc. Symp. XIX URSI Gen. Assembly*, Helsinki, July 1978, D. Reidel Publ. Co., Dordrecht, Holland.

Raadu, M.: 1979, *Plasma Phys.* **21**, 331.

Robson, A. E. and Thonemann, P. C.: 1959, *Proc. Phys. Soc.* **73**, 508.

Schrittwieser, R. and Eder, G. (eds.): 1984, in *Proc. Sec. Symp. on Plasma Double Layers and Related Topics*, July 1984, Innsbruck, Austria.

Takeda, Y. and Yamagiwa, K.: 1985, *Phys. Rev. Letters* **55**, 711.

Torvén, S. and Babič, M.: 1975, 'Current Chopping Space Charge Layers in a Low Pressure Arc Plasma', in J. G. A. Hölscher and D. C. Schram (eds.), *Proc. of the Twelfth Int. Conf. on Phenomena in Ionized Gases*, contributed Papers, p. 124, North Holland Publ. Co., Amsterdam.

Torvén, S. and Lindberg, L.: 1980, *J. Phys. D. Appl. Phys.* **13**, 2285.

v. Engel, A.: 1955, *Ionized Gases*, Clarendon Press, Oxford.

Wang, E. Y., Hershkowitz, N., and Intrator, T.: 1986, *Rev. Sci. Instr.* **57**, 1085.

Yamagiwa, K. and Takeda, Y.: 1987, *J. Phys. E. Sci. Instr.* **20**, 332.

CRITICAL IONIZATION VELOCITY INTERACTION: SOME UNSOLVED PROBLEMS*

N. BRENNING and I. AXNÄS

Department of Plasma Physics, Royal Institute of Technology, Stockholm, Sweden

(Received 4 August 1987)

Abstract. Different problems of current interest regarding the critical ionization velocity (C IV) phenomenon are discussed. The article is divided into five sections corresponding to different aspects of the interaction: velocity, magnetic field strength, geometry, neutral gas density, and time duration. In each section, experiments and theories – microscopic and macroscopic – are discussed.

1. Introduction

The critical ionization velocity (C IV) phenomenon was introduced by Alfvén (1954) as a hypothesis: that a strong interaction should be expected if the relative velocity between a magnetized plasma and a neutral gas exceeds the critical value v_c, defined by

$$m_n v_c^2/2 = eU_i \,, \tag{1}$$

where m_n and U_i are the mass and ionization potential of the neutrals and e is the electronic charge. The process has been observed in a number of experiments, and has been extensively reviewed in the literature (e.g., Danielsson, 1973; Newell, 1985). The purpose of this paper is not to present yet another review, but to focus on some unsolved questions. The following five sections each discuss one aspect of C IV interaction, and in each section the discussion draws upon both experiments and theory. The sections are: 2. Velocity; 3. Magnetic field strength; 4. Geometry; 5. Neutral gas density; 6. Time duration.

The theoretical models for the C IV interaction can be crudely split up into two groups, microscopic and macroscopic. The difference is best illustrated from the simplest form of the electron energy equation

$$\frac{\partial}{\partial t}\,(n_e \langle W_e \rangle) = \nu_i n_e \left(\eta \,\frac{m_n v^2}{2} - eU_i \right), \tag{2}$$

where n_e and $\langle W_e \rangle$ are the density and average energy of the electrons, and ν_i is the electron impact ionization frequency. The energy transfer factor η in Equation (2) denotes the overall efficiency of energy transfer from the neutral atoms to the electrons at ionization. Some authors define it to include other aspects of the interaction, e.g., loss of hot electrons, spectral line excitation, or energetic tail formation in the electron energy distribution.

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 15–30.

By *microscopic* theories we here mean those that are mainly concerned with the efficiency of the energy transfer. The *macroscopic* theories are those that take some value of η as given and solve the energy equation, together with some form of the momentum equation, as a function of time (and sometimes of space).

2. Velocity

All the experimental observations of the value of the critical velocity have been made in laboratory experiments (i.e., none in space). Although the results are commonly referred to in terms of *one* critical ionization velocity, the laboratory observations are best described by splitting them up into three groups.

(1) The *threshold velocity* v_t for interaction to start in impact experiments: an initial velocity that must be exceeded in order for the C IV process to occur. In these experiments, a plasma is shot at a neutral gas cloud of limited spatial extent. Consider the initial phase of plasma penetration, at a time when electron impact ionization dominates over the seed ionization (e.g., photoionization or charge exchange collisions, which trigger the interaction), but before the velocity has begun to decrease ($v = v_0$). In this phase of the interaction Equation (2) applies. A positive energy flow to the electrons is necessary to prevent the interaction from dying out. This is only possible when the right-hand side of Equation (2) is positive, which gives the familiar expression

$$v_0 > v_t = \eta^{-1/2} v_c . \tag{3}$$

Theoretical estimates of η put it somewhere around 0.7 (Raadu, 1978; Galeev, 1981). An inclusion of the energy loss through line excitation would typically reduce the effective η-value by a factor of 2, giving threshold velocities about 70% above v_c. The experiments confirm this estimate. The impact experiments (Figure 1) give threshold velocities that lie 50–100% above v_c.

(2) The *final velocity* v_f in impact experiments is the velocity below which no further deceleration is seen, even though the plasma penetrates further into the neutral gas. If momentum coupling to the walls is neglected, as well as charge exchange and elastic ion-neutral collisions, a value of v_f can be derived as follows. Consider a volume of plasma, with boundaries that follow the plasma stream. Let the volume contain N_i original plasma ions and pick up N_{in} ionized neutrals. We assume that the interaction is local in the sense that there is no flow of energy or momentum across the boundaries. For simplicity we disregard the energy change in the original electron population. Instead of the instantaneous value η (Equation (2)) of the energy transfer factor, we consider the whole interaction time and denote by η_2 the fraction of the released energy that is used for ionization of the neutrals and heating of the new electrons to the average energy $\langle W_e \rangle$ – i.e.,

$$\eta_2(N_i m_i(v_0^2 - v_f^2)/2 - N_{in} m_n v_f^2/2) = N_{in}(eU_i + \langle W_e \rangle) , \tag{4}$$

in which m_i is the ion mass in the plasma stream and m_n is the neutral mass. Equation (4)

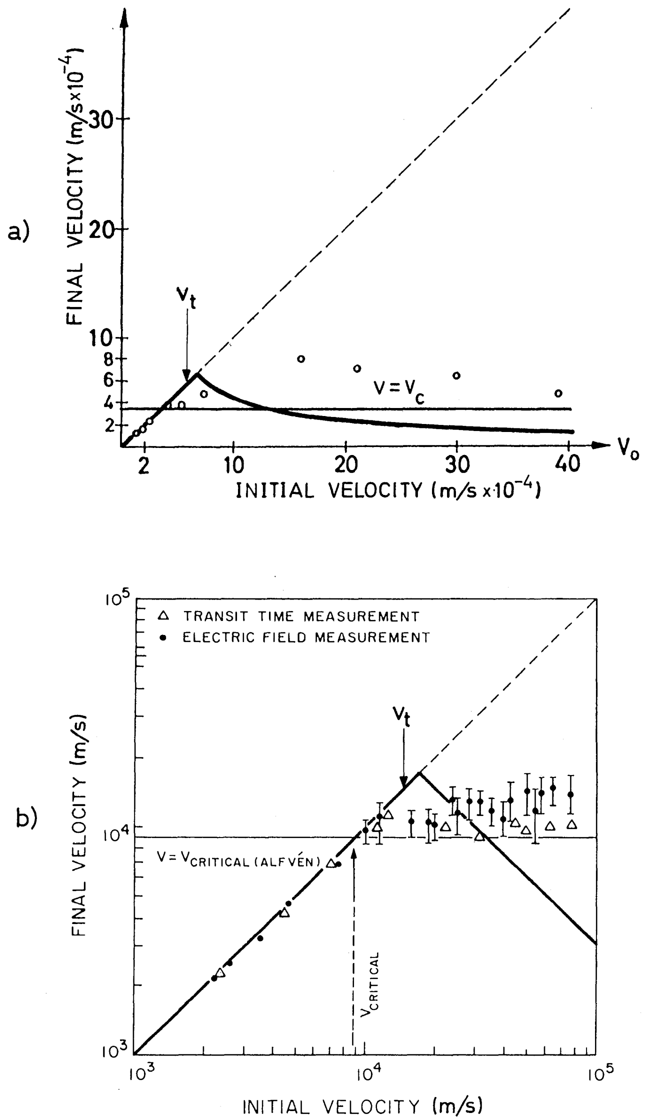


Fig. 1. A comparison between the experimental results and the theoretical expressions for the threshold velocity v_t and the final velocity v_f in impact experiments. The graphs show the velocity inside the neutral gas cloud as a function of the velocity in the undisturbed plasma stream. (a) Danielsson (1970): hydrogen plasma against neutral helium. (b) Mattoo and Venkataramani (1980): hydrogen plasma against neutral argon. The arrows show v_t from Equation (3), with $\eta_{\rm eff} = 0.7/2 = 0.35$ as motivated below Equation (3). The solid lines show v_f from Equation (6), with the factor $((\eta_2)/(1 + \langle W_e \rangle / eU_i))^{1/2} = 0.25$.

together with the condition $N_{in} \geq 0$ and the momentum equation

$$N_i m_i v_0 = N_i m_i v_f + N_{in} m_n v_f \tag{5}$$

gives the final velocity

$$
\begin{aligned}
v_f &= v_0 && \text{when} \quad v_0 < \left(\frac{1 + \langle W_e \rangle / eU_i}{\eta_2}\right)^{1/2} v_c\,, \\
v_f &= \left(\frac{1 + \langle W_e \rangle / eU_i}{\eta_2}\right) \frac{v_c}{v_0} && \text{when} \quad v_0 > \left(\frac{1 + \langle W_e \rangle / eU_i}{\eta_2}\right)^{1/2} v_c\,.
\end{aligned}
\tag{6}
$$

Notice that the ion mass m_i disappears in the final result. In this sense, Equation (6) agrees with Alfvén's original hypothesis and with experimental results. However, there is also a disagreement: Equation (6) predicts very low final velocities, when v_0 is high. For example, $v_0 = 10v_c$ as in the experiments by Danielsson and Brenning (1975) would, with $\eta_2 = 0.5$ and a final value of $\langle W_e \rangle = eU_i$, give a final velocity $v_f = 0.4v_c$. The physical reason for this low final velocity is illustrated in Figure 2 (from Brenning, 1982a) which shows a spatially resolved solution of Equation (2) for thermal electrons, together with the momentum equation, under the assumptions listed before Equation (4). The net energy flow to the electron population becomes negative at the penetration depth indicated by the broken line, where $v = \eta^{-1/2} v_c$. The continued

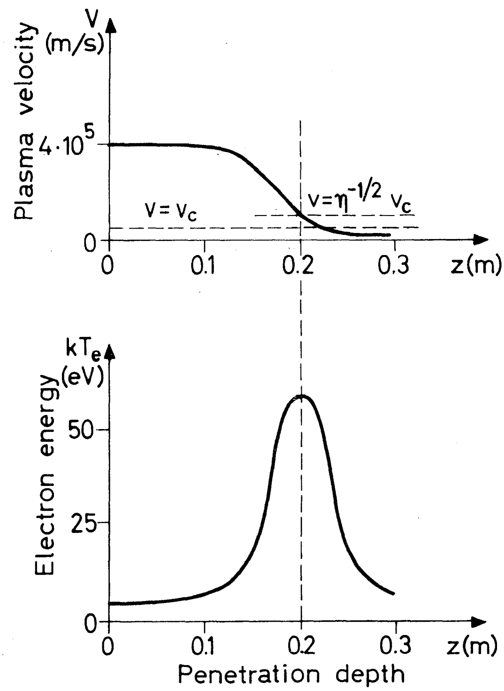


Fig. 2. The variations of plasma velocity and electron temperature as a hydrogen plasma with $kT_e = 5$ eV enters a uniform helium ($n_{He} = 10^{20}$ m^{-3}) cloud. The energy transfer efficiency is $\eta = 0.25$, and the electron population is kept Maxwellian. The broken line indicates the penetration depth at which the energy flow to the electrons becomes negative, $v < \eta^{-1/2} v_c$. The continued ionization and deceleration is an 'afterburn' effect of the large heating in the beginning of the penetration. (From Brenning, 1982a.)

ionization and deceleration is an 'afterburn' effect due to the large electron heating in the initial phase of the interaction. Varma (1978), who was the first to distinguish between the threshold and final velocities, did not, however, include this afterburn effect in his analysis.

A comparison between Equation (6) and some experimental results is made in Figure 1, with the factor $\eta_2/(1 + \langle W_e \rangle / eU_i) = 0.25$.

(3) At the *steady state velocity* v_s in the discharge experiments, the electron energy balance is often dominated by resonant charge exchange collisions between ions and neutrals. Adding these to Equation (2) gives

$$\frac{\partial}{\partial t} (n_e \langle W_e \rangle) = \nu_i n_e \left(\eta \frac{m_n v^2}{2} - eU_i \right) + \nu_{ch} n_i \eta \frac{m_n v^2}{2} , \tag{7}$$

where ν_{ch} is the charge exchange collision frequency. This corresponds to the steady state condition

$$v_s = \left(\frac{\eta}{1 + (\nu_{ch}/\nu_i)} \right)^{-1/2} v_c . \tag{8}$$

The discharge experiments where the ionization and charge exchange frequencies are best known are those by Axnäs (1981, 1982). A value of ν_{ch}/ν_i around ten is not unusual; this would make charge exchange the dominating term in the electron energy balance. Experimental results from Axnäs' (1978) discharge experiments are shown in Figure 3. The observed values usually lie within $\pm 50\%$ of v_c, over a wide range of parameters and for many different gases. In other discharge experiments, values even closer to $_{vc}$ are reported. In view of Equation (8) it is surprising that these experiments so very consistently show Alfvén's predicted value of the critical velocity, even though the value of ν_{ch}/ν_i varies considerably between individual discharges and between gases. This implies that the mechanism that clamps the electric field E at $E/B = v_c$ is rather insensitive to the charge exchange collisions, i.e., insensitive to the *amount* of energy that is released when stationary ions are formed out of neutrals. This could be the case if the efficiency η_{tail} of high-energy ($W_e > eU_i$) tail formation is a function of E/B as illustrated in Figure 4(b); direct heating of the tail has been shown in discharge experiments by Axnäs and Raadu (1983).

To summarize: for the threshold velocity v_t in impact experiments, the deduction of the value $v_t = \eta^{-1/2} v_c$ can be justified on physical grounds. It is also in good agreement with observations. For the final velocity v_f in impact experiments, and the steady-state velocity v_s in discharge experiments, the situation is somewhat confusing: on one hand, it is clear that a deduction *based on energy transfer efficiency* should include the initial velocity in the impact experiments (since v_0 determines the total energy available) and the resonant charge exchange collision in the discharge experiments (where these collisions usually release more energy than electron impact ionization). On the other hand, a first-order inclusion of these effects only makes the agreement of the theoretical values of v_f and v_s with observations worse.

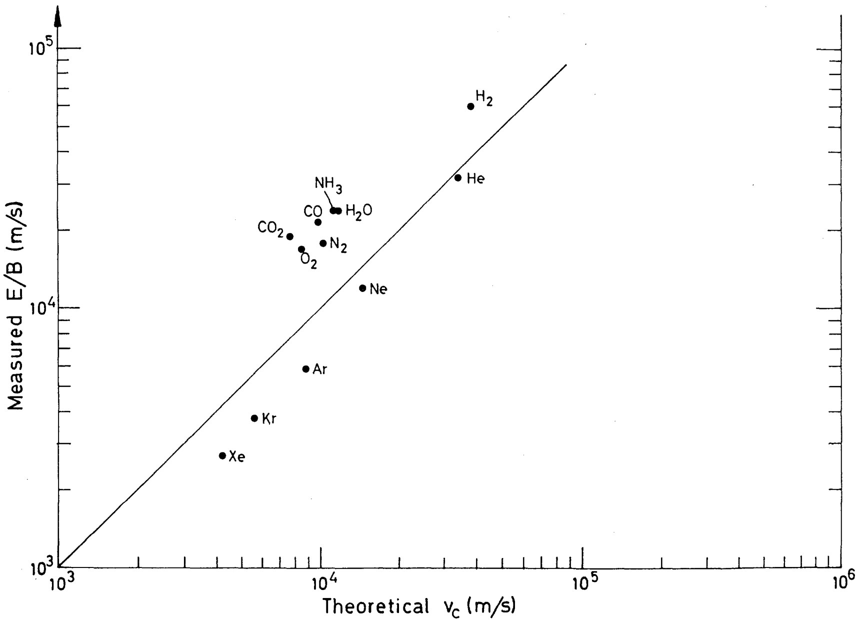


Fig. 3a.

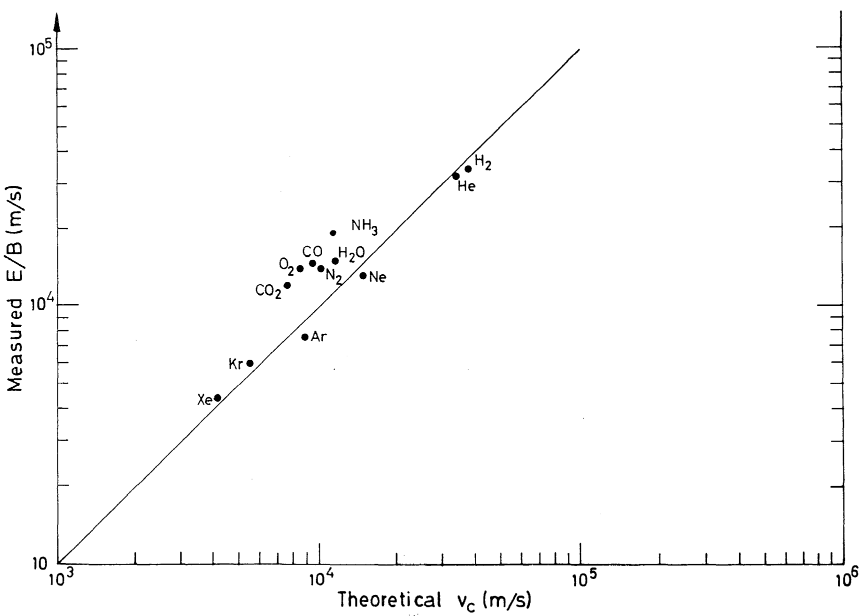


Fig. 3b.

Fig. 3. Experimental values of the steady-state velocity v_s (= E/B) in discharge experiments (from Axnäs, 1978). (a) discharge current 100 A; (b) discharge current 1000 A.

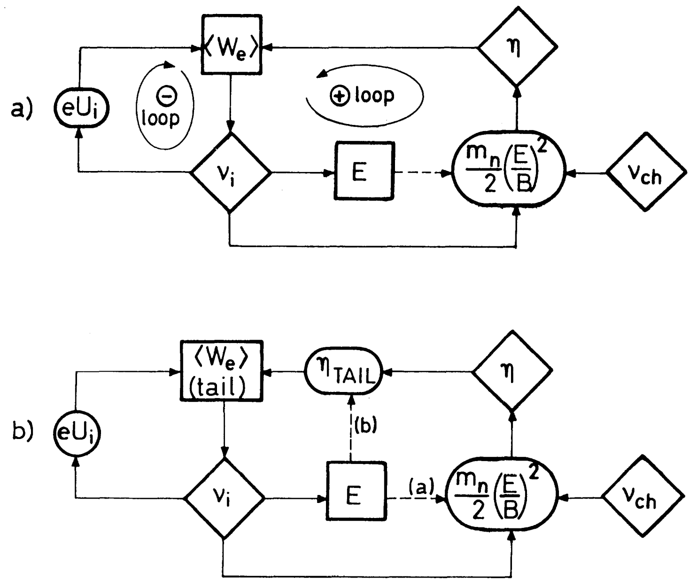


Fig. 4. Two possible ways to regulate the electric field E in discharge experiments. (a) The E/B value is determined by the balance between the positive and the negative feedback loops. The E/B value determines the amount of energy released in each ionization or charge exchange. This is the philosophy behind Equations (3) and (8). (b) The E/B value influences both the amount of energy that is released in each ionization or charge exchange, and the fraction η_{tail} of the energy that goes to electrons energetic enough to ionize. If η_{tail} depends very strongly on E/B as indicated by the broken line denoted by (b), the discharge becomes insensitive of the value of v_{ch}/v_i.

The unsatisfactory situation is that for v_f and v_s, which constitute more than 90% of the observational material, the commonly quoted Equation (3) gives a good estimate of the *numerical value* but does not adequately reflect the *physics* behind that value.

3. Magnetic Field

A transverse magnetic field of some strength was a part already of Alfvéns original scheme for C IV interaction. A survey of laboratory impact experiments (Brenning, 1985) shows that, from the experimental evidence, the relevant condition is that the flow is subalfvénic (this condition was originally proposed by Papadopoulos, 1982):

$$v_0 < v_A \,. \tag{9}$$

In spite of the fact that no carefully diagnosed investigation has been made across the limit $v_0 = v_A$, the experimental evidence is convincing in view of the large differences between the experimental devices: the magnetic field at the limit $v_0 = v_A$ ranges between an externally applied 0.02 T in a thin plasma ($3 \times 10^{17}\,\mathrm{m}^{-3}$) to a frozen-in field of 0.0035 T in a much denser plasma ($10^{19}\,\mathrm{m}^{-3}$); the plasma was in contact with the walls in most devices, but isolated by vacuum in one case; the neutral gas cloud was usually of the same cross section as the plasma flow, but was in one case much smaller; the velocities were above the critical velocity with varying factors up to 10.

Most theories for the laboratory C IV experiments invoke some version of the modified two-stream instability (MTSI) for the energy transfer, although the *direct* experimental evidence is sparse. The experimental determination of $v_0 = v_A$ as a lower limit to the magnetic field should be regarded as another argument in favour of these models, since the MTSI is damped by electromagnetic effects when the flow is superalfvénic. It also shows that, at least in the laboratory impact experiments, no other process provides efficient energy transfer when the MTSI is damped.

The condition (9) applies only to efficient C IV interaction, with a threshold velocity in the vicinity of v_c. A less efficient C IV interaction, with a low efficiency of the energy transfer, is possible also in superalfvénic flows. This situation leads to a threshold velocity much higher than v_c. It is of interest in many astrophysical situations (e.g., Galeev and Chabibrachmanov, 1986).

4. Geometry

It is necessary to include the momentum coupling between the 'C IV-interaction volume' and the surroundings in any macroscopic description. This coupling depends largely on the magnetic field topology. The momentum coupling to the surroundings in the ionosphere is customarily assumed to be 'Alfvén wave coupling' as discussed by, e.g., Scholer (1970) and Haerendel (1982). In laboratory experiments, the magnetic field can close totally (Axnäs, 1981) or partly (Lehnert, 1988) within the interaction volume, connect with walls that are stationary in the neutral gas frame (Venkataramani and Mattoo, 1980; Danielsson and Brenning, 1975), or be isolated from the walls by neutral gas or vacuum (Himmel *et al.*, 1976; Brenning, 1981). In all these laboratory experiments an efficient coupling to the walls would tend to decelerate the plasma flow. A coupling to a large surrounding plasma as, e.g., in the ionospheric release experiments, would work in the opposite direction: a good contact to the surrounding plasma would here counteract the plasma deceleration.

It seems difficult to draw conclusions from laboratory experiments carried out so far about this momentum coupling in the ionospheric experiments; see, however, Lehnert (1988). From theory, the 'Alfvén wave coupling' is expected to give an $\mathbf{i} \times \mathbf{B}$ term in the momentum equation (cf. Goertz *et al.*, 1985)

$$\mathbf{i} \times \mathbf{B} = -2V_{A0}\rho_0\mathbf{v}/H_{\parallel}, \tag{10}$$

where V_{A0} and ρ_0 are the ambient Alfvén speed and mass density, and $H_{\parallel}$ is the extent of the cloud parallel to $\mathbf{B}$. This coupling term leads to an asymptotic state of the ionization, where the mass that is added to a flux tube due to ionization is approximately equal to the ionospheric mass that is swept over by the Alfvén waves that travel outwards along the same flux tube. The ionization rate in this situation is proportional to $(\rho_0)^{1/2}B/H_{\parallel}$ (Goertz *et al.*, 1985). A similar expression is given as a maximum ionization rate by Torbert (1986).

The 'Alfvén wave coupling' above is built on the assumption of infinite conductivity parallel to the magnetic field. The parallel current density that is drawn from the surroundings when an electric field is mapped parallel to **B** in this fashion is given by Mallinckrodt and Carlson (1978)

$$i_{\parallel} = \frac{\operatorname{div} E_{\perp}}{\mu_0 v_{\mathrm{A}}} . \tag{11}$$

A smaller structure thus requires a higher parallel current density for mapping along B. If this current density exceeds some threshold value one can expect instabilities to arise, and maybe also current limitation. So far, there is no firm theoretical prediction of this, particularly in view of the very transient nature of the ionospheric C IV experiments; according to Lysak and Carlson (1981), one can expect the current-driven electrostatic ion-cyclotron (EIC) instability when $v_D/v_i > 13T_i/T_e$, and the Buneman instability when $v_D > v_e$ (where v_D is the electron drift velocity corresponding to the parallel current, and v_e and v_i are the electron and ion thermal velocities). When $v_D > v_e$, double-layer formation could also limit the current.

In case of current limitation, a completely different mechanism of energy transfer could some into play, which resembles a process proposed by Varma (1978). If the electric field within the cloud is so large that it cannot be mapped out along the field lines, there must exist some region with electric field components parallel to **B**. Experience from laboratory plasma stream experiments (e.g., Lindberg, 1978) indicate that the negative side would take the potential of the plasma, while the main part of the stream becomes positive. Electrons could then be accelerated along the field lines into the neutral cloud to energies above the ionization threshold. The energy source is of course the same as in the heating by lower hybrid instabilities: the kinetic energy, in the plasma frame, of the newly ionized neutrals which maintain the potential difference.

Another geometrical characteristic of the ionospheric releases is the injection angle, which has several consequences. One is that it influences the cloud extension $H_{\parallel}$ parallel to **B**, which appears in the Alfvén coupling term of Equation (10), and in the asymptotic ionization rate of Goertz *et al.* (1985). $H_{\parallel}$ might also be important for the electron heating. Most of the recent theories use some version of the modified two-stream instability operating at the 'equal effective mass angle' defined by $\cos\Theta_0 = (\mathrm{m_e/m_i})^{1/2}$. The parallel wavelength $\lambda_{\parallel}$ of the instability at this angle is

$$\lambda_{\parallel} = \pi v_0 m_i/eB , \tag{12}$$

i.e., of the order of the ion gyroradius, calculated with the beam velocity component perpendicular to **B**. A large value of $H_{\parallel}/\lambda_{\parallel}$ would be favourable both for the growth of the instability and for the trapping of the hot electrons in the instability microfields (and prevent them from escaping along **B**). As examples (Newell, 1985), $H_{\parallel}/\lambda_{\parallel}$ is of the order of unity in Star of Lima (0.1% ionization) and more than an order of magnitude larger in Porcupine (30% ionization of the neutrals with $v > v_c$).

5. Neutral Gas

There are two types of constraints on the neutral gas, on the absolute density and on the column density.

The constraint on the *absolute density* is related to the microscopic aspects of the theory, i.e., the energy transfer efficiency that can be achieved. For 'very dense' neutral gas, a resistive heating of the electrons is possible (Machida and Goertz, 1986), and no collective effect is needed for energy transfer to the electrons. Below this density are two distinct regimes of collective C IV interaction: in 'dense' neutral gas, the efficiency η of energy transfer is large, and the limiting velocity is close to the value proposed by Alfvén. In 'thin' neutral gas, there is a less efficient energy transfer, and a higher threshold value for interaction. We will here only discuss the limit between 'thin' and 'dense' neutral gas. The ratio between ionization rate and ion gyro (angular) frequency ν_i/ω_{gi} is here the important parameter (Formisano *et al.*, 1982). In homogeneous steady-state situations, efficient energy transfer requires that

$$\nu_i/\omega_{gi} > 1 \, . \tag{13}$$

The interaction is expected to become gradually weaker when ν_i/ω_{gi} decreases below unity, to reach a constant low value of $\eta = 0.025$ when $\nu_i/\omega_{gi} < (m_e/m_i)^{1/2}$. In the latter case, the ions would form a gyrotropic distribution in velocity space before the instability taps the energy.

If the situation is transient or inhomogeneous (e.g., if the plasma enters a limited region with neutral gas), there is a possibility that the interaction can be triggered, even though the initial value of ν_i/ω_{gi} is far below unity. The question is whether the density of new ions produced during a time much shorter than ω_{gi}^{-1} would give an instability growth rate $\gamma \gg \omega_{gi}$. In this case, a ring-type distribution would not have time to develop (Brenning, 1982b). This gives a limit (Brenning, 1986; a similar condition was derived by Haerendel, 1986)

$$\nu_i/\omega_{gi} > 2(m_e/m_i)^{3/2}\,\kappa^3 \, . \tag{14}$$

Figure 5 shows a schematic illustration of how η can be expected to vary with ν_i/ω_{gi} in steady-state and transient situations. Unfortunately, κ is a rather uncertain parameter: the number of instability growth time-scales required for the modified two-stream instability to tap the energy of a thin ion stream in a plasma. Also, Equation (14) strictly applies only to a fraction of the first gyro-time after ionization has begun. As time goes on, the gyrotropic background of previously ionized neutrals will grow, and there will be a gradual transition to the steady-state case, where Equation (13) applies. A computer simulation of the instability at low values of ν_i/ω_{gi} has been made by Machida and Goertz (1986). With a seed ionization at a rate $\nu_{\text{seed}}/\omega_{gi} = 0.217(m_e/m_m)^{1/2}$, they are well into the regime where one would expect $\eta = 2.5\%$ in a homogeneous steady-state situation. They find a constant value of $\eta = 32\%$ during more than three ion gyro periods ($t = 20\omega_{gi}^{-1}$).

The constraint on the *column density of the neutral gas* is of another nature, and

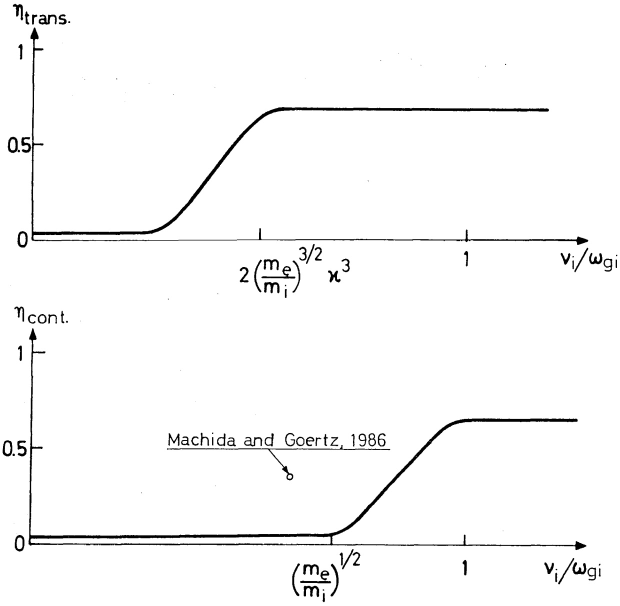


Fig. 5. A schematic illustration of how the energy transfer efficiency can be expected to vary with v_i/ω_{gi}. v_i is here the total production of quasi-stationary neutrals, including electron impact ionization, charge exchange and photoionization. κ is the number of growth time-scales it takes the instability to tap the ion energy (the quasi-linear relaxation time in units of growth time-scales). *Upper curve*: the initial phase of transient interaction. *Lower curve*: homogeneous steady-state interaction (schematically from Galeev, 1981). The circle in the lower graph shows a computer simulation result by Machida and Goertz (1986) for a case between the initial transient and the steady state: they followed the interaction during more than three ion gyro periods ($t = 20\omega_{gi}^{-1}$), and found a constant value $\eta = 32\%$ during this time.

illustrates well the difference between the microscopic and the macroscopic models for C IV interaction. Consider the impact experiments (where the plasma sweeps through a finite neutral gas cloud) or the ionospheric shaped-charge release experiments (where a part of the ionosphere is swept over by a neutral gas cloud). The question is whether the C IV interaction has time to get underway during the transit time (called the injection time by Haerendel). A conservative estimate can be made by requiring that a 'hot' electron shall have time to ionize one neutral atom during the transit time. This is the 'Townsend condition', and it is always fulfilled where C IV interaction is observed.

Another estimate of required neutral gas column density is found by solving the electron energy equation (some extended version of Equations (2) or (7)), together with some form of the momentum equation, to decide whether the interaction has time to be triggered during the transit time. Such calculations were first made by Axnäs (1980), who used the momentum equation including charge exchange and resonant charge exchange collisions for a discussion of a proposed Xe release experiment in the ionosphere. The situation with a limited gas cloud within a larger plasma was treated by Brenning, 1982c. In both these cases, the purpose was to see if the interaction would have time to be *triggered*, i.e., if heating of electrons to the ionization energy could occur during the transit time. Calculations of this type that followed the interaction into the ionization stage were made by Brenning (1982a) for the laboratory impact experiments, and by Goertz *et al.* (1985) and Torbert (1986) for the ionospheric release experiments. The result of these calculations is that the laboratory experiments could be explained well, but that there is a significant discrepancy between the theoretical models and the very efficient ionization seen in the shaped-charge explosion ionospheric Porcupine experiment.

The heart of the problem is that the models predict that the ionization should switch off a few kilometers from the explosion point. The total amount of ionized neutrals is, for energy conservation reasons, limited. It can have at most the same mass as that part of the surrounding ionosphere which is involved in the momentum exchange process described in Section 4. That mass increases as the cube of the distance from the explosion point. If the ionization were to swtich off a few kilometers from the explosion point, the high ionization seen in Porcupine would be energetically impossible. In order to explain Porcupine, some mechanism is needed that keeps the ionization going at longer distances from the explosion point; a distance of around 10–15 km would be sufficient. A common feature in the models discussed above is that they are one-dimensional, and convection-free in the plasma frame. Any process that convects energy towards the front of the neutral stream would give larger ionization than these models, where a large part of the transit time is used to get the electrons hot enough to ionize (Brenning, 1982c).

The conclusions of this discussion of the neutral gas density are two: (1) there is an uncertain condition on v_i/ω_{gi} for efficient energy transfer during the initial phase of a transient C IV situation, and (2) there is a strong disagreement between the results observed in Porcupine and those predicted by one-dimensional models without convection terms.

6. Time Duration

Figure 6 shows that some minimum time duration is necessary for the establishment of steady state in the C IV experiments. Axnäs and Raadu (1983) studied how the E/B value and spectral line intensity responded to a step-wise current increase (Figure 6(a)). The discharge adjusted to the higher current in roughly one ionization time-scale, which in these experiments was also close to the ion gyro time. In the impact experiments (Figure 6(b)), there is a transient when the plasma stream first hits the neutral cloud, followed by a plateau where E/B remains at a constant value. The duration of the transient increases with the ion mass (Venkataramani, 1981); it is of the order of the ion gyro time, which in these experiments is also of the order of the ion transit time.

Figure 7 shows the total time duration of different proposed C IV applications. In some cases, like the shuttle glow (Papadopoulos, 1984) or the production of the Io plasma torus (Galeev and Chabibrachmanov, 1986), the interaction is more or less constant in time. Other cases are of a transient nature: the plasma is a short pulse as in most laboratory impact experiments, or the neutral gas density decays rapidly as in the ionospheric releases.

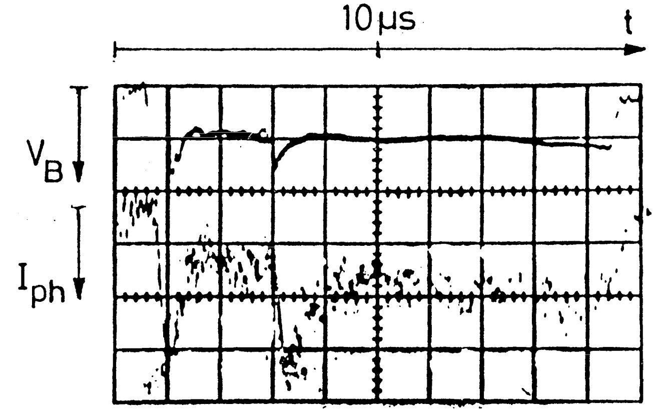


Fig. 6a.

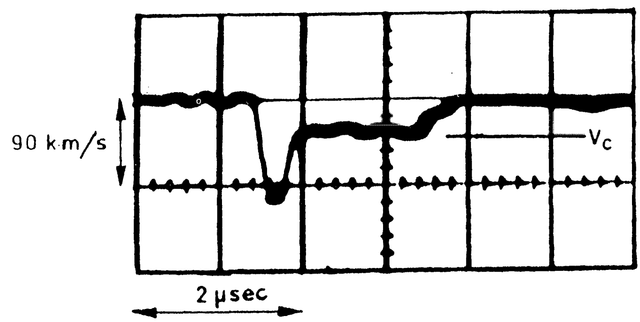


Fig. 6b.

Fig. 6. Transients observed in laboratory C IV experiments. (a) The discharge experiment by Axnäs and Raadu (1983). *Upper curve*: burning voltage (arbitary units); *lower curve*: He I spectral line (3889 Å) intensity. The discharge starts at 2 μs, and the current is abruptly increased (a factor of 2) at 6 μs. (b) The measured transverse electric field in a plasma flow of 4 μs duration which impacts on a neutral cloud. The scale is given in units of velocity (from $v = E/B$). (From Danielsson and Brenning, 1975.)

Regardless of whether one chooses to compare with the ion gyro or the ion transit time, the ionospheric release experiments are by far the most short-lived. The time duration is determined by the rapid decrease in neutral density. It can be estimated from the crude assumption that the neutral cloud is spherically symmetric, expands with a velocity v_{ex}, and has a directed 'beam' velocity v_b. Under these assumptions the neutral gas density decreases by a factor 2 during a time $t_d = (2^{1/3} - 1)R/v_{ex}$, where R is the instantaneous radius of the cloud. With the transit time $t_{tr} = 2R/v_b$, we get

$$\frac{t_d}{t_{tr}} = 0.13 \frac{v_b}{v_{ex}} . \tag{15}$$

Spatial extension ↓ / Time duration →	A few ion transit times (or ion gyro times)	More than ten ion transit times (or ion gyro times)	Steady state
Limited plasma, limited neutral gas.		Laboratory impact experiments.	Laboratory ExB discharge experiments.
Limited neutral gas, larger plasma.	Ionospheric experiments with shaped charge explosions.	Proposed Ionospheric releases from the Space shuttle.	Io plasma torus; Space shuttle glow; Comet interaction with solar wind.
Large plasma and neutral gas.			Alfvéns solar system theory; Solar wind - interstellar gas interaction.

Fig. 7. Overview of the geometrical and temporal differences between some proposed cases of CIV interaction.

Since $v_b = 10v_{ex}$ in a typical shaped-charge explosion, the duration of these experiments is always close to the transit time. Judging from the laboratory experiments, it is, therefore, doubtful if they can ever reach a steady state. The situation would be much improved in releases from orbiting spacecraft (Möbius *et al.*, 1979; Murad *et al.*, 1986), where the expansion velocity can be much lower.

Another effect of the short time duration in the ionospheric experiments was pointed out by Torbert (1986): one can expect a very small degree of ionization in the neutral cloud, even if the CIV interaction is just as efficient as in the laboratory experiments.

7. Summary

The preceding five sections have treated five different aspects of the CIV interaction, and concentrated on remaining unsolved problems concerning the phenomenon. The main conclusions are:

Velocity: all experimental observations of the *value* of the critical velocity come from the laboratory experiments. At least 90% of these observation are of the final velocity v_f in impact experiments or of the steady-state velocity v_s in discharge experiments. For none of these does there exist a clear theoretical explanation why the value is so close to Alfvén's prediction from 1954. Only for the threshold velocity v_t in the impact experiment does there exist a satisfying link between experimental results and theory.

Magnetic field: efficient C IV interaction, with a threshold velocity in the vicinity of v_c, has only been seen in experiments where the plasma flow is subalfvénic. This is in agreement with the theories which use the lower hybrid instabilities to heat the electrons; these instabilities are damped by electromagnetic effects in superalfvénic flows.

Geometry: C IV interaction has been proposed in situations with so widely different topologies that they probably cannot be covered by one theory. A particular example is the ionospheric C IV experiments, where the interaction region is coupled to a much larger volume of surrounding plasma. This might possibly change the whole nature of the interaction: electron acceleration along the magnetic field lines might enhance, or even replace, the heating by the lower hybrid instability.

Neutral gas: the ratio $v_{\text{seed}}/\omega_{gi}$ determines the efficiency of the energy transfer during the initial phase of a transient C IV situation. In order to trigger the interaction, the seed rate $v_{\text{seed}}/\omega_{gi}$ (and, therefore, the absolute density of the neutral gas) has to exceed some limit. There is yet no clear theoretical prediction what this limit is. There is also an uncertain condition on column density, which is directly related to the 'Porcupine Puzzle': one-dimensional models without convection terms predict a degree of ionization in Porcupine which is orders of magnitude below the observed.

Time duration: the time duration of proposed applications of the C IV process varies from steady state to very short time-scales, of the order of the ion gyro time and the transit- (or injection-) time. The ionospheric release experiments are in this respect extremely short-lived; laboratory experiments generally show strong transients on comparable time-scales.

Acknowledgements

We wish to thank Professors C.-G. Fälthammar and M. Kelley for valuable discussions. This work has been financed by the Swedish Natural Science Research Council.

References

Alfvén, H.: 1954, *On the Origin of the Solar System*, Oxford University Press, Oxford.

Axnäs, I.: 1978, *Experimental Comparison of the Critical Ionization Velocity in Atomic and Molecular Gases*, TRITA-EPP-78-04, Royal Institute of Technology, S-100 44 Stockholm, Sweden.

Axnäs, I.: 1980, *Geophys. Res. Letters* 7, 933.

Axnäs, I.: 1981, *The Radial Variation of the Ionization in a Coaxial Plasma Gun Operated under Critical Velocity Conditions*, TRITA-EPP-81-07, Royal Institute of Technology, Stockholm.

Axnäs, I.: 1982, 'The Radial Variation of the Ionization in a Coaxial Plasma Gun Operated under Critical Velocity Conditions', in H. Wilhelmsson and J. Weiland (eds.), *Proc. 1982 Int. Conf. on Plasma Phys.*, p. 267.

Axnäs, I. and Raadu, M. A. R.: 1983, 'Rapid Energization of Electrons in a Critical Velocity Discharge Experiment', *Proc. XVI Int. Conf. on Phen. in Ionized Gases*, Düsseldorf, August 29–September 2, 1983, p. 708.

Brenning, N.: 1981, *Plasma Phys.* **23**, 967.

Brenning, N.: 1982a, 'A Criterion for the Possibility of Critical Ionization Velocity Interaction', in H. Wilhelmsson and J. Weiland (eds.), *Proc. 1982 Int. Conf. on Plasma Phys.*, p. 430.

Brenning, N.: 1982b, *Review of Impact Experiments on the Critical Ionization Velocity*, TRITA-EPP-82-14, Royal Institute of Technology, Stockholm.

Brenning, N.: 1982c, 'Comment on the Townsend Condition', *Proc. 'Workshop on Alfvén's Critical Velocity Effect'*, Max-Planck-Institute für Extraterrestrische Physik, Garching, October 11–13, 1982, p. 313.

Brenning, N.: 1985, *Phys. Fluids* **28**, 3424.

Brenning, N.: 1986, *On the Role of the Ionization Frequency to Gyrofrequency in Critical Ionization Velocity Interaction*, TRITA-EPP-86-10, Royal Institute of Technology, Stockholm.

Danielsson, L.: 1970, *Phys. Fluids* **13**, 2288.

Danielsson, L.: 1973, *Astrophys. Space Sci.* **24**, 459.

Danielsson, L. and Brenning, N.: 1975, *Phys. Fluids* **18**, 661.

Danielsson, L. and Kasai, G. H.: 1968, *J. Geophys. Res.* **73**, 259.

Formisano, V., Galeev, A. A., and Sagdeev, R. Z.: 1982, *Planetary Space Sci.* **30**, 491.

Galeev, A. A.: 1981, 'Weak Turbulence Theory of an Enhanced Gas Ionization by the Plasma Flow, *Proc. of an Int. School and Workshop on Plasma Astrophysics*, Varenna, Italy, ESA SP-161, p. 77.

Galeev, A. A. and Chabibrachmanov I. Ch.: 1986, 'The Critical Ionization Velocity Phenomenon in Astrophysics', *Proc. of the Joint Varenna-Abastumani Int. School and Workshop on Plasma Astrophysics*, held in Sukhumi, U.S.S.R., 19–28 May, 1986, ESA SP-251, p. 129.

Goertz, C. K., Machida, S., and Smith, R. A.: 1985, *J. Geophys. Res.* **90**, 12230.

Haerendel, G.: 1982, *Z. Naturforsch.* **37a**, 728.

Haerendel, G.: 1986, private communication.

Himmel, G., Möbius, E., and Piel, A.: 1976, *Z. Naturforsch.* **31a**, 934.

Lehnert, B.: 1988, *Astrophys. Space Sci.* **144**, 31 (this issue).

Lindberg, L.: 1978, *Astrophys. Space Sci.* **55**, 203.

Lysak, R. L. and Carlson, C. W.: 1981, *Geophys. Res. Letters* **8**, 269.

Machida, S. and Goertz, C. K.: 1986, *J. Geophys. Res.* **91**, 11965.

Mallincrodt, A. J. and Carlson, C. W.: 1978, *J. Geophys. Res.* **83**, 1426.

Mattoo, S. K. and Venkataramani, N.: 1980, *Phys. Letters* **76a**, 257.

Möbius, E., Boswell, R. W., Piel, A., and Henry, D.: 1979, *Geophys. Res. Letters* **6**, 29.

Murad, E., Lai, S. T., and Stair, Jr., A. T., 1986, *J. Geophys. Res.* **91**, 10, 188.

Newell, P. T.: 1985, *Rev. Geophys.* **23**, 93.

Papadopoulos, K.: 1982, 'Electron and Ion Driven Discharges with Collective Dissipation', *Proc. 'Workshop on Alfvén's Critical Velocity Effect'*, Max-Planck-Institut für Extraterrestrische Physik, Garching, October 11–13, p. 178.

Papadopoulos, K.: 1984, *Radio Sci.* **19**, 571.

Raadu, M. A.: 1978, *Astrophys. Space. Sci.* **55**, 125.

Scholer, M.: 1970, *Planetary Space Sci.* **18**, 977.

Torbert, R. B.: 1986, private communication.

Varma, R. K.: 1978, *Astrophys. Space Sci.* **55**, 113.

Venkataramani, N.: 1981, 'Interaction of Magnetized Plasma with Neutral Gas', Thesis, Gujarat University, India.

Venkataramani, N. and Mattoo, S. K.: 1980, *Pramana* **15**, 117.

VELOCITY LIMITATION OF A NEUTRAL DUST CLOUD COLLIDING WITH A MAGNETIZED PLASMA*

B. LEHNERT

The Royal Institute of Technology, Stockholm, Sweden

(Received 11 March, 1987)

Abstract. The problem is considered of a cloud of neutral dust moving into a cloud of static plasma which is confined in a magnetic field. Earlier experiments with rotating plasma devices and plasma guns on critical velocity limitation suggest that such limitation could also arise in the case of plasma-neutral dust interaction in cosmos.

Nevertheless further analysis is required to provide a clear picture of the relations between the cosmical and laboratory conditions for plasma-neutral gas and plasma-neutral dust interaction. In particular this applies to the question how to relate the experiments, which are largely in the plasma-physical MHD range, to the cosmical interaction which appears to be mainly governed by kinetic effects.

1. Introduction

It has earlier been suggested by Alfvén (1954) that there arises an enhanced ionization process when a charged and a neutral gas cloud move into each other at a relative velocity which exceeds the critical value $v_c = (2e\phi_i/m_i)^{1/2}$, where e is the electronic charge, ϕ_i is the ionization potential, and m_i the ion mass. The first experimental verification of such a mechanism was presented by Fahleson (1961) and Angerth *et al.* (1962) who showed that the voltage and the associated velocity of a rotating plasma in a flat homopolar device become limited at a value which is close to the velocity v_c. Later such critical voltage and velocity limitation has been observed in a number of experiments as summarized in review articles (Lehnert, 1971; Danielsson, 1973). Such limitation has also been observed in gun experiments where a plasma blob is forced to collide with a limited cloud of neutral gas, in presence of a transverse magnetic field (Wilcox *et al.*, 1964; Eninger, 1966; Danielsson, 1970, 1973).

The critical velocity limitation was suggested by Alfvén in connection with a theoretical model of the formation process of the solar system. In the model non-ionized (neutral) matter is falling towards a magnetized solar body, thereby moving through a plasma cloud which is kept at rest by a confining solar magnetic field. Thus, the ionized cloud consists of a magnetized plasma, and velocity limitation applies to the motion which is perpendicular to the magnetic field lines.

In the present paper the discussion on velocity limitation will be extended to the case of neutral dust clouds. The rôle of cosmic dust and dusty plasmas has recently become an important subject in cosmical physics. A summary will first be given in Section 2 on some experimental results which are relevant to the process of velocity limitation and

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 31–42.

its connection with plasma-wall interaction. This is followed in Section 3 by a discussion on the case of a neutral dust cloud colliding with a magnetized plasma.

2. Experiments on Velocity Limitation by Plasma-Wall Interaction

A number of experiments have earlier been made in Stockholm with rotating plasmas. The first generation of these experiments was based on a flat homopolar device having a nearly homogeneous magnetic field and where the plasma was moving across the magnetic field at a velocity being nearly constant along each magnetic field line (Fahleson, 1961; Angerth *et al.*, 1962). The flat plasma layer which was situated between the end insulator plates then became subject to velocity limitation, apparently within almost the entire volume of the plasma body.

The second generation of experiments was based on a geometry having a strongly inhomogeneous magnetic field, being similar to that of internal ring systems. In such geometry the velocity limitation of a fully-ionized plasma could be studied in the form of plasma-neutral gas interaction which was localized to layers at the end insulator plates, thus not becoming directly associated with the plasma motion in the midplane of the apparatus. These experiments will be described to some detail here, because they connect plasma-wall interaction with the critical velocity phenomenon, and this becomes important to the discussion on plasma-neutral dust interaction.

The experiments in strongly inhomogeneous magnetic field geometries have been performed with various types of axisymmetric as well as non-axisymmetric boundary and wall conditions.

2.1. Axisymmetric Boundary Conditions

The first type of experiments was carried out with the FI device (Bergström *et al.*, 1966; Lehnert, 1971) the geometry of which is outlined in Figure 1. The radial extensions in the midplane of the confinement region are $r_{01} \simeq 0.19$ m and $r_{02} \simeq 0.26$ m. The magnetic field **B** is purely poloidal, i.e., it is situated in planes through the axis of symmetry. The average magnetic field strength $\bar{B}_0$ in the midplane has been varied in a range up to about 1 T. The plasma is generated from neutral gas which originally fills the discharge chamber, by imposing a potential difference ϕ_{12} and an electric current J_{12} between the anode rings and the cathode plate of Figure 1.

The associated electric field **E** is essentially at right angles to the magnetic field **B** and this puts the plasma into rotation around the axis of symmetry, at a velocity **v**. When a quasi-steady state has been reached, there is a fully-ionized plasma within the shaded confinement region of Figure 1, being defined by the magnetic field lines which are tangent to the electrodes. The plasma body then becomes bounded by partially-ionized boundary layers in the perpendicular direction of the field **B**, and by partially-ionized wall layers in the parallel direction of **B**, where the field lines end upon the insulator surfaces.

The following experimental results with hydrogen are of particular interest to the present discussion:

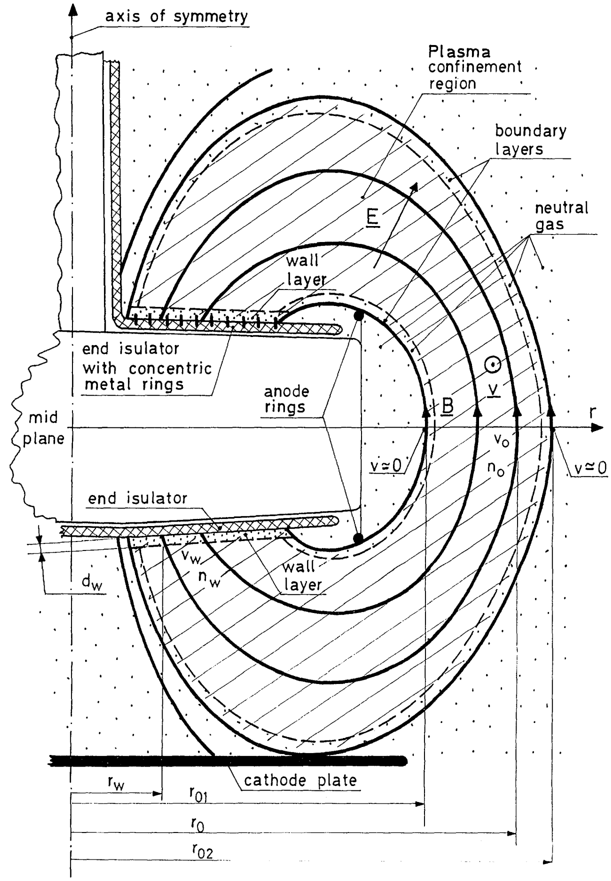


Fig. 1. Outline of the axisymmetric rotating plasma geometry in a device of FI type. The magnetic field **B** is given by the heeavy lines. A potential difference and an electric current are imposed between the anode rings and the cathode plate. This creates an electric field **E** being essentially at right angles to the magnetic field **B**. The plasma within the shaded area is set into rotation around the axis of symmetry, at a velocity **v**. The latter drops nearly to zero at the field lines which are tangent to the anode rings and the cathode plate. Neutral gas surrounds the fully-ionized plasma body. Of special importance in this connection are the thin wall layers at the end insulators, having the thickness d_w. These layers are formed by plasma which recombines at the insulator surfaces and forms a back-stream of neutral gas. In part of the experiments the upper end insulator has been provided with a set of concentric metal rings.

(i) During the first ten microseconds after the start of the discharge, a partially-ionized plasma is created. The voltage ϕ_{12} and velocity v are then bound to ranges which correspond to the velocity limit $v_0 \leq v_c$ of a spoke-shaped plasma body moving through neutral gas in the midplane of the device in Figure 1.

(ii) With an applied electrode voltage exceeding about $\phi_{12} = 5$ kV and a corresponding discharge current $J_{12} \simeq 2$ kA, 'burnout' conditions are reached, under which velocity limitation does no longer occur in the midplane. Thus, the voltage ϕ_{12} begins to increase, and a fully-ionized plasma is created within the shaded region of Figure 1. The plasma has an average ion density $n_0 \simeq 2 \times 10^{21}\ \mathrm{m}^{-3}$ in the midplane. It rotates at angular velocities $\Omega = v/r$ being rather close to those given by the isorotation law. This applies at least to the interior hotter parts of the plasma confinement region. The deviations from isorotation have been estimated by Drake *et al.* (1978) through probe measurements at a point well above the midplane of Figure 1, before and after an abrupt retardation of the plasma rotation. These measurements, which were performed at a power input slightly above the 'burnout' level, indicated that the angular velocity of the plasma near the end insulators can become as much as 40% lower than that in the midplane. Such a deviation from the isorotation law can be explained by the frictional force between the rotating plasma and the end insulators (Lehnert, 1968). However, at a power input being well above the 'burnout' level, and within the interior hotter parts of the plasma, the isorotation law should hold with rather good approximation – i.e.,

$$\Omega \simeq v_0/r_0 \simeq v_w/r_w\,, \tag{1}$$

along a given field line, where (r_0, v_0) and (r_w, v_w) are the corresponding radii and velocities in the midplane and at an insulator surface, respectively. The radial profile of $\Omega_0(r)$ in the midplane of Figure 1 has a maximum somewhere at the centre of this region, and drops almost to zero near the boundaries at $r = r_{01}$ and $r = r_{02}$.

(iii) Due to the isorotation law and the effect of the centrifugal force, the average ion density $\bar{n}_w$ near an insulator surface becomes about an order of magnitude smaller than the average density $\bar{n}_0$ in the midplane. The plasma, which has temperatures up to $T \simeq 10^5$ K at $v_0 \simeq 10^5\ \mathrm{m\ s}^{-1}$ in the central parts of the confinement region, will be partly lost by streaming along the field lines and hitting the end insulator surfaces. Here the ions and electrons recombine and form a backscattered flux of neutral gas having a local temperature $T_{nw} \ll T$. In a steady state this backstreaming neutral gas becomes reionized in the wall layers of Figure 1, thus replacing the plasma which is lost to the insulator surfaces. There is efficient electron heat conduction from the plasma interior along **B** and a correspondingly high ionization rate ξ_w prevails in the wall layers. This fact, in combination with a low temperature ratio T_{nw}/T, results in a wall layer of small thickness

$$d_w \simeq (k_w/\bar{n}_{nw}\xi_w)\,(kT/m_i)^{1/2} \simeq (k_t/\bar{n}_w\xi_w)\,(kT_{nw}/m_n)^{1/2}\,, \tag{2}$$

where k is the Boltzmann constant, $k_w \simeq 1 + k_t$, $k_t = (2\pi)^{-1/2}$, $\bar{n}_{nw}$ and $\bar{n}_w$ are the average neutral and ion densities in the layer and m_n is the mass of a neutral particle (Lehnert, 1968, 1971). In device FI the densities $\bar{n}_w$ and $\bar{n}_{nw}$ are both rather high, thus resulting in $d_w \simeq 10^{-3}$ m. The plasma-neutral gas balance leading to expression (2) and to a small thickness d_w is based on free streaming of charged and neutral particles along **B** and across the wall layer. It should not be confused with the balance conditions leading to the penetration depth (cf. Lehnert, 1974, 1975)

$$L_n \simeq [2kT/m_i\xi\xi_{in}]^{1/2}/n\,. \tag{3}$$

This depth applies to neutrals of mass $m_n \simeq m_i$ which move by diffusion through collisions into a nearly static plasma having the temperature T, the ionization rate ξ, the ion-neutral collision rate ξ_{in}, and the density n. Thus, the parameter d_w becomes related to the wall layers of Figure 1, and L_n to the boundary layers of the same figure.

(iv) By gradually increasing the power input well above the burnout limit, the voltage ϕ_{12} and the velocity v are observed to increase as well, but only to a certain critical level. When this upper level ϕ_{12c} is reached, a further power input does not enhance ϕ_{12} and v but introduces an irregular mode of operation during which large amounts of material are boiled off from the insulator surfaces. Combination of the isorotation law with various experimental data including Doppler shift measurements (Lehnert *et al.*, 1963) indicates that this voltage and velocity limitation sets in as soon as the velocity $v_w = v_0(r_w/r_0)$ reaches Alfvén's critical value v_c within some local region in a wall layer. This occurs in device FI when the maximum of the velocity profile in the midplane reaches the value $v_{0\,\max} \simeq (r_0/r_w)v_c \simeq 3v_c \simeq 1.5 \times 10^5$ m s^{-1}. The onset of the critical velocity phenomenon is clearly demonstrated in these experiments which are performed in a strongly inhomogeneous magnetic field and where the 'radial ratio' r_0/r_w is substantially larger than unity. This makes it possible to allow the electrode voltage ϕ_{12} to approach the critical level ϕ_{12c} slowly from below, where ϕ_{12c} is about three times higher than the voltage ϕ_{12} which would result from critical velocity limitation in the midplane of device FI. A further experimental demonstration of the sharpness of the critical voltage level will be given in Section 2.2.

(v) The velocity limitation in the reported experiments is equivalent to the original picture suggested by Alfvén (1954), but it is here restricted to plasma-neutral gas interaction within thin wall layers.

(vi) Further confirmation of the observations and of the interpretation of the velocity limitation has been provided by a special experiment performed by Berström and Hellsten (1976). This experiment was based on an attempt to suppress the mechanism of critical velocity limitation in the wall layers by short-circuiting an associated azimuthal electric field by means of a set of concentric metal rings, the spacing between which was chosen to be smaller than or equal to the local ion-Larmor radius (Lehnert, 1974). For this purpose 50 ring probes were placed at the upper end insulator surface, as outlined in Figure 1. However, an elimination of the critical velocity limitation process was not achieved in the experiment, but several interesting results were still obtained. First, measurements of the voltage distribution between the rings in a quiescent plasma state revealed a velocity distribution being largely of the form suggested by earlier investigations on discharges being run below the critical voltage level. Second, when voltage limitation was seen to appear, a voltage difference corresponding to a local velocity $v_w = 5 \times 10^4$ m s$^{-1} \simeq v_c$ near the end insulator surface was also recorded at a certain pair of neighbouring rings. Third, the plasma resistance between the same pair was found to decrease drastically when the critical velocity was reached. This reveals that a strong local change of the discharge conditions takes place within the wall layer, just at the position where the critical velocity is first being approached.

To sum up the experimental results with axisymmetric boundary conditions, it is thus found that:

– a *sharply* defined critical velocity phenomenon given by the limit v_c occurs within a *localized* and *thin* partially-ionized wall layer;

– this phenomenon has a substantial effect on the *entire* plasma balance;

– under axisymmetric boundary conditions this phenomenon has been observed in a situation where the plasma **E** × **B** wind is *tangential* to the interacting wall surface and the magnetic field lines *end upon* the same surface.

2.2. Non-axisymmetric boundary conditions

A later series of experiments was performed with the non-axisymmetric FII device (Lehnert *et al.*, 1966),, being outlined in Figure 2. In this geometry the magnetic field lines only end upon an insulator surface within a minor part of the perimeter of the device. Thus, most of the field lines pass freely around the main coil which is supported by an oblique rod covered by an insulator. Movable tubes of ceramics can be inserted in the equatorial plane (midplane). In addition, a turnable tube of ceramics has been used in part of the experiments.

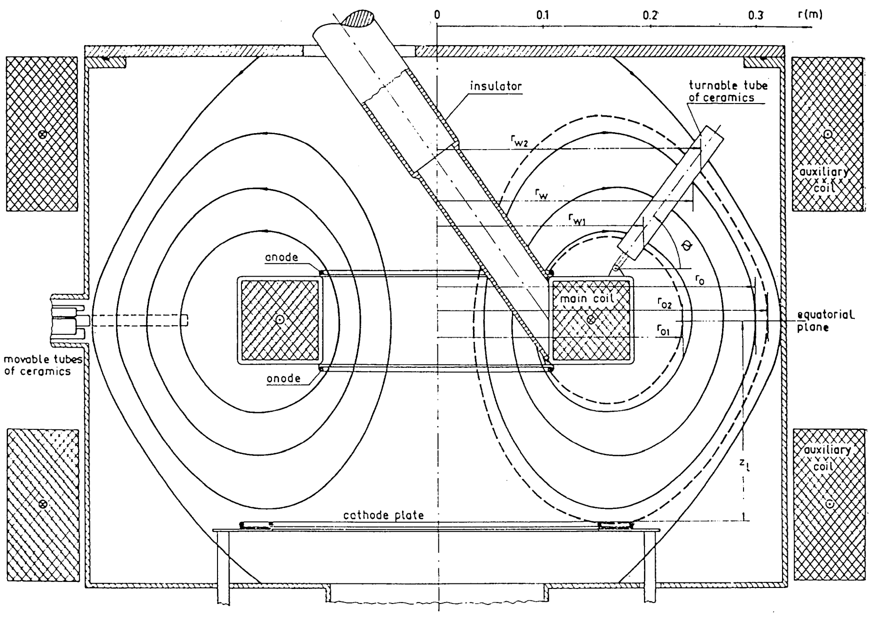


Fig. 2. Outline of the non-axisymmetric rotating plasma device FII. The magnetic field lines pass freely around the main coil, with the exception of $\frac{1}{9}$ of the perimeter where the same lines end upon the support of the main coil. Movable tubes of ceramics can be inserted in the equatorial plane (midplane). In addition, a turnable tube of ceramics has been inserted as an extra wall in part of the experiments.

The following experimental results are of particular interest to the present discussion:

(i) In absence of the movable and turnable tubes, critical voltage and velocity limitation are observed, being of the same character as those occurring in the axisymmetric FI device. Thus, the limitation cannot be avoided, even if $\frac{8}{9}$ of the magnetic flux in the confinement region consist of field lines which do not end upon an insulator surface. The relatively high rotational speed, as compared to the thermal ion velocity along the field lines, makes the oblique rod insulator act like a conical end wall. The strong plasma wind which hits the insulator surface, and the corresponding formation of an immovable back-scattered neutral gas cloud, give rise to a violent plasma-neutral gas interaction and a critical velocity limitation within a thin wall layer at the insulator surface. This occurs as soon as the velocity $v_w = (r_0/r_w)v_0$ reaches v_c at some point along the oblique rod. The transition from a quiescent mode with classical particle and energy losses to a violently disturbed mode with large losses and heavy insulator interaction was found to be extremely sharp, i.e., within a few percent of the corresponding changes in electrode voltage ϕ_{12} and power input into the rotating plasma discharge. This is demonstrated by Figure 3 where burnout is provided by a first condenser bank connected between the electrodes at time $t = 0$. A second bank is connected at a later time $t = t_2$, giving rise to an acceleration of the already fully-ionized plasma. By varying

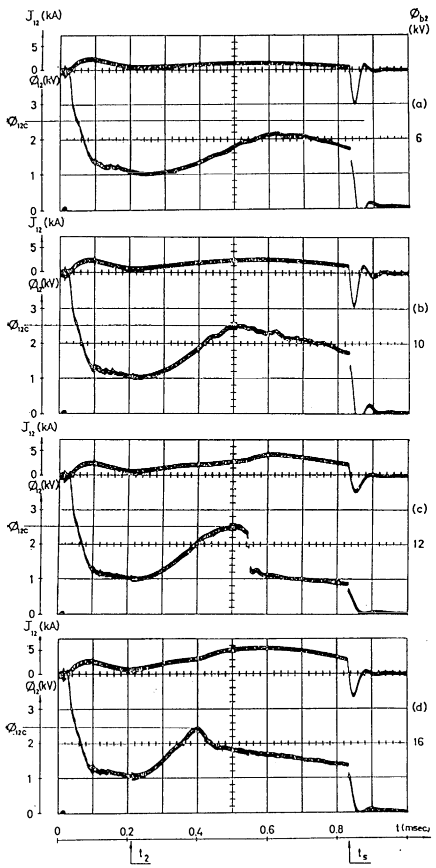


Fig. 3. Plasma voltage ϕ_{12} (lower trace) and current J_{12} (upper trace) in device FII at a fixed voltage of the first condenser bank, and an increasing voltage ϕ_{b2} of the second bank: (a) 6 kV; (b) 10 kV; (c) 12 kV; (d) 16 kV. Turnable tube absent. Critical voltage level ϕ_{12c} indicated by horizontal lines.

the second bank voltage ϕ_{b2} the critical voltage level ϕ_{12c} can be approached more or less rapidly from below. In Figure 3(a) the voltage ϕ_{b2} is too small for reaching the critical level, and the plasma behaves quite regularly, all the way to the time $t = t_s$ where a short-circuit is being applied. In Figures 3(b)–3(d) the voltage ϕ_{b2} has been stepwise increased, to make the plasma reach the critical level ϕ_{12c} at an increasing rate. The plasma is then seen to become strongly irregular and lossy, just after having 'touched' the critical level. This occurs for voltages ϕ_{12} which are seen to be almost exactly determined by the same level $\phi_{12} = \phi_{12c}$ in all Figures 3(b)–3(d).

(ii) When the movable tubes were inserted into the mid-plane, the velocity in this plane immediately became limited to $v_0 \leq v_c$ by violent plasma-wall interaction. This occurred even with such a small tube diameter as 5 mm. The ion-Larmor radius is then somewhat smaller than 1 mm.

(iii) Taking away the movable tubes and inserting the turnable tube, the latter produced a voltage limitation being similar to that caused by the oblique rod. Also in this series of experiments the critical limit was approached slowly from below, but now at various values of the rod angle θ. The results gave a clear picture of a sharply defined velocity limitation at $\phi_{12c} = \phi_{12c}(\theta)$, determined by $v = (r_0/r_w)v_c$ where $r_w = r_w(\theta)$ is the radius of the corresponding field line which ends on the turnable tube. Thus, results similar to those demonstrated by Figure 3 were obtained, but now with a critical level $\phi_{12c}(\theta)$ which decreased from $\phi_{12c}(\theta = 139°) = 2.6$ kV to $\phi_{12c}(\theta = 14°) = 1.1$ kV when the rod angle was varied in the range $\theta = 139°$ to $\theta = 14°$.

To sum up the experimental results with non-axisymmetric boundary conditions, it is thus found that:

– a *sharply* defined critical velocity phenomenon occurs within a localized and thin partially-ionized wall layer, also when this layer is formed around an obstacle of limited *size*;

– even a *small* obstacle gives rise to a substantial effect on the *entire* plasma balance;

– under non-axisymmetric boundary conditions the critical velocity limitation is observed also when the plasma $\mathbf{E} \times \mathbf{B}$ wind is *perpendicular* to parts of the interacting wall surface and only *part* of the magnetic field lines end on this surface.

3. Application to Neutral Cosmic Dust

We now turn to the problem of a cloud of neutral cosmic dust moving into a static cloud of plasma which is confined by a magnetic field, as outlined in Figure 4. Here the consequences of the experiments described in Section 2 will be discussed and some problems of further analysis be identified.

3.1. Ranges of Basic Parameters

3.1.1. *Experimental Conditions*

In the laboratory experiments the mean-free paths for ion-neutral and electron-neutral collisions, the charged particle Larmor radii, the wall layer thickness d_w, the penetration

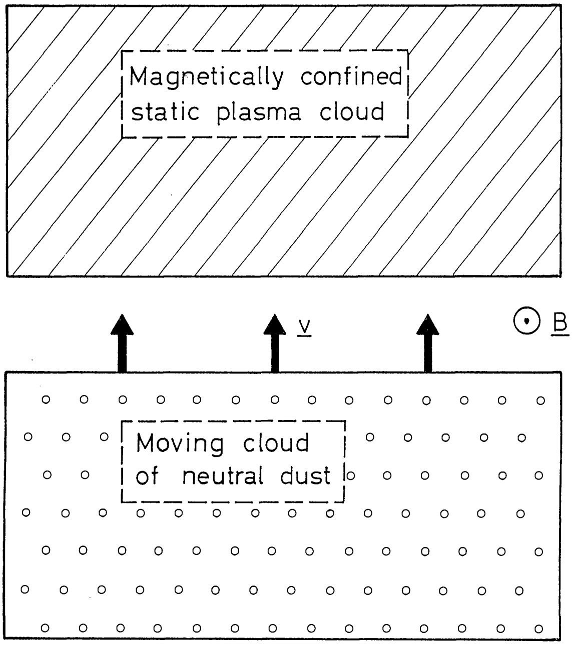


Fig. 4. Schematic picture of a cloud of neutral dust which moves at the velocity $\mathbf{v}$ into a cloud of magnetically confined static plasma.

length L_n of neutrals, and the Debye distance h_D are all comparable to or smaller than the characteristic macroscopic dimensions of the apparatus and the wall geometry. This implies that macroscopic fluid theory can at least be partly adopted when tackling the problems of plasma-neutral gas interaction.

3.1.2. *Cosmical Conditions*

In an application to cosmic dust moving through a plasma, the characteristic parameter ranges may in some cases deviate substantially from those of the described experiments. Thus, with Alfvén (1954) we assume an initial amount of neutral matter per unit volume which corresponds to a neutral density of the order of $n_n = 10^{11}\ \mathrm{m}^{-3}$ in the case of a gaseous state. This figure is related to the amount of matter out of which the solar system and similar systems have to be formed.

In the present discussion on plasma-neutral dust interaction, it is assumed that the same amount of matter now forms a cloud of solid cosmic grains, instead of a cloud of neutral gas. The grains are assumed to have an average radius R, an average separation distance L_g, an average solid state mass density ρ_g, and an average mass number A. The average distance of the grains then becomes

$$L_g = (4\pi\rho_g/3Am_pn_n)^{1/3}\,R\,, \tag{4}$$

where m_p is the proton mass. For $A = 50$ and $n_n = 10^{11}\ \mathrm{m}^{-3}$ we then have $L_g \simeq 10^6\,R$.

Concerning the plasma and neutral gas temperatures T and T_n, there are two possibilities to be observed here:

– In the earlier discussed cases of plasma-neutral gas interaction the electron temperature usually has to be limited, to correspond to a Saha equilibrium with $T \simeq T_n \simeq 2 \times 10^3$ to 10^4 K where the electron densities are in a range of about $n = 10^7$ to 10^9 m^3 (compare Alfvén, 1954). A high plasma temperature would otherwise result in a fully-ionized state, not being subject to plasma-neutral gas interaction by critical velocity limitation.

– In presence of neutral cosmic dust the plasma temperature does not necessarily have to be restricted. It then becomes possible for a dust cloud to move into a fully-ionized plasma cloud of high temperature, and still end up with critical velocity phenomena being somewhat similar to those observed in the experiments of Section 2.

The magnetic field strengths estimated by Alfvén (1954) in connection with the formation process of the solar system are in the range of 10^{-7} to 10^{-5} T.

With these data, and for plasma and neutral gas densities of the order of 10^{11} m^{-3}, the mean-free path of ion-neutral collisions will be of the order of $\lambda_{ni} \simeq 10^8$ m, the ion-Larmor radius a_i about 10 to 10^3 m, the wall layer thickness $d_w \simeq 10^7$ m, and the neutral particle penetration length $L_n \simeq 10^8$ m. Higher local densities will reduce the values of λ_{ni}, d_w, and L_n.

The mechanisms of plasma-neutral dust interaction will depend on the relative magnitude of the grain radii R and separation distances L_g, as compared to the relevant characteristic length parameters, such as λ_{ni}, a_i, d_w, and L_n among others. Introducing L_p as a common symbol of these length parameters, three parameter regimes can be identified with respect to plasma-dust interaction, namely:

(i) $L_p \lesssim R \ll L_g$ (large grain size). There is a nearly mascroscopic fluid behaviour, both for processes near the grain surface and in the space between the grains. This range is equivalent to the conditions of the laboratory experiments described in Section 2.

(ii) $R < L_p < L_g$ (intermediate grain size). There is a nearly macroscopic behaviour of the processes in the space between the grains, but the processes near the grain surfaces are governed by kinetic effects.

(iii) $R \ll L_g \lesssim L_p$ (small grain size). There is a kinetic behaviour of all processes which determine the plasma-dust interaction. To be judged from the estimated cosmical parameter data, this regime is likely to become important in most cases of cosmic interaction.

3.2. The Model of Interaction

When a cloud of neutral cosmic dust grains moves at the velocity $\mathbf{v}$ into a static plasma cloud being confined in a magnetic field $\mathbf{B}$ as demonstrated in Figure 4, the following scenarios can be outlined:

(i) As soon as the grains enter the plasma, ions and electrons will hit the grain surfaces and recombine there. This produces a primary generation of back-scattered neutral gas which then surrounds the grain surface.

(ii) In the case of large-size grains for which the average grain radius is comparable to or larger than both the corresponding wall layer thickness d_w and the ion-Larmor radius, the situation becomes equivalent to that of the plasma-wall interaction studied in the symmetric and non-symmetric experiments described in Section 2. This applies to all possible orientations of the grain surface elements with respect to the direction of the flow velocity $\mathbf{v}$ and the magnetic field $\mathbf{B}$. According to the estimations made in Section 3.1.2, such a situation could also apply to cases where the local plasma and neutral densities substantially exceed $10^{11}\ \mathrm{m}^{-3}$. When the velocity v of the neutral dust cloud approaches the critical value v_c, a violent interaction is thus expected to take place near the grain surfaces.

(iii) In the case of small-size grains, the parameters of dust-plasma interaction will, on the other hand, end up in a regime being quite different from that of so far performed experiments. This is a situation which seems to prevail within the main range of cosmical data outlined in Section 3.1.2. For sufficiently many small grains per unit volume the average separation distance between the grains will become comparable to or smaller than the wall layer thickness d_w. Consequently, the back-scattered neutral gas is then expected to form a cloud which moves with the grains and fills the entire space between them. This leads back to the conditions of a neutral gas cloud moving into a magnetized plasma and being subject to critical vleocity limitation, in the same way as in earlier experiments with plasma guns. The detailed mechanisms which couple the grain motion to the plasma and neutral gas motions should in this case become fully kinetic.

(iv) The case of intermediate-size grains is governed by mechanisms which are, in general terms, a type of mixture between the mechanisms just outlined under (ii) and (iii).

(v) As soon as the mode of critical velocity limitation sets in, matter is expected to be boiled off from the grains at an enhanced rate, thus increasing the amount of neutral gas near the grain surfaces as well as in the spacing between the grains. This 'feedback' mechanism reinforces the plasma-dust interaction, thus leading to further gas release from the grains. As a consequence, there should also arise a retarding force on the neutral dust cloud.

3.3. Further Analysis

Further theoretical and experimental analysis is required to connect the details of the classical type of critical velocity limitation by plasma-neutral gas interaction with those of plasma-neutral dust interaction. Such an analysis includes a number of questions such as:

– the dependence of the plasma-grain interaction and the associated critical velocity limitation mechanisms on the relation between grain radius, wall layer thickness, mean-free paths, Larmor radii, and the Debye distance;

– the time-scale of the onset and development of the interaction process on cosmical and laboratory scale;

– the effects of particle-wall interaction at the grain surface, including the spectrum of recombined and reflected particles, sticking probabilities, and all other wall effects which are involved in the plasma-grain and neutral gas-grain balance (Lehnert, 1970);

– the large Larmor radius (LLR) effects and associated charge separation phenomena which arise in presence of heavy charged particles in a dusty plasma.

4. Conclusions

From the present discussion can be seen that there are experimental as well as theoretical arguments for a cloud of neutral dust to be subject to critical velocity limitation when moving into a static magnetically-confined plasma cloud. Thus, there seem to be many features of this kind of plasma-neutral dust interaction in common with those of the earlier studied plasma-neutral gas interaction at the critical velocity.

The laboratory experiments on critical velocity limitation have been performed within parameter ranges where the plasma mainly exhibits an MHD behaviour. Further analysis is, therefore, required to provide a clear picture of the relations between plasma-neutral gas and plasma-neutral dust interaction, in terms of mixed MHD-kinetic and fully kinetic theory, and by a more detailed analysis of the relevant parameter ranges in cosmos and in the laboratory.

Acknowledgement

The author is indebted to Professor Hannes Alfvén for valuable discussions on the problems treated in this paper.

References

Alfvén, H.: 1954, *On the Origin of the Solar System,* Clarendon Press, Oxford, p. 30ff.

Angerth, B., Block, L., Fahleson, U., and Soop, K: 1962, *Nucl. Fusion* **1**, 39.

Bergström, J. and Hellsten, T.: 1976, *Nucl. Instr. Methods* **133**, 347.

Bergström, J., Holmberg, S., and Lehnert, B.: 1966, *Plasma Physics and Controlled Nuclear Fusion Research*, IAEA, Vienna, Vol. 1, p. 341.

Danielsson, L.: 1970, *Phys. Fluids* **13**, 2288.

Danielsson, L.: 1973, *Astrophys. Space Sci.* **24**, 459.

Drake, J. R., Hellsten, T., and Bonnevier, B.: 1978, The Royal Institute of Technology, Stockholm, TRITA-PFU–78–105.

Eninger, J.: 1966, *Proc. Seventh Int. Conf. on Phenomena in Ionized Gases*, Gradeninska Knijga, Beograd I, p. 520.

Fahleson, U.: 1961, *Phys. Fluids* **4**, 123.

Lehnert, B.: 1968, *Arkiv Fysik* **38**, 499.

Lehnert, B.: 1970, *Cosmic Electrodyn.* **1**, 218.

Lehnert, B.: 1971, *Nucl. Fusion* **11**, 485.

Lehnert, B.: 1974, *Phys. Scripta* **9**, 189.

Lehnert, B.: 1975, *Phys. Scripta* **12**, 327.

Lehnert, B., Bergström, J., and Holmberg, S.: 1966, *Nucl. Fusion* **6**, 231.

Lehnert, B., Bergström, J., Holmberg, S., and Wilner, B.: 1970, *Phys. Scripta* **1**, 39.

Wilcox, J. M., Pugh, E., Dattner, A., and Eninger, J.: 1964, *Phys. Fluids Suppl.* **7**, S51.

DYNAMICAL ASPECTS OF ELECTROSTATIC DOUBLE LAYERS*

MICHAEL A. RAADU

Department of Plasma Physics, The Royal Institute of Technology, Stockholm, Sweden

and

J. JUUL RASMUSSEN

Association Euratom – Risø National Laboratory, Physics Department, Roskilde, Denmark

(Received 30 September, 1987)

Abstract. Electrostatic double layers have been proposed as an acceleration mechanism in solar flares and other astrophysical objects. They have been extensively studied in the laboratory and by means of computer simulations. The theory of steady-state double layers implies several existence criteria, in particular the Bohm criteria, restricting the conditions under which double layers may form. In the present paper several already published theoretical models of different types of double layers are discussed. It is shown that the existence conditions often imply current-driven instabilities in the ambient plasma, at least for strong double layers, and it is argued that such conditions must be used with care when applied to real plasmas. Laboratory double layers, and by implication those arising in astrophysical plasmas, often produce instabilities in the surrounding plasma and are generally time-dependent structures. Naturally occuring double layers should, therefore, be far more common than the restrictions deduced from idealised time-independent models would imply. In particular it is necessary to understand more fully the time-dependent behaviour of double layers. In the present paper the dynamics of weak double layers is discussed. Also a model for a moving strong double layer, where an associated potential minimum plays a significant role, is presented.

1. Introduction

A double layer (DL) is a local region in a plasma which can sustain a potential difference. Essentially it consists of two adjacent layers with equal and opposite net charge. The layer as a whole is globally neutral but has an internal electric field. The structure is determined in a self-consistent manner by the particle dynamics in the electric field set up by the net charge distribution produced by the particles. This is indicated schematically in Figure 1 (see also Figure 3). In general a DL requires both free and reflected ion and electron components. The free particles carry current through the layer and lead to emerging beams of accelerated particles. The hatched areas in Figure 1 denote regimes of ion and electron acceleration. The potential across the DL must in general be maintained by external sources which are the source of energy for the particle acceleration.

Double layers are of interest in astrophysics as a direct means of accelerating particles (Alfvén, 1981, 1986). They can sustain a local region of parallel electric field leading to the magnetohydrodynamic relaxation of a large-scale magnetic field. The globally stored magnetic energy is then released both as accelerated particles and in the mass motions

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 43–71.

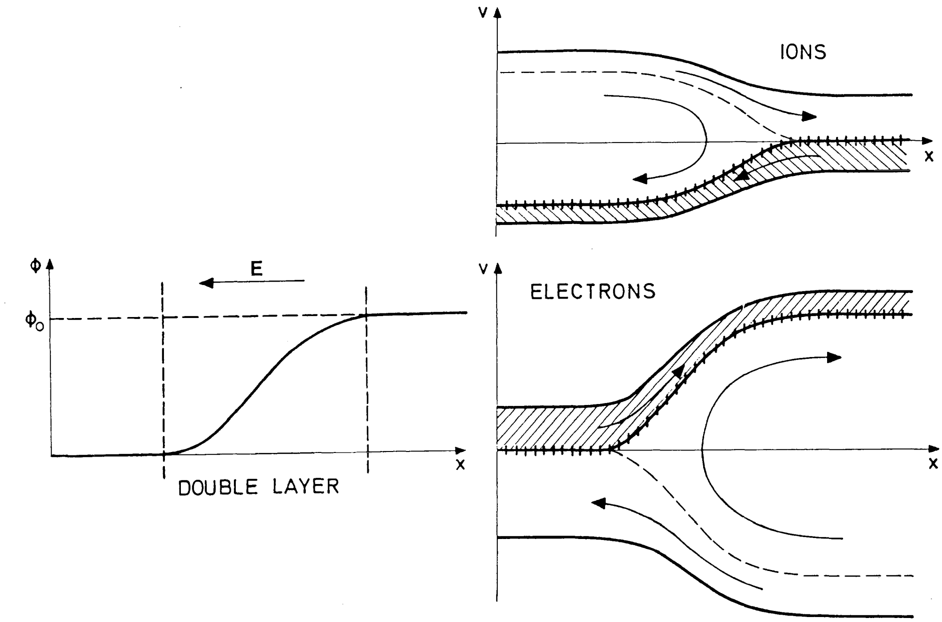


Fig. 1. The potential distribution $\varphi(x)$, ion and electron phase space distributions across a double layer with potential φ_0 and internal electric field E. As indicated by the arrows and the separatrixes ('dashed' and 'barred') there are phase space regions of transmitted and reflected particles for both species. It is physically important to distinguish between particles originating from different sources (at $x \to \pm\infty$) as indicated by the hatched and clear areas in phase space: at the boundaries ('barred' contours) there can be a sharp jump in phase space density, in principle a discontinuity, but in reality smoothed to some extent by fluctuations and diffusion processes. Outside the narrow transition region indicated by the vertical bars the phase space densities on either side have no physical relation, being related to particles from widely separated sources.

set up by the untwisting motions of the magnetised plasma (Raadu, 1984). The power supplied to accelerated particles is simply the product of the current crossing the double layer and the potential across it. The formation of a DL in a current carrying loop in the solar corona has been proposed as a mechanism for energy release in a solar flare (Jacobsen and Carlqvist, 1964; Alfvén and Carlqvist, 1967; Hasan and ter Haar, 1978). Carlqvist (1979) has also argued that the associated motions in the coronal loop can explain the occurrence of solar surges. For suitable conditions the released magnetic energy can go predominantly to the surge mass ejection rather than particle acceleration at the DL. Hénoux (1986) argues that multiple DLs can explain the high rate of energy release observed in solar flares.

In the magnetosphere the first indications of parallel electric fields leading to electrostatic acceleration of particles were provided by electron measurements in a rocket experiment (McIlwain, 1960). Further evidence was provided by the discovery of the inverted-V structures and associated electric field patterns (Frank and Ackerson, 1971; Gurnett and Frank, 1973). There is evidently a close relationship between parallel electric fields and the aurora (e.g., Mozer *et al.*, 1980, 1985; Chiu *et al.*, 1983, Fälthammar, 1983). DLs may well be responsible for sustaining the necessary parallel electric field. There is some observational evidence for multiple weak ion-acoustic DLs (e.g., Temerin *et al.*, 1982; Kellogg *et al.*, 1984). The theory of the acceleration of auroral particles by DLs has been treated in a number of papers (e.g., Block, 1978; Kan *et al.*, 1979; Goertz, 1979). Wagner *et al.* (1980) have made a two-dimensional computer simulation of an ionospheric DL (see also Singh *et al.*, 1987). Applications have also been made to Jupiter's magnetosphere (Shawhan, 1976; Smith and Goertz, 1978) and to accretion on to neutron stars (Williams *et al.*, 1986).

In the laboratory the existence of DLs has been for many years a well-established fact. Recent theoretical and experimental work has been reported at three symposia on DLs (Michelsen and Rasmussen, 1982; Schrittwieser and Eder, 1984; Williams and Moorehead, 1986). See also recent reviews by Sato (1984), Hershkowitz (1985), and Schamel (1986). Clear evidence for the formation of DLs has been shown in numerical simulations, and stationary solutions have been shown to exist theoretically for a variety of choices of the velocity distributions of the involved particle species.

Near the boundaries of a DL self-consistency between the electric field and charges requires that the net charges on the low- and high-potential sides are negative and positive, respectively (for a monotonic DL). These requirements take the form of a generalisation of the Bohm criterion for wall sheaths. For a strong DL the incoming free particles must then have velocities exceeding their thermal velocities (Block, 1972a). In particular on the low-potential side of the double layer this implies that the electron drift velocity should be sufficiently high for electrostatic two-stream instabilities to occur. Similarly ion-acoustic instabilities are expected on the high-potential side. Alfvén and Carlqvist (1967) argued that a current-driven instability (i.e., the two-stream instability in the long wavelength regime (Carlqvist, 1973)) can produce localised extreme reductions in plasma density leading to the formation of a DL in an initially homogeneous plasma. Such a localised evacuation should occur for all temperature ratios and, in one particular case, was followed well into the nonlinear regime (Raadu and Carlqvist, 1981).

Smith and Goertz (1978) have argued that the ponderomotive force of unstable waves produced by current-driven instabilities can lead to DL formation. DeGroot *et al.* (1977) ran a simulation with hot electrons drifting at about their thermal velocity through cool ions. Many weak DLs were formed. The electrons accelerated by the potential jumps were thermalised by electron-electron two-stream instabilities in the intervening field-free regions. Sato and Okuda's (1981) simulations also showed the formation of weak (ion-acoustic) DLs as a result of current-driven instabilities for drift velocities below the electron thermal velocity. Smith and Priest (1972) argued that anomalous resistivity originating from plasma turbulence rather than DL formation should result from current-driven instabilities. At the time it was not appreciated that the evacuation mechanism of Alfvén and Carlqvist (1967) was essentially identical to the two-stream instability in the long wavelength, non-dispersive, regime.

It should be stressed that very few attempts have been made to investigate the stability of the DL itself taking into account the spatial structure. A stationary DL may be regarded as a particular example of a nonlinear wave as for example described by the method of Bernstein *et al.* (1957) for deriving stationary solutions of the Vlasov–Poisson equations (the BGK method). This method was for example used by Knorr and Goertz (1974). In principle, therefore, the stability analysis of nonlinear plasma waves described by Infeld (1972) should be possible to extend to treat DLs. A general study of the stability of stationary solutions of the Vlasov–Poisson equations has been presented by Lewis and Symon (1979). Wahlberg (1979) demonstrated that linearly unstable trapped Langmuir modes may exist within a DL. This leads to a stability requirement that the DL must have a thickness smaller than a critical value. A related study has recently been made by Teichmann (1987). Schamel's (1983) analysis for a very weak DL shows linear stability with respect to longitudinal one-dimensional perturbations, but instability for transverse perturbations.

Furthermore, except for Torvén's (1981) treatment of the dynamics of a weak DL by modelling the evolution with a modified Korteweg–deVries equation (see also Raadu and Chanteur, 1986; Chanteur and Raadu, 1987), very little work has been done to describe the dynamics of DLs. The movement of strong DLs has been treated by applying the stationary description in the frame of reference moving with the DL. The motion is interpreted as a consequence of current imbalance. Moving DLs have, for example, been discussed by Bergeron and Wright (1978), Singh and Schunk (1982), Iizuka *et al.* (1982, 1985), and by Raadu and Rasmussen (1984). Carlqvist (1984) has described a number of mechanisms that can influence the geometry of double layers and in particular lead to expansion. The properties of the surrounding plasma and the effects of the whole circuit containing the DL are here of great significance.

It has become clear that the requirement that DLs be strictly time-independent structures is an overrestriction on the theory (cf. Smith, 1985; Borovsky, 1986). Rather, there is a strong need for theoretical investigations of DL stability and dynamics. A completely stable DL has not yet been observed experimentally. DLs are always subject to fluctuations in their profile, and the current through the system usually also shows some oscillations which, however, become relatively small for extremely strong DLs (Torvén, 1982; Sato *et al.*, 1983). Experiments with an inductive external circuit have been shown to lead to the transient formation of strong DLs maintained by inductive voltage pulses accompanying rapid current decreases (Torvén *et al.*, 1985). Even in computer simulations, where the boundary conditions can be fully controlled to meet the requirements for a theoretically stationary DL, the DLs produced are not at all stationary. Recurring formation and disruptions of a DL accompanied by a series of other phenomena are usually observed (cf. Smith, 1982a, b, 1985; Borovsky, 1984). This indicates on the one hand that stationary DL solutions are most likely unstable and on the other hand that 'real plasmas' can adjust 'self-consistently' to limit these instabilities making the existence of almost stationary structures possible.

In the present paper we consider the existence conditions for a DL and in particular the generalised Bohm criteria. For strong DLs these criteria have implications for the stability of the adjacent uniform plasma regions. We take this as a further indication that DLs should be regarded as dynamical time-dependent structures, so that, for example, the Bohm criteria probably only apply in some average sense. These considerations lead us on to the presentation of a simplified model of a moving dynamical DL, which we hope may lead to some new interpretations of earlier work (e.g., numerical simulations) and indicate the need for further work on DL-dynamics and stability. This paper is not intended to be a thorough review of DLs, and although the reference list is extensive it is far from complete.

2. Stationary Double Layers

In principle the structure of steady-state plane parallel DLs may be found by a straightforward integration of the Poisson equations and the time-independent Vlasov equations for the particle species. In general a first integral may then be obtained in the form

$$\tfrac{1}{2}\varepsilon_0 E^2 + V(\phi) = 0\,, \tag{2.1}$$

where E and ϕ are the electric field and potential. This is possible since stationary solutions of the Vlasov equation are given by arbitrary functions of the particle total energy. The function $V(\phi)$, the classical or Sagdeev potential, is to within an additive constant equal to minus the total particle pressure, defined using the second moments of the distribution functions (cf. Andrews and Allen, 1971; Torvén, 1980). Thus the first integral of the equations is completely equivalent to the condition of stress balance within the DL

$$-\tfrac{1}{2}\varepsilon_0 E^2 + \sum_{i,\,e} \int m_{i,\,e} v^2 f_{i,\,e}(v)\,\mathrm{d}v = \text{const.}\,; \tag{2.2}$$

where by assumption any magnetic field is constant within the DL so that the magnetic stress term arising from the Maxwell stress tensor may be dropped. Here $m_{i,\,e}$ and $f_{i,\,e}$ denote, respectively, the mass and distribution function of ions and electrons. The form of the classical potential $V(\phi)$ is, therefore, determined by the choice of the particle distributions. A second integration gives the spacial structure of the DL. This is the explicit procedure followed by Andrews and Allen (1971), and by Schamel and Bujarbarua (1983). Schamel (1982) argues that it is more natural to specify all the particle distributions and derive the DL structure than to predetermine the potential profile and then to artificially adjust one of the particle distributions to obtain a solution, i.e., BGK method (cf. Knorr and Goertz, 1974). However, to correctly represent the physical situation, care should be taken to define the various distributions at the potentials of their sources.

The properties of the DL solutions depend significantly on the potential ϕ_0 across the DL. In order to classify DLs we introduce a typical effective temperature T which represents the mean energy of the reflected particles on either side of the DL. Then for a strong DL $e\phi_0 \gg T$ and the dynamical pressures of the accelerated particle beams dominate. Such a DL is to a good approximation described by Langmuir's (1929) theory (Levine and Crawford, 1980). For weak DLs $e\phi_0 \simeq T$ and a variety of solutions are possible. Schamel (1983) distinguishes two types of weak DLs (slow electron or ion-acoustic DLs). Strong DLs are included among what he calls beam-type DLs. We note the existence of other types of weak DLs such as those described by Torvén (1981) for cold ions and non-Maxwellian electrons which move at close to an ion-acoustic velocity. Also there are DLs which we may refer to as intermediate DLs. These are characterised by the presence of particle components produced by secondary physical processes. An example is given by the work of Andersson and Sørensen (1983) who include the effects of ionisation. Also Kan and Lee (1980) include an electron population built up by scattering processes in their magnetospheric DL model. We note that what have been referred to as asymmetric electron and ion holes also have some of the characteristics of DL's and for example can sustain a net potential difference (cf. Schamel, 1986).

3. Double-Layer Existence Conditions

We now consider existence conditions for DL solutions in general terms. First since $E^2 \geq 0$, a necessary condition for the existence of DL solutions is that the classical potential $V(\phi) \leq 0$ within the DL. This condition is particularly sensitive at the edges of the DL ($\phi = 0$ and $\phi = \phi_0$, where ϕ_0 is the DL potential). There the normal matching conditions between the external plasma and the DL require $V(0) = V(\phi_0) = 0$ (zero electric field) and $V'(0) = V'(\phi_0) = 0$ (no net charge). A power series expansion then gives $V''(0) \leq 0$ and $V''(\phi_0) \leq 0$ so that V has local maxima at $\phi = 0, \phi_0$. These are the generalised forms of the Bohm criteria which ensure appropriate net charges near the edges of the DL (cf. Introduction) and may also be written $\{Zen_i'(\phi) - en_e'(\phi)\} < 0$ for $\phi = 0$ and $\phi = \phi_0$, where Z is the ion-charge number and $n_{i,e}$ the number density of ions and electrons, respectively. Since $E = 0$ at the edges of the double layer it follows that $V(0) = V(\phi_0) = 0$. This is simply the condition that the total particle pressures on either side of the DL are equal. The DL is in pressure equilibrium. For a strong DL the dynamical pressures of the emerging accelerated particle beams dominate and pressure equilibrium requires that $|j_e/j_i| \approx \sqrt{m_i/m_e}$, the Langmuir condition on j_e and j_i, the electron and ion-current densities through the DL. Thus the pressure balance condition $(V(0) = V(\phi_0))$ may be regarded as the appropriate generalisation of the Langmuir condition.

Figure 2 illustrates the general required form of the classical potential $V(\phi)$, incorporating the existence conditions for DL solutions. The minimum value corresponds to the maximum electric field E inside the DL.

Many of the considerations that apply to DLs have counterparts in the theory of wall sheaths, where it is also of interest to find how the Bohm criterion can be satisfied and the stability properties of the sheath for different relative potentials between the plasma and the wall. A brief discussion of wall sheaths may be found in Appendix A.

Block (1972a) discussed the application of the Bohm sheath criterion to a double layer. Considering the predominant particle species he derives simple conditions for the

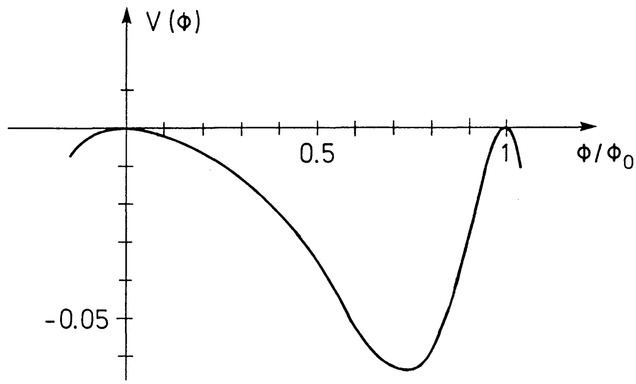


Fig. 2. Example of the classical potential $V(\varphi)$. The function plotted is: $V(\varphi) = -5\varphi^2(1 - \varphi^2)(0.05 + \varphi^4)$, φ is normalized to the DL potential φ_{DL}.

ion and electron injection velocities v_{i0}, v_{e0}: namely,

$$m_i v_{i0}^2 \geq T_e\,, \qquad m_e v_{e0}^2 \geq T_i\,, \tag{3.1}$$

where T_e, T_i are the trapped thermal electron and ion temperatures on the appropriate sides of the double layer. He argues that there must be a certain pre-acceleration of beam particles to meet these conditions before the final acceleration in the DL.

Levine and Crawford (1980) have presented a series of fluid models for double layers using four particle species and successively introducing thermal effects. They note the condition that the classical potential (in their notation $P(\varphi)$ or $\Pi(\varphi)$) must be negative and in particular that the second derivative must be negative at the edges of the DL, i.e., generalised Bohm criteria.

The generalised Bohm criteria may be expressed directly in terms of the distribution functions $f_{i,\,e}(v)$ of the ions and electrons. For time-independent conditions the Vlasov equations are satisfied if the distribution functions are

$$f_i = F_i\,[\mathrm{sgn}(v)\,(v^2 + 2q_i\phi/m_i)^{1/2}]\,, \tag{3.2}$$

$$f_e = F_e\,[\mathrm{sgn}(v)\,(v^2 + 2q_e(\phi - \phi_0)/m_e)^{1/2}]\,, \tag{3.3}$$

where the functions $F_{i,\,e}$ are defined by the ion distribution at $\phi = 0$ and by the electron distribution at $\phi = \phi_0$. The ion and electron particle charges are $q_i = Ze$ and $q_e = -e$, where for hydrogen and singly charged ions $Z = 1$. From the ion and electron densities given by

$$n_{i,\,e} = \int_{-\infty}^{+\infty} f_{i,\,e}\,\mathrm{d}v\,, \tag{3.4}$$

it follows that if $\rho(\phi)$ is the net charge, then

$$\rho'(\phi) = \sum_{i,\,e} \frac{q_{i,\,e}^2}{m_{i,\,e}} \int_{-\infty}^{\infty} \frac{1}{v}\,\frac{\partial f_{i,\,e}}{\partial v}\,\mathrm{d}v\,, \tag{3.5}$$

using the functional dependence of $f_{i,e}$ on ϕ and v defined above through $F_{i,e}$. Now using distribution functions normalised to unit density the Bohm criteria (defined by $\rho'(\phi) < 0$ as $\phi \to 0$ and ϕ_0) are given by

$$\beta \equiv \sum_{i,e} \omega_{pi,e}^2 \int_{-\infty}^{\infty} \frac{1}{v} \frac{\partial f_{i,e}}{\partial v} \, dv < 0 \quad \text{as} \quad \varphi \to 0^+, \ \phi_0^- . \tag{3.6}$$

This is closely related to the sheath criterion found by Allen (1976) (see Appendix A). It is of interest to compare this expression with the dispersion relation for electrostatic waves:

$$K^2 = \sum_{i,e} \omega_{pi,e}^2 \left[P \int_{-\infty}^{\infty} \frac{1}{v - a} \frac{\partial f_{i,e}}{\partial v} \, dv + i\pi f'_{i,e}(a) \right], \tag{3.7}$$

where $a = \omega/K$ is the phase velocity of waves with frequency ω and wave number K. From the expression for the distribution functions inside the DL given above we notice that $f'_{i,e}(0) = 0$. It then follows that for zero-phase velocity ($a = 0$) $K^2 = \beta$. The Bohm criteria are then seen to be equivalent to requiring that the stationary electrostatic wave is spatially evanescent (K is imaginary). For a Maxwellian plasma $\beta = K^2$ $(a = 0) = -\lambda_D^{-2}$ where λ_D is the Debye shielding length. In general we may then say that the Bohm criteria are requirements that the plasma just inside the edges of the DL acts to shield charge.

In general we may allow for discontinuities in the distribution functions at the separatrix between trapped and free particles. The separatrix is given by $v = \pm v_i \equiv \pm(2q_i(\phi_0 - \phi)/m_i)^{1/2}$ for the ions and $v = \pm v_e \equiv \pm(-2q_e\phi/m_e)^{1/2}$ for the electrons. Ions and electrons with velocities v with $|v| < v_i$ and $|v| < v_e$, respectively, are reflected somewhere inside the DL. The distribution functions are, therefore, symmetric inside these ranges and in particular $f_i(v_i - \varepsilon) = f_i(-v_i + \varepsilon)$ and similarly for electrons (where $\varepsilon \to 0^+$). In contrast, for free particles we expect no relationship between the distributions for $v > v_{i,e}$ and $v < -v_{i,e}$, since the particles come in general from widely separated, unrelated sources to the left and right of the DL, respectively. To derive the Bohm criteria including the possible discontinuities in the distributions, the integration range in the definition of $n_{i,e}$ should be divided up before taking derivatives. The final result is then found to be unchanged if we treat $f'_{i,e}(v)$ as a δ-function at the discontinuities so that, for example,

$$f'_i(v) = \{f_i(v_i^+) - f_i(v_i^-)\}\, \delta(v - v_i) \quad \text{for} \quad v \approx v_i .$$

The Bohm criteria are then given by the expression already given above.

If we now consider the Bohm criterion as $\phi \to 0$ the critical velocity for electron trapping tends to zero, i.e., $v_e \to 0$. Depending on the discontinuities in $f_e(v)$ at $v = -v_e$ (Δf_1) and at $v = +v_e$ (Δf_2) the electron contribution to β can be divergent. The

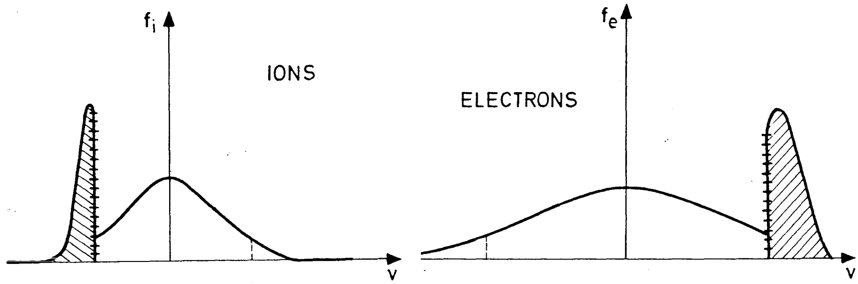


Fig. 3. The ion and electron distribution functions $f_i(v)$ and $f_e(v)$ within a double layer. These should be compared with the phase space plots of Figure 1. The hatched regions here correspond with those of Figure 1. Note the near discontinuities marked by bars which correspond to the barred contours of the previous figure. They mark the boundaries between particles from distinct sources at $\pm\infty$.

δ-function contribution then dominates so that

$$\beta \approx \omega_{pe}^2 \left[\frac{\Delta f_1}{-v_e} + \frac{\Delta f_2}{v_e} \right] = \frac{\omega_{pe}^2(\Delta f_2 - \Delta f_1)}{\sqrt{2e\phi/m_e}} \,, \tag{3.8}$$

and if $\Delta f_2 \neq \Delta f_1$, $\beta \to \pm\infty$ as $\phi \to 0$. Then the Bohm condition $\beta < 0$ reduces to

$$0 > \Delta f_2 - \Delta f_1 = f_e(0^+) + f_e(0^-) - 2f_{et}(0) \,, \tag{3.9}$$

where $f_{et}(0)$ is the value of the trapped electron distribution just inside the DL. This means that either $f_{et}(0)$ must exceed a critical value to satisfy the Bohm criterion (this can lead to a critical trapped particle density requirement as for Kan and Lee's (1980) model) or that $f_e(v) \to 0$ as $v \to \pm 0$. If $\Delta f_2 = \Delta f_1$ or if there are no discontinuities the singularity is removed and we must use the full expression for β to find the Bohm criterion. Similar considerations apply at $\phi = \phi_0$ for the ions.

4. Bohm Criteria and Instabilities

The existence of a double layer in current carrying plasma imposes requirements on the incoming particle distributions (e.g., generalised Bohm criteria) and also implies the presence of emerging particle beams accelerated by the DL. It is, therefore, necessary to ask if the uniform plasma outside the DL is stable. A distinction should be drawn between two possible sources of instability. Those instabilities, which may be driven by emerging beams, propagate away from the DL and affect it only indirectly by modifying the incoming particle distributions, even if the beams themselves are destroyed. In this respect such instabilities are relatively harmless to the continued existence of the DL. On the other hand, instabilities generated by incoming particle beams might be dangerous as they may substantially modify or even disperse the beams so that they no longer fulfill the matching conditions, e.g., generalised Bohm criteria, at the edges of the DL.

Having derived the Bohm criteria in terms of the distribution functions it is now clear from a comparison with the dispersion relation that there is no simple relationship with the criteria for instability (Penrose, 1960) (see Equations (3.6) and (3.7)). We now proceed instead to consider several one-dimensional DL models which have been presented in the literature and comment on the stability and Bohm criteria.

4.1. Strong Double Layers

In view of the special role that discontinuities can play we note that a fluid component in a DL model (e.g., Block, 1972b) is in one dimension completely equivalent to a water bag distribution (Bertrand and Feix, 1968).

Block (1972b) presented a DL model with fluid and Maxwellian distributed particle components. This model explicitly shows the matching of a strong DL onto the surrounding plasma. Collisions are included, but for the present purposes we discuss the collisionless limit only. In this limit the fluid components are completely equivalent to water bag distributions (Bertrand and Feix, 1968) in one dimension and for the ratio of specific heats $\gamma = 3$. On the low-potential side from the general analysis presented here we then find the Bohm criterion

$$0 > \beta = -\frac{n_{i0}e^2}{\varepsilon_0 T_{i0}} + \frac{\omega_{pe}^2}{v_e^2 - c_e^2} - \frac{n_{e0}e^2\, e^{-e\varphi_0/T_{i0}}}{\varepsilon_0 T_{e0}} + \frac{\omega_{pib}^2}{v_{ib}^2 - c_{ib}^2} . \tag{4.1}$$

The notation is as follows: n_{i0}, n_{e0} are the densities of the Maxwellian distribution of ions and electrons on the low ($\varphi = 0$) and high ($\varphi = \varphi_0$) potential sides of the DL, respectively, with corresponding temperatures T_{i0}, T_{e0}; v_e, c_e, and ω_{pe} are the velocity, sound velocity, and plasma frequency of the incoming fluid electrons on the low ($\varphi = 0$) potential side; v_{ib}, c_{ib}, and ω_{pib} are the corresponding quantities for the fluid ions which have been accelerated by the DL potential φ_0 and emerge as a beam on the low-potential side. For a strong DL the last two terms in the Bohm criterion are negligible as well as the corresponding densities. The Bohm condition may be reduced and rewritten as

$$v_e > \sqrt{\frac{T_{i0}}{m_e} + c_e^2} , \tag{4.2}$$

in agreement with Block's (1972b) equation in the collisionless limit. An analogous result holds on the high-potential side.

The emerging low-density ion beam on the low-potential side of the DL should drive instabilities with phase velocities close to the beam velocity, propagating away from the DL. More dangerous instabilities may be driven by the incoming fluid electrons moving through the Maxellian distribution of ions, i.e., just those components which dominate the Bohm criterion.

Schamel and Bujarbarua (1983) have derived several DL models by self-consistently solving the stationary Vlasov equations together with Poisson's equation. They prescribe trapped electron and ion-distribution functions in terms of Maxellians and free electron- and ion-distribution functions in terms of drifting Maxwellians. They further assume that the distributions are continuous everywhere, i.e., without a jump at the separatrix between reflected and free particles (cf. the discussion in Section 3). This assumption provides an extra restriction on the solutions.

In the cases where the DL potential is much greater than the trapped particle temperatures (measured in electron volts) their distributions in the ambient plasma are essentially fully Maxwellian. The drift velocities of the injected distributions are much larger than their thermal velocities because they have to match onto the small tails of the distributions with high-energy particles that have managed to overcome the DL potential.

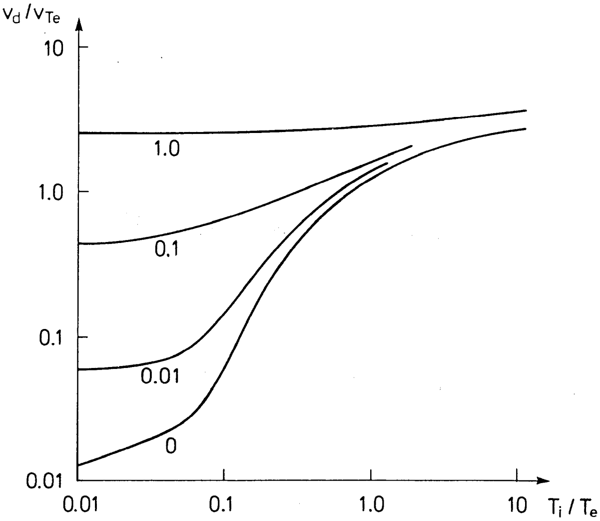


Fig. 4. Stability diagram for electrostatic current-driven instabilities taken from Stringer (1964), $v_{Te} \equiv \sqrt{T_e/m_e}$. The contours show the growth rate of the fastest growing mode. The number labelling the curves is $\gamma_{\max}/\Omega_i$. Ω_i is the ion-plasma frequency. Note that a correction has here been made for a small error in the original scale on the vertical axis.

Hence, looking for instabilities driven by the injected distributions the situation is that of two relative drifting Maxwellian distributions (neglecting the low-density emerging beams). The results of Stringer (1964) for electrostatic instabilities in a uniform plasma may then be applied to the ambient plasma regions. In Figure 4 the instability diagram of Stringer (1964) is plotted. Clearly the models of Schamel and Bujarbarua (1983) imply instability in the ambient plasma at least for strong DLs. Such instabilities are also to be expected for models involving distributions with discontinuities at the separatrix between reflected and free particles (cf. Section 3). The instabilities certainly appear when the incoming electrons have truncated distributions, with finite truncation velocity. Note that Jentsch (1976) has generalized the instability analysis of Stringer (1964) by allowing the relatively drifting distribution functions to have a continuous range of forms including Maxwellian, waterbag, and resonance functions as limiting cases.

4.2. Intermediate double layers

The auroral DL proposed by Kan and Lee (1980) is an example of an intermediate DL. Their DL is 'strong' if we consider only the primary particles, i.e., ions and electrons from the plasma sheet and ionosphere. The DL potential energy $e\varphi_0$ is not limited and may be much greater than the typical energies of the primary particles. Kan and Lee note that with no extra particle component the Bohm conditions derived by Block (1972b) would imply that on the low-potential side the incoming electrons would have to be pre-accelerated by some mechanism up to energies corresponding to the temperatures of the hot plasma sheet ions, i.e., a few keV. (This is already the energy of auroral electrons so that the DL acceleration is secondary.) They argue for an extra electron population trapped between the DL and the magnetic mirror. These electrons consist of degraded beam electrons and secondary electrons produced by collisions with neutral particles. This 'extra' population has energies up to $e\varphi_0$. They may, therefore, penetrate the DL and if there are enough they ensure that there is a net negative charge at the low-potential side of the DL. This ensures that the Bohm criterion is satisfied. The DL is 'weak' with respect to these secondary electrons. Kan and Lee's DL model successfully explains how low-energy incoming electrons can be accelerated up to auroral energies. Since they assume a water bag distribution extending down to zero-velocity for incoming electrons there are no dangerous instabilities propagating toward the DL. The emerging ion beam may be expected to drive instabilities, but these propagate upwards away from the DL. However, depending on the effective external circuit global instabilities are possible and, for example, a negative resistance instability of the Kan and Lee model was shown by Silevitch (1981) and Raadu and Silevitch (1983).

In some experiments the energetic electrons emerging from a DL significantly ionise the neutral gas on the high-potential side. This allows the production of DLs with only one plasma source located on the low-potential side of the DL: energetic electrons produce the plasma on the other side (Torvén and Andersson, 1979; Torvén and Lindberg, 1980; Andersson, 1981). Andrews and Allen (1971) discussed the structure of a DL pre-sheath with ionisation. We note that if their integrals are rewritten (using integration by parts) to avoid singularities it follows from the analysis that the ionisation rate has to go smoothly down to zero at the edge of the DL. In principle we might expect the Bohm condition to apply just at the inner edge of such a sheath at the DL. There is then a better chance that the pre-sheath and external plasma region may be stable. This may also apply if collisions (classical or anomalous) in the external plasma are taken into account (cf. Borovsky, 1986). Comparisons may here be made with work on plasma wall sheaths as referred to in Appendix A.

Andersson and Sørensen (1983) have obtained time-independent solutions of DLs where the ions on the high-potential side are created by ionisation due to the accelerated electrons. They use their results to explain experiments and to suggest the possibility of DLs with ionisation effects in the ionosphere. The density and velocity distribution of the free ions are found assuming that they are created by ionisation at zero velocity and then freely accelerated in the electric field. The authors remark that the predominance of low-energy ions in the distribution would seem to violate the Bohm criterion; but since ionisation is included here the densities are not functions of the potential only and the collisionless Bohm criteria are not applicable. However, on the low-potential side the incoming electrons must satisfy the normal Bohm criterion and, as for strong DLs, we can expect instabilities propagating towards the DL. In the relevant experiments (Torvén and Lindberg, 1980), however, there is a potential minimum on the low-potential side, beyond which the potential is gently increasing. Then even if the Bohm criterion is applicable at the DL it need not imply instability in the adjacent plasma where the particle distributions are functions of the changing potential.

The strong potential DLs discussed by Knorr and Goertz (1974) are only 'strong' in the sense applied to Kan and Lee's (1980) models. Following the method of Bernstein *et al.* (1957), Knorr and Goertz derive a necessary distribution of trapped electrons for a given potential profile and free-particle distribution functions. There are no trapped ions in their models. The trapped electrons have distributions extending up to the DL potential energy. The solutions have the property that the uniform plasmas on either side of the DL are Penrose stable, i.e., there are no electrostatic current-driven instabilities. The electrons have everywhere a mean velocity equal to zero. The generalised Bohm criterion on the low-potential side is met by the trapped electrons which penetrate through the DL as in the model of Kan and Lee. In fact the strong DL of Knorr and Goertz have some of the properties characterising Schamel's (1983) slow electron- and ion-acoustic DLs (SEADL and SIADL). On the high-potential side there is a minimum in the electron distribution at zero-velocity as for a SEADL. However, the density is largest on the high-potential side (since there are no trapped ions) as is the case for a SIADL. Knorr and Goertz (1974) also obtain weak DL solutions which may be made asymptotically Penrose stable. These are constructed so that the electron distribution has a single maximum. Chosing an ion frame of reference on the low-potential side, the velocity of these weak DLs must be less than the ion-acoustic velocity (cf. their Equation (43), for the case of zero-temperature free ions).

It is apparent that the extra particle distributions assumed in intermediate DLs help to avoid strict Bohm criteria. This has the consequence that the external plasmas may be less unstable or even completely stable. A departure from strictly one-dimensional models would lead to an even stronger modification of DL properties. In experiments radial losses are probably important and may even be affected by instabilities resulting in a complex interdependence between the time-averaged structure and time-dependent fluctuations. Borovsky and Joyce (1983) have simulated DL structures in two dimensions including the effects of magnetic fields. They also present a simple one-dimensional oblique DL model including a uniform magnetic field. In particular they derive explicit Bohm criteria which clearly show a dependence on the angle between the DL electric field and the magnetic field.

5. Weak Double Layers

Weak double layers are characterized by having a potential jump $e\varphi_0/T \leq 1$ implying that the densities and thereby the function $V(\varphi)$ may be expressed as a power series in

φ. This of course simplifies the analysis considerably, and it has for instance been possible to derive evolution equations for such weak DLs (cf. Torvén, 1981). While the strong DL-solutions require the existence of four distinct particle types (see Section 4.1) for the weak DL-solutions three types of particles suffice. We shall also see that these solutions in general are stable with respect to beam instabilities, i.e., the generalized Bohm criterion do not requir inflow of particles with velocities exceeding the critical velocities for instabilities.

Recently the existence of weak DLs has been shown for many models corresponding to different combinations of the involved particle species. These models may roughly be divided into two main types.

(i) Models based on a fluid description, where typically the plasma consists of two groups of electrons with distinctly different temperatures and cold ions (cf. Torvén, 1981; Goswami and Bujarbarua, 1985; Raadu and Chanteur, 1986; Bharuthram and Shukla, 1986a; Chanteur and Raadu, 1987).

(ii) Models based on a kinetic description of at least one of the involved particle species (cf. Perkins and Sun, 1981; Kim, 1983; Schamel, 1983; Torvén, 1986).

In addition we should also mention the so-called ion-acoustic double layers (Sato and Okuda, 1981; Hasegawa and Sato, 1982; Schamel, 1982) which essentially should be classified as weak DLs in the sense that the net potential jump across them: $e\varphi_0/T_e \leq 1$. However, by employing the terminology of Schamel (1983, 1986) these structures should rather be referred to as asymmetric ion holes. Similarly asymmetric electron holes also exist (Schamel, 1982, 1986) (see also Pécseli (1984) for a review of electron and ion holes). We shall not give a detailed description of these structures here, but just mention that, e.g., the asymmetric ion holes are observed in simulations of current carrying plasmas (cf. Sato and Okuda, 1981; Borovsky, 1984). Furthermore, they are observed experimentally (Chan *et al.*, 1984; Sekar and Saxena, 1985) and above the aurora (cf. Temerin *et al.*, 1982; Mozer *et al.*, 1985), where a series of them are believed to be responsible for the acceleration of the auroral electrons.

To illustrate how the weak DL-solutions are constructed and how the different restrictions come about we consider a model of type (i). The DL-solution, stationary in a frame moving with the velocity M, is obtained from the first integral of the Poisson equation (see Section 2) in the normalized form

$$\frac{1}{2}\left(\frac{\partial\varphi}{\partial x}\right)^2 + V(\varphi) = 0\,, \tag{5.1}$$

where $-\mathrm{d}V/\mathrm{d}\varphi \equiv (n_e(\varphi) - n_i(\varphi))$. Here φ is normalized with T/e, $n_{e,i}$ with the undisturbed densities n_0, x with $(\varepsilon_0 T/e^2 n_0)^{1/2}$. For small amplitude DLs both n_e and n_i may be expanded in powers of φ resulting in

$$-V(\varphi) = A_1\varphi^2 + A_2\varphi^3 + A_2\varphi^3 + A_3\varphi^4 + \cdots\,, \tag{5.2}$$

where we have taken $V(0) = 0$. The coefficients A_i are functions of the velocity M and the parameters of the involved particle species.

By reference to Section 3 it is easily seen that we have to retain up to the fourth power in the expansion of $V(\varphi)$, i.e., up to the third power in the expansions of n_e, n_i. Furthermore, a DL-solution to (5.2) requires that

$$A_3 > 0 \tag{5.3}$$

(the generalized Bohm criterion) and that the polynomial $\varphi^2 + (A_2/A_3)\varphi + (A_1/A_3) = 0$ has a double root – i.e.,

$$A_2^2 = 4A_1A_3 . \tag{5.4}$$

The DL potential is then given by

$$\varphi_0 = -\frac{A_2}{2A_3} ; \tag{5.5}$$

and for consistency we must require that

$$\left|\frac{A_2}{2A_3}\right| \ll 1 . \tag{5.6}$$

For the case where the plasma consists of cold ions and a hot and cold electron component it is possible to fulfill the conditions (5.3), (5.4), and (5.6) (cf. Torvén, 1981; Goswami and Bujarbarua, 1985), and the DL-solution is expressed as

$$\varphi = \frac{1}{2}\varphi_0\left\{1 - \tanh\left[\left(\frac{A_3}{2}\right)^{1/2}\varphi_0(x - Mt)\right]\right\}, \tag{5.7}$$

where φ_0 as given by (5.5) is positive (negative) when A_2 is negative (positive). Thus both compressional ($\varphi_0 > 0$) and rarefactive ($\varphi_0 < 0$) DLs may exist. It is evident that these DL-solutions do not require streaming particle populations which could lead to instabilities. However, certain restrictions must be fulfilled for their existence, and in particular (5.6), essentially ensuring a sufficiently small DL-potential, must be satisfied for the expansion (5.2) to be consistent. The relation (5.6) is satisfied for certain ratios between the temperatures and the densities of the hot and cold electrons, respectively. Note that Bharuthram and Shukla (1986b) recently constructed DL-solutions with finite potential (i.e., without employing the expansion (5.2) for the above discussed model).

When the condition (5.6) applies, it has been shown that the dynamical evolution of small amplitude disturbances, is governed by a modified KdV equation (cf. Torvén, 1981; Raadu and Chanteur, 1986; Chanteur and Raadu, 1987, and references given therein)

$$\partial_t\varphi + \alpha\varphi\,\partial_x\varphi + \beta\varphi^2\,\partial_x\varphi + \tfrac{1}{2}\,\partial_{xxx}\varphi = 0 , \tag{5.8}$$

where α corresponds to $-3A_2$ and β to $-6A_3$ in (5.2), which may easily be seen by looking for the stationary solution of (5.8). Note that (5.8) may be transformed into the canonical form of the mKdV by using the linear transformation

$$\varphi = \left(\frac{3}{|\beta|}\right)^{1/2} \psi - \frac{1}{2}\frac{\alpha}{\beta}, \tag{5.9}$$

together with a change of the frame of reference and rescaling the time by dividing by two – i.e.,

$$\partial_t \psi + 6 \operatorname{sign}(\beta) \psi^2 \partial_x \psi + \partial_{xxx} \psi = 0 . \tag{5.10}$$

The transformation (5.9) simply corresponds to a shift of the reference level of the potential. Only when $\beta < 0$ will the mKdV equation have DL-type solutions (cf. Equation ((5.3)), and for consistency $\alpha = o(\beta)$ for the same parameters, corresponding to (5.6). In that connection we should mention that for a plasma consisting of positive and negative ions and isothermal electrons an mKdV equation has also been derived (Watanabe, 1984). However, in the validity region of this mKdV equation (i.e., $\alpha \simeq o(\beta))\beta > 0$ and DL-type solutions do not exist.

Chanteur and Raadu (1987) have presented a detailed study of the mKdV equation (5.10) with $\beta < 0$ solving it analytically, by means of the inverse-scattering transformation, as well as numerically. The general results are briefly summarized as follows: suppose that initially $\psi(-\infty) = \psi_0^-$ and $\psi(+\infty) = \psi_0^+$. Then a DL-like structure forms if and only if $\psi_0^- \psi_0^+ < 0$ and has a potential jump $2\,|\psi_0^+|$. In that case, for $|\psi_0^-| < |\psi_0^+|$ a train of solitons moving away from the DL appears and for $|\psi_0^-| > |\psi_0^+|$ a non-dispersive ramp forms behind the DL. Individual solitons may also appear depending on the form of the initial profile (*ibid*). The DL-like solutions to the mKdV equation are stable since they are obtained as the asymptotical evolution of a given initial condition by means of the inverse scattering method. The stability of these weak DL-solutions with respect to perturbations transverse to their propagation is, however, still an open question. Such a stability study could be based on a two-dimensional generalization of the mKdV equation as, e.g., derived by Bharuthram and Shukla (1986a) for a magnetized plasma.

For the models of type (ii) the plasma usually consists of free ions and electrons and reflected ions and/or reflected electrons. Schamel (1983) has characterized these weak DLs as either slow electron acoustic DL (SEADL), when reflected electrons are involved, i.e., a tuning-fork-like phase space configuration of the electrons, or slow ion-acoustic DL (SIADL) when reflected ions are involved. These DLs may even exist without any net current (Perkins and Sun, 1981). The SEADL at rest in the ion-frame exist for $T_i/T_e > 3.5$ while the SIADL correspondingly require $T_e/T_i > 3.5$. We point out that the SEADL (SIADL) essentially are 'half' electron- (ion-) holes. To obtain the solutions for the small amplitude DLs both the electron and ion densities may be expanded in powers of φ resulting in

$$-V(\varphi) = A_1 \varphi^{3/2} + A_2 \varphi^2 + A_3 \varphi^{5/2} + A_4 \varphi^3 + \cdots . \tag{5.11}$$

Due to the presence of reflected particles governed by a kinetic description the expansion includes half-powers in φ. The coefficients are complicated functions of the parameters of the involved particle species. A_1 in particular is related to the jump in the distribution function at the separatrix between free and reflected particles (cf. Torvén, 1986). Schamel (1983) assumes that the distributions are continuous everywhere and argues that this is a necessary requisite for a physically acceptable theory. In this case $A_1 = 0$ and $A_4\varphi^3$ must be retained to have DL-solutions. Torvén (1986) argues that if the free and reflected particles have different origins (and the plasma is collisionless) then a discontinuity at the separatrix is most probable (cf. Figure 3). We should stress that even if the distribution has a discontinuity the 'physical' quantities, the moment of the distribution, may certainly be continuous as they of course should be. Referring to the discussion of model (i) it is easily seen that a DL-solution requires that

$$A_3 > 0 \quad (A_1 \neq 0); \qquad A_4 > 0 \quad (A_1 = 0)\,. \tag{5.12}$$

Furthermore,

$$A_2^2 = 4A_1A_3 \quad (A_1 \neq 0)\,; \qquad A_3^2 = 4A_2A_4 \quad (A_1 = 0) \tag{5.13}$$

and

$$\varphi_0 = \left(\frac{A_2}{2A_3}\right)^2 \quad (A_1 \neq 0)\,; \qquad \varphi_0 = \left(\frac{A_3}{2A_4}\right)^2 \quad (A_1 = 0)\,. \tag{5.14}$$

The DL-solution then becomes

$$\begin{aligned} \varphi &= \varphi_0 \tanh^4(\gamma(x - Mt)) \quad (A_1 \neq 0)\,, \\ \varphi &= \tfrac{1}{4}\varphi_0(1 + \tanh(\kappa(x - Mt)^2) \quad (A_1 = 0)\,, \end{aligned} \tag{5.15}$$

where the coefficient γ may be found in Torvén (1986) and $\kappa = (\varphi_0/8)^{1/2}$. Note that the first solution in (5.15) is defined in the half plane $-\infty < x - Mt \leq 0$. Torvén (1986) derived a modified KdV equation with half power nonlinearities describing the DL-evolution (for $A_1 \neq 0$).

Concerning the stability of these types of DLs, Schamel (1983) argues that both the SEADL and SIADL are stable with respect to beaming instabilities but unstable with respect to transverse perturbations.

In closing this section we mention that a weak DL of the SIADL type has been observed experimentally (Chan *et al.*, 1986), while there is evidence for the existence of the SEADL in numerical simulations (Schamel, 1984).

6. Dynamics of Strong Double Layers

Double layers are inherently dynamical structures as has been strongly emphasized in the previous sections. It has also become evident that due to their strongly nonlinear nature it is hard to conceive of an analytical model for the dynamical evolution of strong and intermediate DLs. For weak DLs, on the other hand, tractable evolution equations have been derived taking the form of modified KdV-equations as discussed in Section 5.

Certainly a lot of insight into the dynamics of strong DLs has been gained from numerous computer simulations (e.g., Smith, 1982b; Singh and Schunk, 1982; Borovsky, 1984; Iizuka and Tanaca, 1985; Yamamoto, 1985; and references given in these papers) and from experimental investigations (e.g., Sato, 1984; Hershkowitz, 1985). Furthermore, some attempts to describe the dynamics of strong DLs have been made by investigating the consequences of varying some of the parameters of the stationary DL-solutions. For example, the movement of a DL was treated by applying the stationary description in the moving frame of reference (e.g., Singh, 1979; Singh and Shunk, 1982; Raadu and Rasmussen, 1984), and Carlqvist (1984) considered different effects, that would give rise to an expansion or contraction of a DL. In this section we discuss some of these attempts and relate the results to the dynamical features observed in simulations and experiments. We emphasize that the dynamics of the DL is strongly dependent on the properties of the surrounding plasma and the whole circuit, in the general sense, that contains the DL. We intend to demonstrate this through specific examples in what follows.

In particular we consider the movement of a strong DL. Moving double layers have been observed in most of the numerical simulations performed (see the above references) and also in many experiments (e.g., Lutsenko *et al.*, 1976; Leung *et al.*, 1980; Sato, 1984; Iizuka *et al.*, 1985). In most cases the DL moves towards the high-potential side, 'rarefactive movement'; but also movement towards the low-potential side, 'compressional movement' has been observed (Coakley and Hershkowitz, 1981). In the following we describe a simple model for the rarefactive movement of a strong DL. We shall restrict ourselves to a one-dimensional model and assume that the DL moves as a 'rigid body' in accordance with the observations in experiments and simulations. Our argument is based on the pressure balance across the DL, which for strong DLs (non-relativistic) reduces to the Langmuir condition (see Section 3)

$$|j_e/j_i| = \sqrt{m_i/m_e} \equiv \mu \,. \tag{6.1}$$

The DL is allowed to move in the laboratory frame with the velocity v_{DL}, positive in the direction of the electron beam velocity, i.e., from the low- to the high-potential side of the DL (see Figure 5). Since the DL is assumed to conserve its shape during the propagation (6.1) must be satisfied in the moving frame, resulting in

$$v_{\mathrm{DL}} = \frac{n_{e1}v_{e1} - \mu n_{i0}v_{i0}}{n_{e1} + \mu n_{i0}} \,, \tag{6.2}$$

where 0 labels the parameters on the high-potential side and 1 those on the low-potential side, $v_{e1,\,i0}$ are the averaged drift velocities of electrons and ions entering the DL, respectively, while $n_{e1,\,i0}$ are the respective densities. When Equation (6.1) is satisfied for the currents injected at the boundaries of the DL region the DL is at rest, but when a current imbalance exists the DL may compensate by moving. Similar arguments were presented by several authors (see, e.g., Singh, 1979; Singh and Schunk, 1982).

Imagine now a stationary DL formed in the interior of a plasma produced by two

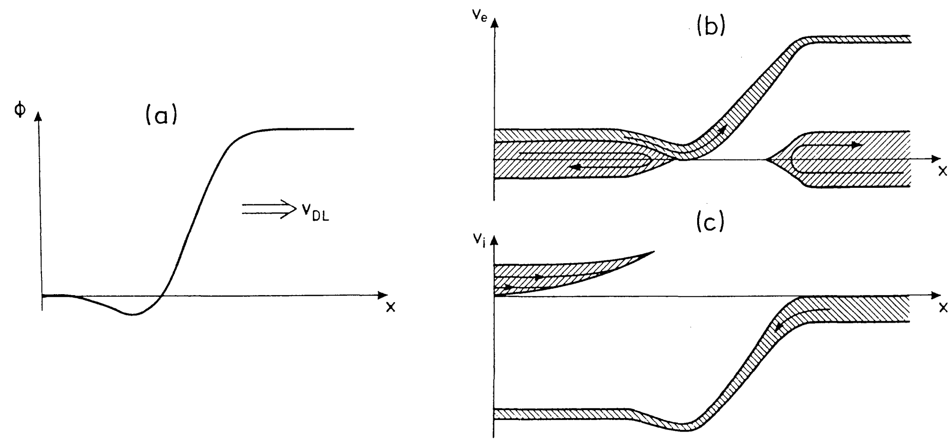


Fig. 5. Snapshot of a moving DL with negative potential dip followed by an expanding plasma. (a) potential; (b) electron phase space. v_{DL} is much smaller than the thermal velocities of both streaming and reflected electrons; (c) ion phase space, showing the expanding ions coming in from the left. It should be noted that everything is in the laboratory frame of reference.

sources at different potentials at each end of a linear device. Then the plasmas emitted by the two sources satisfy: $n_{e1}v_{e1} = \mu n_{i0}v_{i0}$. At, say $t = 0$, the density on the high-potential side is decreased, but the rest is left unchanged – i.e.,

$$n'_{i0} = n_{i0} - \delta n_0 \quad \text{and} \quad n'_{et} = n_{et} - \delta n_0$$

and we still have the plasma condition on the high-potential side $n'_{i0} = n'_{et} + n'_{eb}$, where t labels the trapped (reflected) population and b the beam particles accelerated through the DL. A very small adjustment of the densities on the low-potential side is also necessary to maintain plasma conditions since the density of the beam ions n_{ib} here is decreased by $\delta n_0 v_{i0}/(2e\varphi_{\mathrm{DL}}/m_i)^{1/2}$, which for the very strong DL is negligible. Neglecting the influence of any transient phenomena we see from (6.2) that the DL starts moving with the velocity

$$v_{\mathrm{DL}} = \frac{\mu \delta n_0 v_{i0}}{n_{e1} + \mu n'_{i0}}\ ; \tag{6.3}$$

i.e., the DL moves towards the high-potential side. How does this movement affect the other particle species? The trapped electron population n'_{et} will be compressed, but the high-potential source may easily regulate that, and the effects on the trapped electrons will be negligible. The trapped ions, on the other hand, may lag behind the moving DL if v_{DL} is larger than the expansion velocity of the plasma produced by the low-potential source. This velocity will be around the ion-acoustic velocity c_s. Thus the moving DL will be followed by an expanding plasma. For $c_s < v_{\mathrm{DL}}$ an electron surplus will appear between the expansion front and the DL resulting in the formation of a negative potential dip (see also the more detailed discussion in Appendix B) at the low-potential tail of the

DL as sketched in Figure 5 and clearly observed in simulations and experiments (e.g., Singh and Schunk, 1982; Iizuka and Tanaca, 1985; Iizuka *et al.*, 1985). This dip in turn reflects some of the electrons leading to a decrease in the electron current. Denoting the decrease in the electron beam density and mean velocity at the bottom of the dip by δn_{e1} and δv_{e1}, respectively, we obtain

$$v_{\mathrm{DL}} = \frac{-\delta n_{e1} v_{e1} - \delta v_{e1} n_{e1} + \delta n_{e1} \delta v_{e1} + \mu \delta n_0 v_{i0}}{n'_{e1} + \mu n'_{i0}} . \tag{6.4}$$

If the distribution of the beam electrons are known then δv_e and δn_e can be calculated as functions of the depth of the dip. For our present argument it is sufficient to note that v_{DL} decreases. The potential dip cannot exist self-consistently in a stationary frame because the adjacent DL is very strong ($e\varphi_{\mathrm{DL}} \gg T$), and the ions will cancel the negative space charge connected with the dip (see Appendix B). Thus, when v_{DL} decreases, the dip potential decreases too and the electron current increases. From these considerations we can argue that the DL velocity cannot be much faster than the expansion velocity of the low-potential plasma. The potential dip regulates the electron current through the system and is an effective current limiting mechanism. However, the dip is formed because of the movement of the DL which is initiated by the current imbalance. By using (6.2) we see that the (electron-) current through the DL is given by

$$j_e \equiv e n_{e1} v_{e1} = e\mu n_{i0}(v_{i0} + v_{\mathrm{DL}}) + e v_{\mathrm{DL}} n_{e1} ; \tag{6.5}$$

thus for $n_{i0} > \mu^{-1} n_{e1}$ the current is mainly determined by the density of the high-potential plasma and the DL velocity. We remark that the Bohm condition, in the generalized sense, ensuring that there is a net charge of the correct (self-consistent) sign on entering the DL (cf. Section 3), is easily verified on either side of the moving DL. This further supports the assumptions of 'rigid' DL movement.

This simple model describes the essential features of the rarefactive moving DLs observed in experiments (e.g., Leung *et al.*, 1980; Sato, 1984; Iizuka *et al.*, 1985) and simulations (e.g., Singh and Schunk, 1982; Iizuka and Tanaca, 1985; or Yamamoto, 1985). In particular it emphasizes the role of the negative potential dip which always seems to form on the low-potential tail of a moving DL, and which effectively limits the current through the system. Related investigations of the role of an associated potential minimum for current disruptions caused by a DL have been reported by Torvén *et al.* (1985) and Carpenter and Torvén (1986).

In closing the discussion of the rarefactive moving DL we point out that the negative potential dip forming at the low-potential side of the DL is of quite different origin than the solitary potential structure of the so-called ion-acoustic DL (Sato and Okuda, 1981; Hasegawa and Sato, 1982; Chanteur *et al.*, 1983) (cf. Section 5) although they act in a similar way with respect to reflecting electrons and limiting the current. The dip following the moving strong DL can only exist in the moving frame and appears an ambipolar potential in front of an expanding plasma region; it does not contain trapped ions and is usually relatively broad (many Debye lengths λ_{De}). The potential dip connected with the ion-acoustic double layer is narrow (a few λ_{De}) and may exist in a stationary frame where ion-trapping results in the formation of an ion hole (cf. Pécseli, 1984). The scenario for the formation of an ion-acoustic DL preceeded by electron reflection from a developed ion hole is clearly revealed in numerical simulations (cf. Sato and Okuda, 1981; Borovsky, 1984). Actually even strong transient DLs have been observed to develop due to electron reflection presumably from an ion hole developed in a Buneman unstable plasma (Belova *et al.*, 1980; Galeev *et al.*, 1981). Note that while the simulations resulting in ion-acoustic DLs employed periodic boundary condition, Belova *et al.* (1980) used 'open' boundary conditions.

It is obvious that the model described for the moving DL assumes a particular form of the boundary condition imposed on the plasma and essentially also of the external circuit. That is, we assumed that the inflow of plasma could be regulated by some external agency. Bergeron and Wright (1978), on the other hand, argue that a pressure difference across the DL will not lead to any acceleration of the DL in a collisionless plasma, but rather be compensated by regulating the electron and/or ion beam through the DL. Thus they implicitly assume that their boundary conditions and external circuit allows for such a 'self-consistent' regulation of the plasma influx. They further argue that an acceleration of the DL will be the result of other processes such as, for example, collisions or beam plasma interactions. With the above discussion in mind we add that external constraints fixing, for example, the current through the system would also imply a movement.

The compressional movement of a DL as a 'rigid' body is much more rarely observed than the above described rarefactive movement. Coakely and Hershkowitz (1981) explain their observations in terms of the theory for electrostatic shocks taking into account the effects of reflected electrons on the high-potential side of the shock (Montgomery and Joyce, 1969). In that connection we also refer to the paper by Hershkowitz (1981) where shocks are distinguished from DLs by the direction of the flow of free ions: for shocks they flow toward the high-potential side in the shock frame, while for DLs they flow toward the low-potential side. Attempting to apply a model similar to the one used for the rarefactive movement, to the compressional movement of a strong DL, it may become difficult for the ions to fulfill the Bohm condition on the high-potential side of the DL. If this condition is not fulfilled we expect the rigidity of the DL to break down and the DL profile to relax. For the electrostatic shock Herskowitz (1981) argues that the fulfillment of the Bohm criterion on the high-potential side restricts the Mach number M (shock velocity normalized with the ion-acoustic velocity)

$$M^2 \geq 1 + 2e\varphi_{\text{shock}}/T_e\,. \tag{6.6}$$

This implies that the ions on the high-potential side must be injected with a velocity

$$v_{i0} \geq c_s + (1 + 2e\varphi_{\text{shock}}/T_e)^{1/2}\, c_s\,.$$

Extrapolation to strong DLs obviously leads to unrealistically high-ion velocities.

Finally, we briefly discuss the work of Carlqvist (1984), who suggests different effects leading to relaxation of the DL-profile. Again we emphasize that these results are the outcome of particular assumptions on the whole circuit including the DL-structure. For instance Carlqvist considers a situation in which a strong DL is suddenly formed in a current-carrying plasma. It is assumed that the current is kept constant by the external circuit. By requiring the fulfillment of the Langmuir condition (6.1) it is argued that the high-potential edge of the DL moves into the high-potential plasma while the low-potential edge is at rest. As a result the DL expands with velocity

$$v_{\exp} \simeq |v_{e1}|/\mu \,, \tag{6.7}$$

where v_{e1} is the velocity of the electrons entering on the low-potential side. Carlqvist also proposes an expansion mechanism due to collisions between the emitted beams and the surrounding plasma. Furthermore, electromagnetic forces are shown to provide a contraction of the DL under particular conditions.

7. Discussion

DLs and turbulent plasma conditions are alternative outcomes for the same initial unstable conditions. The generalised Bohm criteria and the condition for electrostatic instability (Penrose criterion) are often identical; but there is a class of asymptotically stable DLs (e.g., the models of Knorr and Goertz, 1974). However, strong DLs appear to always require unstable external plasmas and are, therefore, always time-dependent in some way.

The arguments present in this paper show that many strictly time-independent DL solutions seem impossible to realise because the incoming particle beams lead to two-stream instabilities when they are required to satisfy the Bohm criteria. Smith and Priest (1972) presented similar arguments and concluded that anomalous resistivity should occur rather than DL formation once the critical current drift velocity was exceeded. We agree with their diagnosis but with hindsight, knowing that DLs may, in fact, be produced both in the laboratory and in computer simulations, we instead argue for the DL as a dynamical, intrinsically time-dependent phenomenon. The stability and dynamics of DLs should take into account the effects of realistic boundary conditions and appropriate effective external circuits.

Future developments of the theory of DL dynamics will probably include more realistic treatments of extra effects of the type we have discussed here, e.g., ionisation. More three-dimensional models are needed including a magnetic field. Global effects of the external medium should be considered. Electromagnetic aspects have received little attention apart from discussions concerning the mhd-relaxation of the external plasma. Goertz (1983) has argued for electron acceleration by kinetic Alfvén waves in the magnetosphere (see also Goertz and Boswell, 1979). There may perhaps be some direct interrelationship between kinetic Alfvén waves and DLs through the electric field component parallel to the magnetic field.

Finally, we note that so far there has been no complete stability analysis for a strong DL taking into account its spatial structure. One reasonable approach would be to use a computer simulation starting from carefully prepared initial conditions determined by a stationary DL solution.

Acknowledgements

The authors would like to thank their colleagues at Risø and Stockholm for many useful discussions on topics taken up in the present work. This work was in part supported by the Swedish Natural Science Research Council.

Appendix A. The Bohm-Criterion for the Wall-Sheath

The criterion for the formation of a stationary sheath connecting a biased (or floating) conducting wall to a plasma was considered in the classic papers of Langmuir (1929). It was later precisely stated in the work by Bohm (1949), from where the often quoted name 'Bohm-criterion' originates. The criterion demands that the accelerated mono-energetic particles at the sheath edge stream towards the wall, having an energy greater than the thermal energy of the reflected particles.

We shall here briefly illustrate how a Bohm-criterion is derived. We also discuss the stability of the wall-sheath and discuss the dynamical evolution when the Bohm-criterion is not satisfied. Consider a sheath connecting a biased, perfectly absorbing wall with a one-dimensional, semi-infinite plasma. The plasma is unmagnetized, collisionless, and without internal ionization. Using the Poisson equation the Bohm-criterion for both an electron and ion sheath takes the form

$$[\mathrm{d}(n_e(\varphi) - n_i(\varphi))/\mathrm{d}\varphi]_{\varphi = \varphi(x_0)} > 0 \,, \tag{A.1}$$

where x_0 is the location of the sheath edge. Outside the sheath region in the uniform plasma with density n_0 we take $\varphi = 0$, thus $\varphi(x_0) = 0$. The condition (A.1) ensures the right sign of the charge when the sheath is entered. Let us consider the ion sheath (i.e., for a negatively biased wall). The electrons are assumed to be in Boltzmann equilibrium with a temperature T_e, thus

$$n_e(\varphi) = n_0 \exp(e\varphi/T_e) \,. \tag{A.2}$$

The ion density is determined from the distribution function, which is a stationary solution of the Vlasov equation

$$f_i(u, \varphi) = n_0 F\left(u^2 + \frac{2e\varphi}{m_i}\right) H\left(u - \sqrt{u_{tr}^2 - \frac{2e\varphi}{m_i}}\right), \tag{A.3}$$

where H is the unit step-function. Assuming that the truncation velocity $u_{tr} > 0$ we obtain

$$n_i'(\varphi = 0) = \frac{n_0 e}{m_i}\left[\frac{F(u_{tr}^2)}{u_{tr}} + \int_{u_{tr}}^{\infty} 2F'(u^2)\,\mathrm{d}u\right]. \tag{A.4}$$

By means of partial integrations we get

$$n_i'(\varphi = 0) = \frac{n_0 e}{m_i} \int_{u_{tr}}^{\infty} \frac{F(u^2)}{u^2}\, \mathrm{d}u \equiv \frac{n_0 e}{m_i} \left\langle \frac{1}{u^2} \right\rangle . \tag{A.5}$$

On using (A.5) and (A.2) the Bohm criterion becomes

$$\langle u^{-2} \rangle > T_e/m_i . \tag{A.6}$$

The Bohm-criterion in this form was first derived by Harrison and Thompson (1959). Allen (1976) argues that the criterion (A.6) should have an equality sign replacing the inequality sign. He further provides an alternative physical meaning of the criterion, showing that the criterion implies that an ion-acoustic wave is unable to travel from the wall into the plasma, because its velocity in the laboratory frame is reduced to zero. With the inequality sign in (A.6) an ion-acoustic wave is certainly also hindered in propagating into the plasma.

We emphasize that the criterion (A.6) not only demands a certain averaged drift velocity of the ions but also requires that the distribution function goes rapidly to zero for $u \to 0$. This is certainly the case for truncated distributions as considered by, e.g., Hu and Ziering (1966). For collisional plasmas with internal ionization the analysis gets much more involved than the simple arguments presented above (cf. Riemann, 1981). Almost all analyses of sheath formation have been concerned with ion sheaths although the Bohm-criterion for the formation of a stationary electron sheath is in principle straightforward to obtain by interchanging the roles of electrons and ions in the analysis of ion sheaths.

Now we turn to the question of the stability of the wall sheaths. It is generally found that an electron-rich sheath (i.e., for a positive wall-bias) is unstable at least in a collisionless plasma. This is clearly seen in, for example, the simulations of Skøelv and Rasmussen (1984) and also in the simulations of the plasma diode (Gray *et al.*, 1984; Iizuka and Tanaca, 1985). The ion-rich sheath (i.e., for negative wall-bias) on the other hand, is found to be generally stable, in particular when the Bohm-criterion is fulfilled (Skøelv and Rasmussen). Even when this criterion is not satisfied an almost stationary sheath, a few Debye-lengths wide, is formed. But a quasi-neutral region connecting the sheath to the undisturbed plasma develops. This region expands continuously and the velocity of its leading edge is the ion-acoustic velocity (cf. Cipolla and Silevitch, 1981; Skøelv and Rasmussen, 1984). This quasi-neutral region is usually referred to as the pre-sheath. It accelerates the ions to a velocity at the sheath edge which is just sufficient to fulfill the Bohm criterion. In an ideal semi-infinite collisionless plasma the pre-sheath will never become stationary but continue to expand. However, in a bounded plasma the pre-sheath will reach an extension determined by the geometry (cf. Emmert *et al.*, 1980; Meassick *et al.*, 1985). In a collisional plasma the extension of a stationary pre-sheath will be determined by the mean-free path.

Appendix B

First we show that a one-dimensional time-independent strong DL preceded by a potential dip on the low-potential side (together forming a 'triple layer') cannot be described by using the Vlasov–Poisson system. Consider the first integral of this system, i.e., the stress balance condition. On the low-potential side within the plasma outside the DL $\varphi = 0$ and there is no electric field. Between the potential minimum and the DL there is a point where again $\varphi = 0$. ($V(\varphi)$ is now a multi-valued function.) Here the electron pressure must be lower since some electrons are reflected by the minimum and do not reach the point $\varphi = 0$ (those that do have the same distribution as they had in the plasma) and for a strong DL there are practically no trapped electrons that can penetrate the DL from the high-potential side. If there is an electric field it contributes a negative pressure increasing the stress inbalance further. Hence, the assumption of a potential minimum is untenable since for a strong DL stress balance cannot be maintained. Intuitively the stress inbalance might be expected in a dynamical development to steepen and close up the potential minimum region.

To model a strong DL with a potential minimum it is, therefore, necessary to relax the idealised assumptions. For example, radial ion losses from the plasma in a DL experiment are often significant; the situation is not strictly one-dimensional. Fujita *et al.* (1984) argue that ions are scattered inside the DL in the region beyond the potential minimum. This, they model by reducing the corresponding ion density by a factor depending exponentially on the increasing potential on the high-potential side of the minimum. This introduces an asymmetry in the ion density about the minimum and self-consistent solutions are then found for a DL with a potential minimum. Also other processes than those implied by the Vlasov–Poisson system must apply for broad potential minima with quasi-neutrality (Carpenter and Torvén, 1986).

Here and in a previous paper (Raadu and Rasmussen, 1984) we have discussed a simple non-steady model to explain the essential features of moving DLs. During one phase of the propagation we argue that ions lag behind the DL. Since the ion-dynamical time-scale greatly exceeds that for the electrons, on intermediate time-scales the ion density may be treated as fixed and the electrons may be assumed to have steady approximately time-independent motion. As a simple model we will now consider the situation indicated in Figure 6. The ion density n_i is constant except in a region around the potential minimum where it is approximated to zero. The injected electrons entering from the low-potential sidde have a water-bag (WB) distribution and density $n_{e\mathrm{WB}}$. On the high-potential side there are reflected thermal electrons with low temperatures ($T \ll e\Delta\varphi$) and density n_{eT}.

Suppose that the injected electron water-bag distribution function $f_{e\mathrm{WB}}(v)$ has a value f_0 for $v_{10} < v < v_{20}$ and is zero outside this range. This is equivalent to a fluid (in one dimension) with velocity $(v_{10} + v_{20})/2$ and sound speed $(v_{20} - v_{10})/2$ (cf. Bertrand and Feix, 1968). The density is then given by

$$n_{e\mathrm{WB}} = f_0[(v_{20}^2 + 2e\varphi/m)^{1/2} \mp (v_{10}^2 + 2e\varphi/m)^{1/2}], \tag{B.1}$$

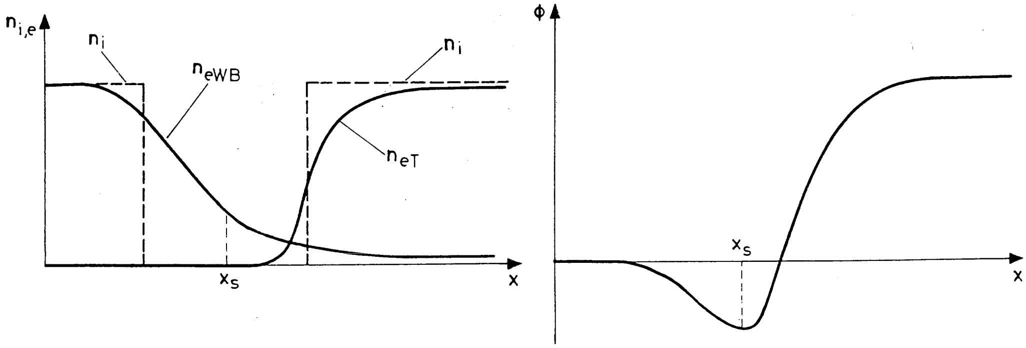


Fig. 6. For a time shorter than that for ion dynamics a local region of low-ion density ($n_i \approx 0$) is crossed by electrons with a water-bag distribution ($n_{e\mathrm{WB}}$) moving in a self-consistent potential $\varphi(x)$. The potential minimum at x_s acts as a Laval nozzle for the fluid electrons. The electrons reach their sound speed at x_s beyond which their density drops as they are accelerated by the increasing potential. Thermal electrons (n_{eT}) reflected by the potential maintain charge neutrality to the right.

where the water-bag extends over $v_1 < v < v_2$ and the velocities v_2, v_1 are given, respectively, by the square-root terms in Equation (B1). For $v_1 < 0$ the fluid motion is subsonic and here we assume $v_{10} < 0$ so that the injected electrons are subsonic. This corresponds to a plus sign in Equation (B1). As the potential φ decreases the density also decreases until $v_1 = 0$ at the sonic point $x = x_s$ in Figure 6. If this also corresponds to the minimum potential φ_s the potential increases beyond this point and choosing $v_1 > 0$, i.e., supersonic motion, minus sign in Equation (B1), the density will continue to fall. The potential minimum acts as a Laval nozzle in aerodynamics allowing a transition from sub- to supersonic flow. The electrons density is, therefore, asymmetric about the potential minimum as is also the total pressure

$$P_{e\mathrm{WB}} = \tfrac{1}{3}mf_0[(v_{20}^2 + 2e\varphi/m)^{3/2} \mp (v_{10}^2 + 2e\varphi/m)^{3/2}]\,. \tag{B.2}$$

However, the total pressure has a minimum value at the sonic point where the electric field $E = 0$. Now, neglecting the other particle populations, the steady-state dynamics of the water-bag electrons in the self-consistent electric field is given by Equation (2.2) as for a steady-state DL. On the high-potential side of the minimum when the motion is supersonic (minus sign in Equation (B.2)) the pressure is less than on the subsonic side at the same potential. From Equation (2.2) the electric field E is correspondingly larger.

The model so far is a particular case of a 'virtual cathode'. To produce a DL potential profile as shown in Figure 6 it is only necessary to add the contributions of the other particle species. Thus the fixed ions on the low-potential side provide a region of net positive charge ($n_i > n_{e\mathrm{WB}}$) cancelling the electric field and providing a match to plasma conditions ($E = 0, n_i = n_e$). On the high-potential side the electron density is essentially given by the thermal electrons $n_e \simeq n_{eT}$ since the water-bag electrons emerge as a low-density beam. These thermal electrons then provide a Debye shielding of the DL electric field and produce the required match to plasma conditions over a few Debye lengths λ_{De}.

References

Alfvén, H.: 1981, *Cosmic Plasma,* D. Reidel Publ. Co., Dordrecht, Holland.

Alfvén, H.: 1986, *IEEE Trans. Plasma Sci.* **PS-14**, 779.

Alfvén, H. and Carlqvist, P.: 1967, *Solar Phys.* **1**, 220.

Allen, J. E.: 1976, *J. Phys. D: Appl. Phys.* **9**, 2331.

Andersson, D.: 1981, *J. Phys. D: Appl. Phys.* **14**, 1403.

Andersson, D. and Sørensen, J.: 1983, *J. Phys. D: Appl. Phys.* **16**, 601.

Andrews, J. G. and Allen, J. E.: 1971, *Proc. Roy. Soc. London* **A320**, 459.

Belova, N. G., Galeev, A. A., Sagdeev, R. Z., and Sigov, Yu. S.: 1980, *Pis'ma Zh. Eksperim. Teor. Fiz.* **31**, 551 (*J. Eksperim. Theor. Phys. Letters* **31**, 518).

Bergeron, K. D. and Wright, T. P.: 1978, *Phys. Fluids* **21**, 1578.

Bernstein, I. B., Greene, J. M., and Kruskal, M. D.: 1957, *Phys. Rev.* **108**, 546.

Bertrand, P. and Feix, M. R.: 1968, *Phys. Letters* **28A**, 68.

Bharuthram, R. and Shukla, P. K.: 1986a, *Phys. Scripta* **34**, 732.

Bharuthram, R. and Shukla, P. K.: 1986b, *Phys. Fluids* **29**, 3214.

Block, L. P.: 1972a, 'Acceleration of Auroral Particles by Electric Double Layers', in B. M. McCormac (ed.), *Earth's Magnetospheric Processes,* D. Reidel Publ. Co., Dordrecht, Holland, p. 258.

Block, L. P.: 1972b, *Cosmic Electrodyn.* **3**, 349.

Block, L. P.: 1978, *Astrophys. Space Sci.* **55**, 59.

Bohm, D.: 1949, 'Minimum Ion Kinetic Energy for a Stable Sheath', in. A. Guthrie and R. K. Wakerling (eds.), *The Characteristics of Electrical Discharges in Magnetic Fields*, McGraw-Hill, New York, p. 77.

Borovsky, J. E.: 1984, 'A Review of Double-Layer Simulations', in R. Schrittewieser and G. Eder (eds.), *Proc. Second Symp. on Plasma Double Layers and Related Topics*, Innsbruck, July 5–6, p. 33.

Borovsky, J. E.: 1986, *Astrophys. J.* **306**, 451.

Borovsky, J. E. and Joyce, G.: 1983, *J. Plasma Phys.* **29**, 45.

Carlqvist, P.: 1973, *Double Layers and two-Stream Instability in Solar Flares,* TRITA-EPP-73–05, Dept. of Plasma Phys., The Royal Inst. of Technol., Stockholm, Sweden.

Carlqvist, P.: 1979, *Solar Phys.* **63**, 353.

Carlqvist, P.: 1984, 'On the Expansion of Electrostatic Double Layers', in R. Schrittwieser and G. Eder (eds.), *Proc. Second Symp. on Plasma Double Layers and Related Topics*, Innsbruck, July 5–6, p. 340.

Carpenter, R. T. and Torvén, S.: 1986, *The Current-Voltage Characteristic and Potential Oscillations of a Double Layer in a Triple Plasma Device*, TRITA-EPP-86–05, Dept. Plasma Physics, The Royal Institute of Technology, Stockholm, Sweden.

Chan, C., Cho, M.-H., Hershkowitz, N., and Intrator, T.: 1984, *Phys. Rev. Letters* **52**, 1782.

Chan, C., Cho, M.-H., Hershkowitz, N., and Intrator, T.: 1986, *Phys. Rev. Letters* **57**, 3050.

Chanteur, G., Adam, J. C., Pellat, R., and Volokhitin, A. S.: 1983, *Phys. Fluids* **26**, 1584.

Chanteur, G. and Raadu, M. A.: 1987, *Phys. Fluids* **30**, 2708.

Chiu, Y. T., Cornwall, J. M., Fennel, J. F., Gorney, D. J., and Mizera, P. F.: 1983, *Space Sci. Rev.* **35**, 211.

Cipolla, J. W. and Silevitch, M. B.: 1981, *J. Plasma Phys.* **25**, 373.

Coakley, P. G. and Herhkowitz, N.: 1981, *Phys. Letters* **83A**, 131.

DeGroot, J. S., Barnes, C., Walstead, A. E., and Buneman, O.: 1977, *Phys. Rev. Letters* **38**, 1283.

Emmert, G. A., Wieland, R. M., Mense, A. T., and Davidson, J. N.: 1980, *Phys. Fluids* **23**, 803.

Fälthammar, C.-G.: 1983, *ESA J.* **7**, 385.

Frank, L. A. and Ackerson, K. L.: 1971, *J. Geophys. Res.* **76**, 3612.

Fujita, H., Yagura, S., and Matsuo, K.: 1984, *Phys. Fluids* **27**, 2948.

Galeev, A. A., Sagdeev, R. Z., Shapiro, V. D., and Shevchenko, V. I.: 1981, *Zh. Eksperim. Teor. Fiz.* **81**, 572 (*Soviet Phys. JETP* **54**, 306).

Goertz, C. K.: 1979, *Rev. Geophys. Space Phys.* **17**, 418.

Goertz, C. K.: 1983, 'Nonstationary Coupling Between the Magnetosphere and Ionosphere', *Sixth ESA Symp. on European Rocket and Balloon Programmes*, Interlaken, 11–15 April, ESA SP-183, p. 221.

Goertz, C. K. and Boswell, R. W.: 1979, *J. Geophys. Res.* **84**, 7239.
Goswami, K. S. and Bujarbaru, S.: 1985, *Phys. Letters* **108A**, 149.
Gray, P. L., Kuhn, S., Crystal, T. L., and Birdsall, C. K.: 1984, 'Particle Simulations of the Unstable States of a Collisionless Single-Ended Plasma Device', in R. Schrittwieser and G. Eder (eds.), *Proc. Second Symp. on Plasma Double Layers and Related Topics*, Innsbruck, July 5–6, p. 266.
Gurnett, D. A. and Frank, L. A.: 1973, *J. Geophys. Res.* **78**, 145.
Harrison, E. R. and Thompson, W. B.: 1959, *Proc. Phys. Soc. London* **74**, 145.
Hasan, S. S. and ter Haar, D.: 1978, *Astrophys. Space Sci.* **56**, 89.
Hasegawa, A. and Sato, T.: 1982, *Phys. Fluids* **25**, 632.
Hénoux, J.-C.: 1986, 'Dynamo Theories of Solar Flares', in V. E. Stepanov and V. N. Obridko (eds.), *Solar Maximum Analysis*, VNU Science Press, p. 109.
Hershkowitz, N.: 1981, *J. Geophys. Res.* **86**, 3307.
Hershkowitz, N.: 1985, *Space Sci. Rev.* **41**, 351.
Hu, P. N. and Ziering, S.: 1966, *Phys. Fluids* **9**, 2168.
Iizuka, S. and Tanaca, H.: 1985, *J. Plasma Phys.* **33**, 29.
Iizuka, S., Michelsen, P., Rasmussen, J. J., Schrittwieser, R., Hatakeyama, R., Saeki, K., and Sato, N.: 1982, *Phys. Rev. Letters* **48**, 145.
Iizuka, S., Michelsen, P., Rasmussen, J. J., Schrittwieser, R., Hatakeyama, R., Saeki, K., and Sato, N.: 1985, *J. Phys. Soc. Japan* **54**, 2516.
Infeld, E.: 1972, *J. Plasma Phys.* **8**, 105.
Jacobsen, C. and Carlqvist, P.: 1964, *Icarus* **3**, 270.
Jentsch, V.: 1976, *Phys. Fluids* **19**, 919.
Kan, J. R. and Lee, L. C.: 1980, *J. Geophys. Res.* **85**, 788.
Kan, J. R., Lee, L. C., and Akasofu, S.-I.: 1979, *J. Geophys. Res.* **84**, 4305.
Kellogg, P. J., Monson, S. J., and Whalen, B. A.: 1984, *Geophys. Res. Letters* **11**, 515.
Kim, K. Y.: 1983, *Phys. Letters* **97A**, 45.
Knorr, G. and Goertz, C. K.: 1974, *Astrophys. Space Sci.* **31**, 209.
Langmuir, I.: 1929, *Phys. Rev.* **33**, 954.
Leung, P., Wong, A. Y., and Quon, B. H.: 1980, *Phys. Fluids* **23**, 992.
Levine, J. S. and Crawford, F. W.: 1980, *J. Plasma Phys.* **23**, 223.
Lewis, H. R. and Symon, K. R.: 1979, *J. Math. Phys.* **20**, 413.
Lutsenko, E. I., Sereda, N. D., and Kontsevoi, L. M.: 1976, *Fiz. Plasmy* **2**, 72 (*Soviet J. Plasma Phys.* **2**, 39).
Meassick, S., Cho, M. H., and Hershkowitz, N.: 1985, *IEEE Trans. Plasma Sci.* **PS-13**, 115.
McIlwain, C. E.: 1960, *J. Geophys. Res.* **65**, 2727.
Michelsen, P. and Rasmussen, J. J. (eds.): 1982, *Symposium on Plasma Double Layers*, Risø National Laboratory, June 16–18, 1982 (Risø–R–472).
Montgomery, D. and Joyce, G.: 1969, *J. Plasma Phys.* **3**, 1.
Mozer, F. S., Boehm, M. H., Catell, C. A., Temerin, M., and Wygant, J. R.: 1985, *Space Sci. Rev.* **42**, 313.
Mozer, F. S., Cattel, C. A., Hudson, M. K., Lysak, R. L., Temerin, M., and Torbert, R. B.: 1980, *Space Sci. Rev.* **27**, 155.
Pécseli, H. L.: 1984, 'Electron and Ion Phase Space Vortices', in R. Schrittwieser and G. Eder (eds.), *Proc. Second Symp. on Plasma Double Layers and Related Topics*, Innsbruck, July 5–6, p. 81.
Penrose, O.: 1960, *Phys. Fluids* **3**, 258.
Perkins, F. W. and Sun, Y. C.: 1981, *Phys. Rev. Letters* **46**, 115.
Raadu, M. A.: 1984, 'Global Effects of Double Layers', in R. Schrittwieser and G. Eder (eds.), *Proc. Second Symp. on Plasma Double Layers and Related Topics*, Innsbruck, July 5–6, p. 3.
Raadu, M. A. and Carlqvist, P.: 1981, *Astrophys. Space Sci.* **74**, 189.
Raadu, M. and Chanteur, G.: 1986, *Phys. Scripta* **33**, 240.
Raadu, M. A. and Rasmussen, J. J.: 1984, 'On the Dynamics of Double Layers', in R. Schrittwieser and G. Eder (eds.), *Proc. Second Symp. on Plasma Double Layers and Related Topics*, Innsbruck, July 5–6, p. 334.
Raadu, M. A. and Silevitch, M. B.: 1983, *J. Plasma Phys.* **30**, 249.
Riemann, K.-U.: 1981, *Phys. Fluids* **24**, 2163.
Sato, N.: 1984, 'Formation of Electrostatic Potential in a Magnetized Plasma', in M. Q. Tran and R. J. Verbeek (eds.), *Proc. 1984 International Conference on Plasma Physics*, Lausanne, Switzerland, June 27–July 3, Invited Papers, Vol. II, p. 555.

Sato, T. and Okuda, H.: 1981, *J. Geophys. Res.* **86**, 3357.
Sato, N., Hatakeyama, R., Iizuka, S., Mieno, T., Saeki, K., Rasmussen, J. J., Michelsen, P., and Schrittwieser, R.: 1983, *J. Phys. Soc. Japan* **52**, 875.
Schamel, H.: 1982, 'The Theory of Double Layers', in P. Michelsen and J. J. Rasmussen (eds.), *Symposium on Plasma Double Layers*, Risø, June 16–18, Risø–R–472, p. 13.
Schamel, H.: 1983, *Z. Naturforsch.* **38a**, 1170.
Schamel, H.: 1984, 'Identifications of Electrostatic Structures in Plasma Situations', in R. Schrittwieser and G. Eder (eds.), *Proc. Second Symp. on Plasma Double Layers and Related Topics*, Innsbruck, July 5–6, p. 302.
Schamel, H.: 1986, *Phys. Reports* **140**, 161.
Schamel, H. and Bujarbarua, S.: 1983, *Phys. Fluids* **26**, 190.
Schrittwieser, R. and Eder, G. (eds.): 1984, *Proc. Second Symp. on Plasma Double Layers and Related Topics*, Innsbruck, July 5–6 (Institut für Theoretische Physik, R. Schrittwieser, Innrain 52, A-6020, Innsbruck).
Sekar, A. N. and Saxena, Y. C.: 1985, *Plasma Phys. Controlled Fusion* **27**, 181.
Shawhan, S. D.: 1976, *J. Geophys. Res.* **81**, 3373.
Silevitch, M. B.: 1981, *J. Geophys. Res. 86*, 3573.
Singh, N.: 1979, *Phys. Letters* **75A**, 69.
Singh, N. and Schunk, R. W.: 1982, *J. Geophys. Res.* **87**, 3561.
Singh, N., Thiemann, H., and Schunk, R. W.: 1987, *Planetary Space Sci.* **35**, 353.
Skøelv, Å. and Rasmussen, J. J.: 1984, 'A Numerical Investigation of Wall-Sheath Stability in a One-Dimensional Plasma', in R. Schrittwieser and G. Eder (eds.), *Proc. Second Symp. on Plasma Double Layers and Related Topics*, Innsbruck, July 5–6, p. 309.
Smith, D. F. and Priest, E. R.: 1972, *Astrophys. J.* **176**, 487.
Smith, R. A.: 1982a, *Phys. Scripta* **25**, 413.
Smith, R. A.: 1982b, *Phys. Scripta* **T2/1**, 238.
Smith, R. A.: 1985, 'On the Role of Double Layers in Astrohysical Plasmas', in M. R. Kundu and G. D. Holman (eds.), *IAU Symp.* **107**, 113.
Smith, R. A. and Goertz, C. K.: 1978, *J. Geophys. Res.* **83**, 2617.
Stringer, T. E.: 1964, *Plasma Phys.* **6**, 267.
Teichmann, J.: 1987, *Phys. Letters* **A121**, 25.
Temerin, M., Cerny, K., Lotko, W., and Mozer, F. S.: 1982, *Phys. Rev. Letters* **48**, 1175.
Torvén, S.: 1980, 'Double Layers – Theory and Experiment', *Proc. Int. Conf. Plasma Physics*, Nagoya, Japan, April 7–11, 1980, Vol. II, Fusion Research Association of Japan, p. 89.
Torvén, S.: 1981, *Phys. Rev. Letters* **47**, 1053.
Torvén, S.: 1982, *J. Phys. D: Appl. Phys.* **15**, 1943.
Torvén, S.: 1986, *Phys. Scripta* **33**, 262.
Torvén, S. and Andersson, D.: 1979, *J. Phys. D: Appl. Phys.* **12**, 717.
Torvén, S. and Lindberg, L.: 1980, *J. Phys. D: Appl. Phys.* **13**, 2285.
Torvén, S., Lindberg, L., and Carpenter, R. T.: 1985, *Plasma Phys. Controlled Fusion* **27**, 143.
Wagner, J. S., Tajima, T., Kan, J. R., Leboeuf, J. N., Akasofu, S.-I., and Dawson, J. M.: 1980, *Phys. Rev. Letters* **45**, 803.
Wahlberg, C.: 1979, *J. Plasma Phys.* **22**, 303.
Watanabe, S.: 1984, *J. Phys. Soc. Japan* **53**, 950.
Williams, A. C. and Moorehead, T. W. (eds.): 1986, *Proceedings of the Workshop on Double Layers in Astrophysics*, G. C. Marshall Space Flight Center, Huntsville, Alabama, U.S.A., March 17–19, NASA Conference Publ., No. 2469.
Williams, A. C., Weisskopf, M. C., Elsner, R. F., Darbro, W., and Sutherland, P. G.: 1986, *Astrophys. J.* **305**, 759.
Yamamoto, T.: 1985, *J. Plasma Phys.* **34**, 271.

COSMIC ELECTRIC CURRENTS AND THE GENERALIZED BENNETT RELATION*

PER CARLQVIST

Department of Plasma Physics, The Royal Institute of Technology, Stockholm, Sweden

(Received 15 June, 1987)

Abstract. A generalized form of the Bennett pinch is studied in both cylindrical geometry and plane-parallel geometry. In this kind of pinch electromagnetic forces, kinetic pressure gradient forces, centrifugal forces, and gravitational forces may act. For each of the two geometries considered a generalized Bennett relation is derived. By means of these relations it is possible to describe among other things the pure Bennett pinch, Jean's criterion in one and two dimensions, force-free magnetic fields, gravitationally balanced magnetic pressures, and continuous transitions between these states. The theory is applied to electric currents in the magnetosphere, in the solar atmosphere, and in the interstellar medium. It is pointed out that the currents in the solar atmosphere and in the interstellar medium may lead to pinches that are of vital importance to the phenomena of solar flares and star formation, respectively.

1. Introduction

Almost all cosmic plasmas that have been studied in detail seem to be penetrated by magnetic fields. The presence of the magnetic fields implies that considerable electric currents must exist in the cosmic plasmas. As has been stressed by Alfvén (1968, 1981, pp. 16) such currents often have a pronounced tendency to flow in relatively thin filamentary and sheet structures. Nearby examples of this are found in the ionosphere and magnetosphere of the Earth. *In situ* measurements show that currents here flow in a complex network of filaments and sheets. Also in more distant plasmas like the chromosphere and corona of the sun, the solar wind, and the interstellar medium we find thin structures in the form of filaments and sheets. There is strong evidence that many of these narrow structures are subject to the pinching action of electric currents.

A fundamental type of pinch of special interest in this connection is the Bennett pinch (Bennett, 1934). This kind of pinch consists of a cylindrical and fully ionized plasma which carries an axially directed current. The current together with the consequent toroidal magnetic field gives rise to a pinching force directed radially inward towards the axis of the plasma cylinder. In the steady state this force is balanced by a kinetic pressure gradient force directed outward. Bennett showed that in the case of constant ion and electron temperatures T_i and T_e, respectively, the relation

$$\mu_0 I^2 = 8\pi N_e k(T_i + T_e) \tag{1.1}$$

must be valid where I is the total current in the pinch and $N_e = N_i$ denotes the number of electrons (or ions) per unit length of the cylinder.

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 73–84.

The Bennett pinch constitutes a simple but important example of a current carrying plasma filament. In cosmic plasmas, however, other and more complex types of plasma pinches may also exist which differ from the Bennett pinch as regards both geometry and operating forces. In the following we shall study a more generalized type of pinch in both cylindrical and plane-parallel geometry which may be expected to be applicable to a broader class of current-carrying cosmic plasmas.

2. The Cylindrical Pinch

We consider a current-carrying, cylindrical plasma of radius a which consists of ions, electrons, and neutral gas having the densitites, n_i, n_e, and n_n and the temperatures T_i, T_e, and T_n, respectively. A current of density $\mathbf{i}_z$ flows in the plasma along the axis of the cylinder which coincides with the z-axis (cylindrical coordinates r, φ, z are adopted). As a result of the axial current a toroidal magnetic field $\mathbf{B}_\varphi$ is induced. There also exists an axial magnetic field $\mathbf{B}_z$ consistent with the toroidal current density $\mathbf{i}_\varphi$. Furthermore, the cylindrical plasma may rotate about the z-axis with the angular velocity ω. It is assumed that in the current carrying cylinder all quantities vary with r only and that there is a close coupling between the charged particles and the neutral gas. At the boundary of the cylinder the kinetic pressure is $P_k(a)$ while the axial magnetic field is $B_z(a)$.

In the steady state the plasma may be described by the force equation

$$-\varrho\omega^2\mathbf{r} = \mathbf{i} \times \mathbf{B} + \mathbf{F}_g - \operatorname{grad} p_k \,, \tag{2.1}$$

where ϱ is the mass density, $\mathbf{F}_g$ is the gravitational force per unit volume, and p_k is the total kinetic pressure of the particles. In the geometry considered all the forces have radial components only. The gravitational force at r is

$$\mathbf{F}_g(r) = -\frac{2G\overline{m}n}{r^2}\,\mathbf{r}\int_0^r 2\pi r m n \,\mathrm{d}r \,, \tag{2.2}$$

where $n = n_i + n_e + n_n$ is the total particle density, $\overline{m} = (m_i n_i + m_e n_e + m_n n_n)/n$ is the mean particle mass, while G is the gravitational constant.

The electromagnetic force may be rewritten as

$$\mathbf{i} \times \mathbf{B} = \mathbf{i}_\varphi \times \mathbf{B}_z + \mathbf{i}_z \times \mathbf{B}_\varphi \,. \tag{2.3}$$

The first term on the righthand side of this equation includes only straight and parallel magnetic field lines and may, hence, be expressed as

$$\mathbf{i}_\varphi \times \mathbf{B}_z = -\operatorname{grad} p_{Bz} \,, \tag{2.4}$$

where $p_{Bz} = B_z^2/2\mu_0$ is the magnetic pressure due to B_z. Using Equations (2.2) to (2.4) and the toroidal magnetic field

$$B_\varphi(r) = \frac{\mu_0}{2\pi r}\int_0^r 2\pi r i_z \,\mathrm{d}r \,, \tag{2.5}$$

we can express Equation (2.1) as

$$\frac{\mu_0 i_z}{2\pi r} \int_0^r 2\pi r i_z \, \mathrm{d}r + \frac{2G\overline{m}^2 n}{r} \int_0^r 2\pi r n \, \mathrm{d}r = -\frac{\mathrm{d}p_k}{\mathrm{d}r} - \frac{\mathrm{d}p_{Bz}}{\mathrm{d}r} + \overline{m} n \omega^2 r \,. \tag{2.6}$$

We now introduce the axial current inside the radius r

$$I(r) = \int_0^r 2\pi r i_z \, \mathrm{d}r \,, \tag{2.7}$$

and the total number of particles per u.l. (unit length) of a cylinder of radius r

$$N(r) = \int_0^r 2\pi r n \, \mathrm{d}r \,, \tag{2.8}$$

implying $\mathrm{d}I/\mathrm{d}r = 2\pi r i_z$ and $\mathrm{d}N/\mathrm{d}r = 2\pi r n$, respectively.

Inserting these relations into Equation (2.6) and integrating from the axis to the outer boarder of the current carrying plasma cylinder we find

$$\mu_0 \int_0^{I(a)} I \, \mathrm{d}I + 4\pi G\overline{m}^2 \int_0^{N(a)} N \, \mathrm{d}N = -4\pi^2 \int_0^a r^2 \left(\frac{\mathrm{d}p_k}{\mathrm{d}r} + \frac{\mathrm{d}p_{Bz}}{\mathrm{d}r} - \overline{m} n \omega^2 r \right) \mathrm{d}r,$$

which yields

$$\begin{aligned} \mu_0 I^2(a) + 4\pi G\overline{m}^2 N^2(a) = 8\pi[\, & W_k(a) - \pi a^2 p_k(a) + W_{Bz}(a) - \\ & - \pi a^2 B_z^2(a)/2\mu_0 + W_{\mathrm{rot}}(a)] \,; \end{aligned} \tag{2.9}$$

where

$$W_k(a) = \int_0^a 2\pi r p_k \, \mathrm{d}r \tag{2.10}$$

denotes the thermal energy of the particles per u.l. of the cylinder, while

$$W_{Bz}(a) = \int_0^a 2\pi r \frac{B_z^2}{2\mu_0} \, \mathrm{d}r \tag{2.11}$$

is the magnetic energy per u.l. of the cylinder due to the B_z-component, and

$$W_{\mathrm{rot}}(a) = \int_0^a 2\pi r \frac{\overline{m} n \omega^2 r^2}{2} \, \mathrm{d}r \tag{2.12}$$

is the kinetic energy of rotation per u.l. of the cylinder.

Putting

$$\Delta W_k(a) = W_k(a) - \pi a^2 p_k(a) \tag{2.13}$$

and

$$\Delta W_{Bz}(a) = W_{Bz}(a) - \pi a^2 B_z^2(a)/2\mu_0 , \tag{2.14}$$

we may write Equation (2.9) in the form

$$\mu_0 I^2(a) + 4\pi G \overline{m}^2 N^2(a) = 8\pi[\Delta W_k(a) + \Delta W_{Bz}(a) + W_{\rm rot}(a)] . \tag{2.15}$$

This equation is the analogue of the Bennett relation (1.1). A related type of equation has been considered by Witalis (1986). It is to be noticed that the distributions of the currents, the magnetic fields, the densities of the particles, the kinetic pressures, and the angular velocity do not appear explicitly in Equation (2.15). Only integrated quantities enter just as in the Bennett relation.

It may be of some interest to depict Equation (2.15) in a graphic form. In Figure 1 we show $I = I(a)$ as a function of $N = N(a)$ for some discrete values of $\Delta W_{Bz} = \Delta W_{Bz}(a)$ in the case of a non-rotating ($\omega = 0$) and fairly cold ($T_i = T_e = T_n = T = 20$ K) plasma mainly consisting of hydrogen molecules ($\overline{m} \approx 3 \times 10^{-27}$ kg) where $W_k(a) \gg \pi a^2 p_k(a)$. Several physically different regions may be distinguished in the figure.

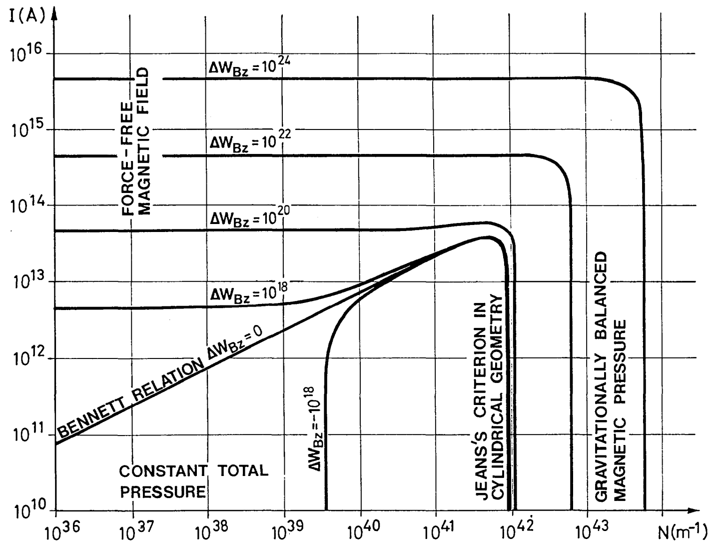


Fig. 1. The total current I in a generalized Bennett pinch of cylindrical geometry is shown as a function of the number of particles per unit length of the pinch N. The temperature of the plasma is $T = 20$ K while the mean particle mass is $\overline{m} = 3 \times 10^{-27}$ kg. It is assumed that the plasma does not rotate ($\omega = 0$) and that the kinetic pressure is much smaller at the border of the pinch than in the inner parts. The parameter of the curves is ΔW_{Bz} representing the excess magnetic energy per unit length of the pinch due to an axial magnetic field B_z. The diagram includes several physically distinct regions representing the Bennett relation, Jeans's criterion in cylindrical geometry, gravitationally balanced magnetic pressure, constant total pressure, and force-free magnetic field.

First there is a region in the upper left-hand part of the figure where the pinching force due to I and the magnetic pressure force due to B_z constitute the dominating forces. Equation (2.15) then reduces to

$$\mu_0 I^2 \approx 8\pi \Delta W_{Bz}\,, \tag{2.16}$$

representing a state of an almost force-free magnetic field (Boström, 1973).

A second region is situated well below the curve labeled $\Delta W_{Bz} = 0$. Here ΔW_{Bz} assumes negative values and an outwardly directed kinetic pressure force is mainly balanced by an inwardly directed magnetic pressure force. Hence the total pressure is constant and Equation (3.15) reduces to

$$NkT + \Delta W_{Bz} \approx 0\,. \tag{2.17}$$

A third region is found well to the right of the vertical part of the curve $\Delta W_{Bz} = 0$. In this region the magnetic pressure force neutralizes the gravitational force and Equation (2.15) is simplified to

$$G\overline{m}^2 N^2 \approx 2\Delta W_{Bz}\,. \tag{2.18}$$

In addition to the three regions mentioned above there are also two special relations which are connected with the curve $\Delta W_{\mathrm{Bz}} = 0$. On the left-hand slope of this curve the gravitational force is negligible and Equation (2.15) reduces to the Bennett relation

$$\mu_0 I^2 \approx 8\pi NkT\,. \tag{2.19}$$

On the right-hand, almost vertical part of the curve the pinching force of the current may be neglected. The kinetic pressure force and the gravitational force here balance one another so that

$$G\overline{m}^2 N^2 \approx 2NkT\,. \tag{2.20}$$

This is the Jeans's criterion in a cylindrical geometry (cf. McCrea, 1957).

In Figure 1 we have chosen to show a special case of Equation (2.15). However, if the parameter of the curves ΔW_{Bz} is replaced by $\Delta W_{Bz}(a) + W_{\mathrm{rot}}(a) - \pi a^2 p(a)$ the figure will illustrate the general relation (2.15) including rotation and external pressure. It is to be noticed that the designations of the various regions of the figure are not valid in this latter case.

3. The One-Dimensional Pinch

We now consider a stationary plasma pinch of plane parallel geometry (orthogonal coordinates x, y, z are adopted). A current of density $i_z(x)$ flows in the z-direction in a symmetrical slab limited by the planes $x = \pm d$. This current induces a magnetic field $B_y(x)$. There also exists a magnetic field in the z-direction, $B_z(x)$, consistent with the current density $i_y(x)$. The plasma consists of ions, electrons, and neutrals of densities n_i, n_e, and n_n and temperatures T_i, T_e, and T_n, respectively. As before it is assumed that there is a close coupling between the charged particles and the neutrals. All quantities are supposed to vary with the x-coordinate only.

In a similar way as Equation (2.15) was derived for the cylindrical pinch in Section 2 we can by using the force equation (without the centrifugal term) derive the following relation

$$\mu_0 \mathscr{I}^2(d) + 4\pi G\overline{m}^2 \mathscr{N}^2(d) = 8(\Delta p_k + \Delta p_{Bz}) \tag{3.1}$$

for the one-dimensional pinch, where

$$\mathscr{I}(d) = 2 \int_0^d i_z \, \mathrm{d}x \tag{3.2}$$

is the current per u.l. of the slab, and

$$\mathscr{N}(d) = 2 \int_0^d (n_i + n_e + n_n) \, \mathrm{d}x \tag{3.3}$$

is the number of particles per u.a. (unit area) of the slab. Furthermore,

$$\Delta p_k = p_k(0) - p_k(d) \tag{3.4}$$

is the kinetic pressure difference in the slab, while

$$\Delta p_{Bz} = p_{Bz}(0) - p_{Bz}(d) \tag{3.5}$$

is the difference of magnetic pressure due to B_z.

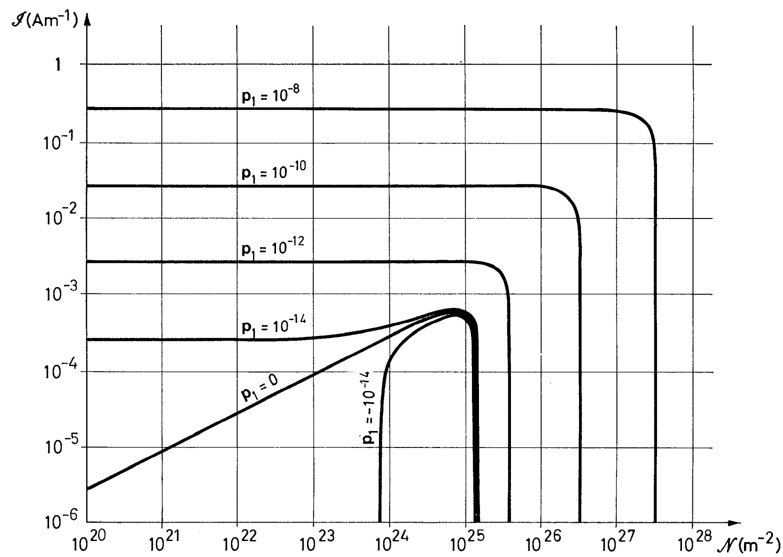


Fig. 2. The current per unit length $\mathscr{I}$ as a function of the number of particles per unit area $\mathscr{N}$ in a generalized Bennett pinch of plane parallel geometry forming a slab. The temperature and mean particle mass are the same as in Figure 1. It is assumed that $n(0) \approx \mathscr{N}/d$. The parameter of the curves is $p_1 = \Delta p_{Bz} - p_k(d)$ where $p_k(d)$ is the kinetic pressure at the borders of the slab while Δp_{Bz} denotes the difference of the magnetic pressure due to B_z between the centre of the slab and the borders.

Equation (3.1) is the one-dimensional analogue of Equation (2.15). In Figure 2 we have depicted $\mathscr{I} = \mathscr{I}(d)$ as a function of $\mathscr{N} = \mathscr{N}(d)$ as obtained from Equation (3.1) for some different values of the quantity, $p_1 = \Delta p_{Bz} - p_k(d)$, when $\overline{m} = 3 \times 10^{-27}$ kg and $T_i = T_e = T_n = T = 20$ K. It is assumed that $n(0) \approx \mathscr{N}/d$. As a comparison reveals there is a striking similarity between Figure 1 and Figure 2.

By means of the theory discussed in this and the previous sections we shall in the next section study the physical conditions in a few different, current-carrying, cosmic plasmas.

4. Application of the Theory to Cosmic Currents

4.1. Birkeland Currents in the Magnetosphere of the Earth

From magnetometer measurements performed from rockets and satellites it is known that considerable electric currents can flow along the magnetic field lines in the ionosphere and magnetosphere of the earth (Zmuda *et al.*, 1966, 1967; Vondrak *et al.*, 1969). The currents, which are usually called Birkeland currents, often exist in sheets in the auroral zones where the current density may be as large as $i_{\|} \approx 10^{-4}$ A m^{-2}. The thickness of such sheets detectable by magnetometers is found to range from a few kilometres to several hundred kilometres (see, e.g., Burke, 1984; Primdahl *et al.*, 1984). Often the sheets exist in pairs with oppositely directed currents.

There is strong evidence that auroral structures such as arcs and rays are intimately connected with upwardly directed Birkeland currents. Since auroral rays are observed to have a thickness down to some hundred metres it is likely that the Birkeland currents may also flow in structures of equally small dimensions.

(A) We first consider a model of a Birkeland current with the current density $i_z \approx 3 \times 10^{-5}$ A m^{-2} flowing in a magnetospheric plasma slab of half-thickness $d = 10^4$ m. It is assumed that the plasma in the slab consists mainly of electrons and protons with the mean mass $\overline{m} \approx 10^{-27}$ kg. The density and temperature of the plasma is $n_e = n_i \approx 10^{10}$ m^{-3} and $T \approx 2 \times 10^3$ K, representative for the plasma at a height of some thousand kilometres above the Earth. The current per u.l. of the slab is then $\mathscr{I}(d) \approx 6 \times 10^{-1}$ A m^{-1} while the number of particles per u.a. is $\mathscr{N}(d) \approx 4 \times 10^{14}$ m^{-2}. Furthermore, the magnetic field is $B_z \approx 4 \times 10^{-5}$ T.

Assuming the plasma to be in the steady-state we can apply the theory of Section 3 and find the magnitude of the terms in Equation (3.1): $\mu_0 \mathscr{I}^2(d) \approx 5 \times 10^{-7}$ Pa, $4\pi G \overline{m}^2 \mathscr{N}^2(d) \approx 1 \times 10^{-34}$ Pa, $8\Delta p_k < 8p_k(0) \approx 4 \times 10^{-9}$ Pa, $8\Delta p_{Bz} < 8p_{Bz}(0) \approx$ $\approx 5 \times 10^{-3}$ Pa. From this it is clear that both the gravitational term and the kinetic pressure term are negligible as compared with the pinching current term. Hence, the pinching force due to the current $\mathscr{I}(d)$ has to be balanced by a magnetic pressure force due to a slight difference between the pressures $p_{Bz}(0)$ and $p_{Bz}(d)$. We are, therefore, here dealing with an almost force-free magnetic field configuration.

(B) We now turn to a model of a thin Birkeland current filament in the magnetosphere constituting a cylinder with the radius $a = 10^2$ m. As in the previous case (A) we assume that $i_z \approx 3 \times 10^{-5}\,\mathrm{A\,m^{-2}}$, $\overline{m} \approx 10^{-27}$ kg, $n_e = n_i \approx 10^{10}\,\mathrm{m^{-3}}$, $T \approx 2 \times 10^3$ K, and $B_z = 4 \times 10^{-5}$ T which yield $I(a) \approx 9 \times 10^{-1}$ A and $N(a) \approx 6 \times 10^{14}\,\mathrm{m^{-1}}$. We apply Equation (2.15) and find the magnitude of the terms: $\mu_0 I^2(a) \approx 1 \times 10^{-6}\,\mathrm{J\,m^{-3}}$, $4\pi G\overline{m}^2 N^2(a) \approx 3 \times 10^{-34}\,\mathrm{J\,m^{-3}}$, $8\pi\Delta W_k(a) < 8\pi N(a)kT \approx 4 \times 10^{-4}\,\mathrm{J\,m^{-3}}$, $8\pi\Delta W_{Bz}(a) < 8\pi W_{Bz}(a) \approx 5 \times 10^2\,\mathrm{J\,m^{-3}}$. Again the gravitational term is negligible. However, in contrast to the previous case the pinching current term may now be balanced in several ways. It may for instance be neutralized either by the kinetic pressure term or by the magnetic pressure term. It may also be balanced by the centrifugal term if the cylindrical plasma rotates with the characteristic velocity $v_c \approx 4 \times 10^2\,\mathrm{m\,s^{-1}}$. Such a rotatory velocity may readily be caused by a double layer or some other resistive mechanism in the filament (Carlqvist and Boström, 1970). Hence, we may draw the conclusion that the magnetic field need not necessarily be force-free in this case.

It should be pointed out that in both examples discussed above the currents are limited to relatively thin structures by processes high up in the magnetosphere (probably at heights where the currents are generated). The pinching action of the currents on the plasma and magnetic field is in these examples very small. In the following we shall consider cosmic currents which may give rise to more pronounced pinches.

4.2. Currents in the Solar Atmosphere

It is well-known that the solar atmosphere consists of a highly conducting plasma in which considerable electric currents flow. Here we shall discuss two different types of current systems in the solar atmosphere.

(A) Using a magnetograph Moreton and Severny (1968) were able to show that vertical currents of more than $\sim 2 \times 10^{11}$ A flow up and down through the photosphere of the Sun. The areas in which the currents were detected had linear dimensions of a few thousand kilometers nearly corresponding to the resolution of the instrument used. It was also found that the regions of concentrated currents coincided with bright flare knots. Later similar measurements of vertical currents have been made by Hagyard *et al.* (1984). From the measurements it can be concluded that vertical currents of about 10^{11}–10^{12} A are fairly common in active regions where solar flares occur.

We now consider a model of such an electric current in the solar atmosphere, $I = 3 \times 10^{11}$ A, flowing in a filament with the radius of cross-section $a \lesssim 10^6$ m. The filament has a length of $l \approx 10^8$ m and passes mainly through the lower corona in a loop connecting two foot-points in the photosphere. Assuming the plasma in most of the filament to be of coronal density $n_e = n_i \lesssim 10^{16}\,\mathrm{m^{-3}}$ and temperature $T \approx 10^6$ K we find from Equation (2.15) that the kinetic pressure term cannot balance the pinching current term $\mu_0 I^2 \approx 10^{17}\,\mathrm{J\,m^{-1}}$. Neither is the rotatory term (for reasonable rotational velocities) or the gravitational term of any importance. Instead $\mu_0 I^2$ must be neutralized by the magnetic pressure term $8\pi\Delta W_{Bz}$ implying a force-free magnetic field. Denoting the characteristic axial magnetic field inside the filament by B_{zc} and assuming the magnetic field outside to be much smaller we find from Equation (2.15)

$$aB_{zc} = \frac{\mu_0 I}{2\pi} \,. \tag{4.1}$$

Inserting the current we get $aB_{zc} \approx 6 \times 10^4$ T m. This gives, with a radius $a \approx 10^5$–10^6 m, a magnetic field $B_{zc} \approx 6 \times 10^{-1}$–$6 \times 10^{-2}$ T and an axial magnetic field flux $\sim 2 \times 10^{10}$–2×10^{11} Wb in the filament.

With the above values of B_{zc}, a, and l the magnetic energy stored in the filament is of the order $\sim 10^{24}$ J. This energy corresponds well to the typical energy of solar flares. It has been suggested that the stored magnetic energy might quickly (~ 10–10^2 s) be released as a solar flare by a double layer occurring in a current filament (Alfvén and Carlqvist, 1967; Carlqvist, 1969, 1986). Support for these ideas is found in basic observations from Skylab and the Solar Maximum Mission satellite indicating that the flare is primarily a high-energy phenomenon originating in loop-like structures.

(B) Next we turn to the more extensive heliospheric current system which is generated by the solar rotation combined with the solar wind (Alfvén, 1977, 1981, pp. 53). The heliospheric current mainly exists in the solar wind plasma. It flows in towards the Sun (or out from it – depending on the magnetic polarity of the Sun) in a thin and wavy layer near the equatorial plane of the Sun. It then crosses the Sun, and flows out from (or in towards) the two polar regions of the Sun. The circuit closes at large distances from the Sun – probably close to the heliopause. The magnitude of the current in each of the polar regions is $I_p \approx 1.5 \times 10^9$ A.

It is not known whether the polar current is concentrated to filaments or smeared out more evenly in the polar regions. We shall here study the conditions for the current to pinch to one or more *pronounced* filaments in the lower corona ($1.1\ R_\odot$). We assume that the polar current is divided into M filaments, each carrying the current $I = I_p/M$. Because of the relatively low mass of the filaments the gravitational and centrifugal forces are of negligible importance. The pinching current term $\mu_0(I_p/M)^2$ in Equation (2.15) can then be balanced only by the kinetic pressure term $8\pi\Delta W_k(a)$ or by the magnetic pressure term $8\pi\Delta W_{Bz}(a)$. In order for a pronounced filament to exist we require either the magnetic field or the density inside the filament to be well above the corresponding quantities outside the filament.

If on the one hand $\mu_0(I_p/M)^2$ is balanced by $8\pi\Delta W_{Bz}(a)$ we obtain from Equation (2.15) the filamentary radius

$$a \approx \frac{\mu_0 I_p}{2\pi M B_{zc}}, \tag{4.2}$$

where B_{zc} represents a characteristic value of the axial magnetic field in the filament. Putting $B_{zc} \gtrsim 10^{-3}$ T and inserting the value of I_p we find that the filamentary radius must be $a \lesssim 3 \times 10^5$ m for $M = 1$ and $a \lesssim 3 \times 10^3$ m for $M = 10^2$.

If on the other hand $\mu_0(I_p/M)^2$ is balanced by $8p\Delta W_k(a)$ the radius of the filament as obtained from Equation (2.15) is

$$a \approx \frac{I_p}{4\pi M} \sqrt{\frac{\mu_0}{n_{ec} kT}}, \tag{4.3}$$

where $n_{ec} = n_{ic}$ denotes the characteristic density of the particles in the filament. Putting

$n_{ec} = n_{ic} \gtrsim 10^{15}\ \mathrm{m}^{-3}$, $T \approx 10^6$ K, and $I_p \approx 1.5 \times 10^9$ A we get $a \lesssim 10^6$ m for $M = 1$ and $a \lesssim 10^4$ m for $M = 10^2$.

From these results we see that if pronounced filaments exist in the polar corona as a result of the pinching action of the heliospheric current then they must be so thin that they are almost impossible to observe at present. This does not mean, however, that thicker filamentary currents of heliospheric origin are excluded in the polar corona. It only implies that such currents cannot appreciably influence the coronal magnetic field and density.

4.3. Currents in the Interstellar Medium

We finally consider a current system on a galactic scale which has been suggested by Alfvén (1977, 1978). In this current system a rotating and magnetized Galaxy serves as a unipolar generator. Electric currents flow in the equatorial plane of the Galaxy, follow along the axis of rotation, and close at large distances outside the Galaxy. Hence, the galactic current system, although enlarged by a factor $\sim 10^8$–10^9, resembles the helicospheric current system. The magnitude of the galactic current has been estimated to be of the order 10^{17}–10^{19} A (Alfvén and Carlqvist, 1978).

Are there now any signs of such currents in our Galaxy? One type of phenomenon which may indicate the presence of electric currents is, as we have seen, the filamentary structure. An inspection of optical, infrared, and radio data reveals that filamentary structures are fairly common in the interstellar medium of our Galaxy. Many of these structures are intimately connected with supernova remnants and, therefore, represent relatively small systems. However, there also exist filaments which are not obviously associated with supernovas or other similar phenomena. In this latter category we find many of the molecular clouds (or dust clouds) out of which stars may form. One problem connected with these clouds consists in explaining how they once could be collected into filaments (or slabs) before the gravitational forces were strong enough to contract them. Here the pinching effect due to electric currents offers an attractive mechanism (Alfvén and Carlqvist, 1978).

To exemplify this we consider a model of a weakly ionized and filamentary plasma cloud which mainly consists of hydrogen molecules of mass $\sim 3 \times 10^{-27}$ kg and temperature $T \approx 20$ K. The radius of the filament is $a = 2 \times 10^{16}$ m $= 0.65$ pc, while the mean density is $n \approx 3 \times 10^8\ \mathrm{m}^{-3}$ yielding $N \approx 4 \times 10^{41}\ \mathrm{m}^{-1}$. It is assumed that the external pressure is much less than the internal pressure. If the cloud does not rotate and the axial magnetic field is zero, so that $\Delta W_{Bz} = 0$, we find from Figure 1 that an electric current of $I \approx 4 \times 10^{13}$ A is required to pinch the plasma.

In the case of an internal axial magnetic field, $B_{zic} = 6 \times 10^{-10}$ T, and a corresponding external field, $B_{zec} = 3 \times 10^{-10}$ T, yielding $\Delta W_{Bz} \approx 1 \times 10^{20}\ \mathrm{J\,m}^{-3}$ we find from Figure 1 that the current needed for pinching is $I = 6 \times 10^{13}$ A. A current of this magnitude induces a toroidal magnetic field of up to $B_{\varphi m} = 6 \times 10^{-10}$ T. This field together with the axial field considered above are well below the upper limits of the magnetic fields, 1.5×10^{-9}–5×10^{-9} T measured for a few molecular clouds (Crutcher *et al.*, 1975, 1981).

If the plasma in the filament rotates the filamentary current I has to be larger than that in a similar, non-rotating filament to pinch the plasma. In the model above the effect of rotation becomes comparable with the effect of the kinetic pressure when the rotatory velocity is $v_c \approx 4 \times 10^2 \text{ m s}^{-1}$.

Matter may be collected to the filament either along the filamentary axis or transverse to it. As long as $N \ll 10^{42} \text{ m}^{-1}$ the pinching of the filament is controlled mainly by the current I while for $N \gtrsim 10^{42} \text{ m}^{-1}$ gravitational forces dominate. The pinching by filamentary currents may thus be an important mechanism in the formation phase of molecular clouds. It is to be noticed that the currents needed for this process constitute only a small fraction of the total galactic current. Hence, a great number of cloud-forming current filaments may exist simultaneously in the Galaxy. When a cloud has become sufficiently condensed stars may be formed in it. Dust existing in the cloud is then probably of vital importance for the formation process (Alfvén and Carlqvist, 1978).

5. Conclusions

In the preceding sections we have studied the generalized Bennett pinch in one and two dimensions and considered a few cosmic applications of it. By integrating the force equation and using Maxwell's equations we derived the basic relations (2.15) and (3.1). The relations found might also be interpreted in terms of the virial theorem (cf. Chandrasekhar and Fermi, 1953; McCrea, 1957; Mestel, 1965a, b). By means of the relations (2.15) and (3.1) it is possible to describe among other things the pure Bennett pinch, the Jeans's criterion in one and two dimensions, force-free magnetic fields, gravitationally balanced magnetic pressures, and continuous transitions between all these states.

In the models treated we have limited ourselves to two simple geometries – the cylindrical geometry and the plane geometry. In cosmic plasmas it is likely that there also exist more complex geometries. Thus it is possible that a filament may consist of a bundle of thinner filaments or that a slab may be split up in a row of filaments. Such subfilaments may in principle be treated similarly as the models considered.

Another interesting possibility is that the filaments or slabs may be subject to instabilities of various kinds. For instance the sausage instability may occur in cases where $B_z = 0$ or the kink instability where $B_z \neq 0$ (see, e.g., Dattner *et al.*, 1958). Such instabilities might be of importance in the processes of star formation. A more thorough discussion on this interesting subject is, however, beyond the scope of the present paper.

Acknowledgement

I wish to thank Dr M. Raadu for reading the manuscript and for several valuable suggestions.

References

Alfvén, H.: 1968, *Ann. Geophys.* **24**, 1.
Alfvén, H.: 1977, *Rev. Geophys. Space Phys.* **15**, 271.
Alfvén, H.: 1978, *Astrophys. Space Sci.* **54**, 279.
Alfvén, H.: 1981, *Cosmic Plasma*, D. Reidel Publ. Co., Dordrecht, Holland.
Alfvén, H. and Carlqvist, P.: 1967, *Solar Phys.* **1**, 220.
Alfvén, H. and Carlqvist, P.: 1978, *Astrophys. Space Sci.* **55**, 487.
Bennett, W. H.: 1934, *Phys. Rev.* **45**, 890.
Boström, R.: 1973, *Astrophys. Space Sci.* **22**, 353.
Burke, W. J.: in T. Potemra (ed.), *Magnetospheric Currents,* AGU Geophysical Monograph 28, Washington, DC, Amer. Geophys. Union, p. 294.
Carlqvist, P.: 1969, *Solar Phys.* **7**, 377.
Carlqvist, P.: 1986, *IEEE Trans. Plasma Sci.* **PS-14**, 794.
Carlqvist, P. and Boström, R.: 1970, *J. Geophys. Res.* **75**, 7140.
Chandrasekhar, S. and Fermi, E.: 1953, *Astrophys. J.* **118**, 116.
Crutcher, R. M., Evans, N. J., II, Troland, T. H., and Heiles, C.: 1975, *Astrophys. J.* **198**, 91.
Crutcher, R. M., Troland, T. H., and Heiles, C.: 1981, *Astrophys. J.* **249**, 134.
Dattner, A., Lehnert, B., and Lundquist, S.: 1958, *Proc. 2nd UN Int. Conf. on the Peaceful Uses of Atomic Energy*, Geneva, Vol. 31, Pergamon Press, p. 325.
Hagyard, M. J., West, E. A., and Smith, J. B.: 1984, No. 84–117, Space Science Laboratory, Marshall Space Flight Center, Alabama 35812, U.S.A. To be published in the Proceedings of the Kunming Workshop on Solar Physics and Interplanetary Travelling Phenomena.
McCrea, W. H.: 1957, *Monthly Notices Roy. Astron. Soc.* **117**, 562.
Mestel, L.: 1965a, *Quart. J. Roy. Astron. Soc.* **6**, 161.
Mestel, L.: 1965b, *Quart. J. Roy. Astron. Soc.* **6**, 265.
Moreton, G. E. and Severny, A. B.: 1968, *Solar Phys.* **3**, 282.
Primdahl, F., Bahnsen, A., Ejiri, M., Høeg, P., Marklund, G., Maehlum, B. N., Olesen, J. K., Ungstrup, E., and Zanetti, L. J.: 1984, *Planetary Space Sci.* **32**, 561.
Vondrak, R. R., Anderson, H. R., and Spiger, R. J.: 1969, *Trans. AGU* **50**, 663.
Witalis, E. A.: 1986, *IEEE Trans. Plasma Sci.* **PS-14**, 842.
Zmuda, A. J., Martin, J. H., and Heuring, F. T.: 1966, *J. Geophys. Res.* **71**, 5033.
Zmuda, A. J., Heuring, F. T., and Martin, J. H.: 1967, *J. Geophys. Res.* **72**, 1115.

ELEMENTARY IDEAS BEHIND PLASMA PHYSICS*

WALTER J. HEIKKILA

Center for Space Science, University of Texas at Dallas, Richardson, Texas, U.S.A.

(Received 7 October, 1987)

"... dubious principle frozen into dogma"
Willy Brandt

Abstract. This contribution is in support of Alfvén's use of circuit theory to advance the understanding of complex plasma physical problems, such as magnetic reconnection; these ideas have often been misunderstood. Circuit analysis is not a full description of the physics, being a scalar relationship. However, it is suitable for dealing with energy relationships and cause and effect, and it is fully capable of showing fallacies behind various fashionable ideas, including steady-state reconnection.

1. Introduction

The above quotation (Newsweek, May 20, 1974) did not, of course, refer to the subject at hand, but it does seem appropriate. The concept of frozen field convection of magnetic field lines in a plasma (Alfvén, 1950) is not a fundamental principle of physics. Nevertheless, it has become so widely used and accepted as to have attained nearly the status of dogma (despite the protests of Alfvén). "In low density plasmas the concept of frozen-in lines of force is questionable. The concept of frozen-in lines of force may be useful in solar physics where we have to do with high and medium density plasmas, but it may be grossly misleading if applied to the magnetosphere of the Earth. To plasma in interstellar space it should be applied with care" (Alfvén and Fälthammar, 1963, p. 161). Alfvén has repeated these words of caution on countless occasions. "I thought that the frozen-in concept was very good from a pedagogical point of view, and indeed it became very popular. In reality, however, it was not a good pedagogical concept but a dangerous 'pseudo-pedagogical concept'. By 'pseudo-pedagogical' I mean a concept which makes you believe that you understand a phenomenon whereas in reality you have drastically misunderstood it" (Alfvén, 1986).

There is another characteristic of Alfvén's that is far from being generally appreciated, that of using circuit diagrams to analyze plasma problems. In a review of his book *Cosmic Plasma* (1981), a reviewer (Kulsrud, 1983) said "It is not easy to understand how pieces of plasma are to behave like wires while the rest of the plasma is to be ignored". Similar statements have been voiced by others (e.g., Vasyliunas, 1984).

I will come to Hannes Alfvén's defense on these two questions, as I too have had difficulties with referees, and readers. In regard to circuit analysis, I will quote from text books on electricity and magnetism to show that there should be nothing controversial here. Then, I will concentrate on the phenomenon of magnetic reconnection.

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

2. Electrical Circuits

The use of circuit concepts will be discussed first, since we can utilize them to try to understand reconnection theories. It should be said immediately that an electrical circuit for a plasma is not a complete description; circuit analysis leads to a scalar equation, useful for comprehending energy relationships in a plasma. In particular, circuit equations contain no information about the forces at work in a plasma. Equations of motion for the plasma are vector relationships, which play no part in circuit theory; they do play a part in determining the appropriate circuit, a feat which may require good intuition.

Actually, there should be no disagreement whatsoever, since this is all text book material; Ramo and Whinnery (1953, second edition) discuss circuit concepts and their derivation from the field equations in Chapter 5. "In a circuit problem there is often an applied voltage, and there are currents in the conductors of the circuit, charges on condensors in the circuit, ohmic losses, and power losses by radiation. These effects include almost everything that can happen when electric currents, charges, and conductors are let loose. The circuit problem is also one of the commonest problems illustrating the ideas of cause and effect relationships... From the rigorous starting point of the fundamental laws, it will be found that for circuits which are small compared with wavelength, this exact approach leads directly to the familiar circuit ideas based upon Kirchhoff's laws, and the concepts of lumped inductances and capacitances are sufficient for analysis."

Jordan and Balmain (1968) have made similar statements. "The electric circuit laws of Ohm, Faraday, and Kirchhoff were based on experimental observations and antedated the electromagnetic theory of Maxwell and Lorentz. Indeed, the theory was developed as a generalization from these simpler and more restricted laws. It is interesting, but not surprising, then, to find that the circuit relations are just special cases of the more general field relations, and that they may be developed from the latter when suitable approximations are made. Nevertheless, the importance of the simpler (and approximate) circuit relations should not be underestimated... The extension of circuit concepts to higher frequencies is accomplished in practice by the addition of appropriately located lumped-circuit constants. For example, 'distributed' inductance and capacitance effects are accounted for by suitably located series inductors and shunt capacitors, and radiation effects by the inclusion of a 'radiation resistance'. An outstanding example in electrical engineering of the extension of circuit concepts to systems not necessarily small in wavelengths is the ordinary transmission line." We physicists must admit that engineers are clever.

Circuit ideas for a plasma stem directly from the continuity equation for charge; in particular, the fact that the divergence of an electrical current in a plasma is small, approximately zero, makes it easy to visualize a plasma current with no beginning or end, in other words, an electric current circuit. This is illustrated in Figure 1(a). These currents may be field-aligned in part; they can be called 'force transfer currents'. It is immediately obvious that the circuit can be divided into three parts depending on the

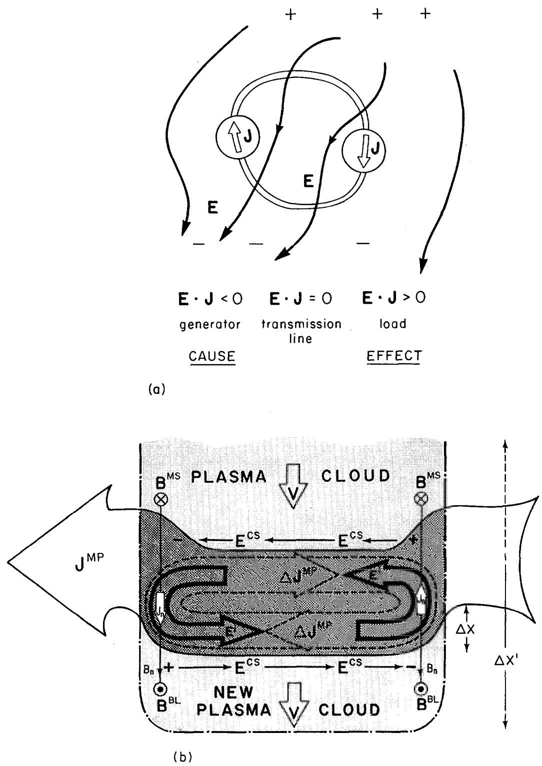


Fig. 1. (a) Magnetospheric currents must be closed in general, unless we are considering radiation phenomena, so that $\operatorname{div} \mathbf{J} = 0$. In the presence of an electric field, regions in which the current opposes the electric field with $\mathbf{E} \cdot \mathbf{J} < 0$ behave as generators or dynamos (the cause), while regions in which $\mathbf{E} \cdot \mathbf{J} > 0$ act as electrical loads (the effect). (b) Here the current perturbation (and the associated inductive electric field) forms a localized circuit, with an electric load preceded by a dynamo.

sign of $\mathbf{E} \cdot \mathbf{J}$, and making use of Poynting's theorem,

$$\int_{\text{vol}} \mathbf{E} \cdot \mathbf{J} \, d\tau = -\frac{1}{\mu_0} \int_s \mathbf{E} \times \mathbf{B} \, d\mathbf{S} - \frac{1}{\mu_0} \int \mathbf{B} \cdot \partial \mathbf{B}/\partial t \, d\tau - \varepsilon_0 \int_{\text{vol}} \mathbf{E} \cdot \partial \mathbf{E}/\partial t \, d\tau,$$

with a dynamo connected to a load. The term on the left in Poynting's theorem is the energy dissipation in the plasma (or release, depending on the sign). For the steady state, only the first term on the right is non-zero; then the electric field is curl-free such that $\mathbf{E} = -\operatorname{grad} \phi$. Using a well-known vector formula (Heikkila *et al.*, 1979), the equation becomes

$$\int_{\text{vol}} \mathbf{E} \cdot \mathbf{J} \, d\tau = -\int_s \phi \mathbf{J} \cdot d\mathbf{S}.$$

This shows that energy can be transferred into the volume by means of a current flowing in at the higher potential and out at the lower, just as in resistive network (see also Plonsey and Collin, 1961, p. 340). This is entirely equivalent to its representation as a Poynting flux through the bounding surface (Vasyliunas, 1968; Heikkila *et al.*, 1979). We can go a step further and define lumped constant circuit elements, following Plonsey and Collin (1961). The capacitance, inductance, and resistance are defined in terms of the energy stored in the electric field W_e, and the magnetic field W_m, and the energy dissipation P_e, respectively, according to

$$C = \frac{II^*}{4\omega^2 W_e} = \frac{4W_e}{VV^*}, \qquad L = \frac{4W_m}{II^*} = \frac{VV^*}{4\omega^2 W_m}, \qquad R = \frac{2P_e}{II^*} = \frac{VV^*}{2P_e};$$

where I and V denote current and voltage at a particular frequency ω, and * denotes the complex conjugate.

To quote from Plonsey and Collin (1961, p. 337): "Under static conditions, it has been shown in earlier chapters that the above definitions for R, L, and C are equivalent to the geometric definitions. However, the above definitions are more general in that they recognize the fact that ideal circuit elements do not exist physically... In other words, by defining capacitance in terms of electric energy storage, account is taken of all portions of the physical structure that contribute to the capacitance of the overall device, and similarly for the inductance and resistance... We emphasize that it is necessary to be able to define unique terminal currents and voltages in order for these parameters to have unique values." Thus, a certain amount of intuition is required to apply these concepts to a plasma, but success brings an ability to look at cause and effect relations in a very simple manner.

In Figure 1(b) a sheet current is shown, which is being indented locally with a constant velocity (for a more complete description, see Heikkila, 1982, 1984b). An electric field with a curl is shown, i.e., $\oint \mathbf{E} \cdot d\mathbf{l}$ is finite, representing an electromotove force; thus, this model is both localized and time-dependent. Inclusion of the electromotive force permits drawing on the stored energy within the volume (Heikkila, 1982, 1984a), using the second term of the right-hand side in Poynting's theorem. We can use Lenz's law to understand the nature of this inductive electric field. Where the current is being reduced, Lenz states that an electric field will be induced which will try to maintain the current (think of a resistive medium); $\mathbf{E} \cdot \mathbf{J}$ will be positive, and that region behaves as an electrical load. On the other side of the current sheet, the magnitude is increasing (by assumption), and according to Lenz's law $\mathbf{E} \cdot \mathbf{J}$ is negative here; the plasma behaves as a dynamo in this region. The magnitude can be calculated by using the integral form of Faraday's law. The current perturbation $\Delta\mathbf{J}$ behaves as an electrical circuit, showing that the energy dissipated in the load comes from the adjacent region just ahead, in fact from the momentum of the plasma, and also from stored magnetic energy (Heikkila, 1982).

This is an attractive model, since the energy for the load is immediately available; there is little time delay involved in the transmission line (where the energy coupling should occur with the Alfvén velocity). More will be said later on this subject, in connection with steady-state reconnection where the energy must come from a distant dynamo.

3. Reconnection

There is no doubt that in a plasma with little spatial or temporal structure we can use the frozen-in theorem (the precise meaning of this phrase is discussed in many plasma physics text books, including Alfvén and Fälthammar, 1963). One can say that a given magnetic field line, or flux tube, is defined by the (low temperature) plasma on it, which we can call a fluid element. These fluid elements may be in motion, associated with an electric field in a fixed reference frame. Thus, as is common, we can discuss the meaning of magnetic reconnection using a relatively simple diagram (Axford, 1984), as in Figure 2. At time t_1 the magnetic field line defined by fluid elements A and B approaches a more or less oppositely-directed field line defined by fluid elements C and D. It may be possible that, at some time, t_2 the two field lines touch at some intermediate point, where the frozen-in theorem may not hold because of the finite gyroradii of the plasma particles when the magnitude of the field becomes small. Meanwhile, the other fluid elements $ABCD$, still in motion, continue to define field lines, as at time t_3, but now with different interconnections. We can speak of the ends of the field line at the point of contact being reconnected at time t_2 at what has come to be called a reconnection line, or an X-line. By symmetry, the X-line must be a stagnation point for the plasma flow, with the plasma being deflected or ejected by the reconnection process in other directions. The 'reconnection rate' is given by the electric field E_r along the X-line, and

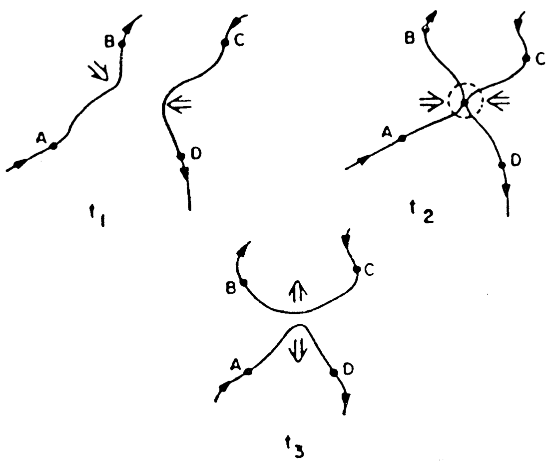


Fig. 2. At time t_1 the magnetic field line defined by fluid elements A and B approaches a more or less oppositely-directed field line defined by fluid elements C and D. At time t_2 the field lines touch at some intermediate point between A and B and between C and D. At time t_3 the fluid elements define new field lines, with $A-D$ moving away from $B-C$. We can say that the field lines are reconnected (Axford, 1984).

is equal to the amount of magnetic flux transported per unit time across a unit length of the X-line (see, e.g., Sonnerup, 1985). The definition emphasizes flux transport between different topological cells (Sonnerup, 1985). Vasyliunas (1975) uses a different definition based on plasma flow across the separatrix (the surface defined by field lines coming from the X-line), but with equivalent results (see Vasyliunas, 1983).

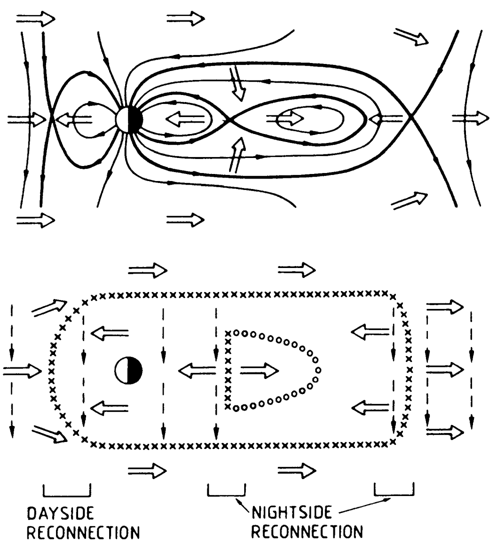


Fig. 3. Sketch of a possible topology of the magnetosphere (Vasyliunas, 1983), often called the reconnection model of the magnetosphere. The top view is in the noon-midnight meridian plane, showing magnetic field lines, separatrices (heavy lines), and plasma flow directions (open arrows). The bottom shows a projection onto a surface containing the magnetic singular lines or reconnection lines (the equatorial plane in a case of ideal symmetry), both X- and O-type. The plasma flow directions are in accord with an assumed electric field (dashed arrows) from dawn to dusk.

The big question is: does all this apply to the magnetosphere? Figure 3 is a sketch of a possible topology of the magnetosphere (after Vasyliunas, 1983); many would argue that this topology must hold, given that the solar wind flow is always present (Vasyliunas, 1983). The topologically different classes of field lines include (1) closed field lines with two feet on the ground, (2) open field lines, with only one foot on the ground while the other is connected to the interplanetary magnetic field (IMF), and (3) the IMF. This implies an X-line completely around the magnetosphere (easily understood in the case of a southward IMF). A fourth class of field lines might occur, (4) an isolated bubble or magnetic island where the field lines close on themselves. The general flow pattern shown by the open arrows requires an electric field that is everywhere from dawn to dusk, out of the plane of the upper diagram (see, e.g., Cowley, 1980, his Figure 1), a sense that is consistent with the electric field in the solar wind with a southward IMF.

Dayside reconnection, directly driven by the solar wind (or rather, by the shocked magnetosheath flow) would produce a jet of plasma toward open field lines, while nightside reconnection would result in some plasma within the plasma sheet flowing Earthward, with the rest flowing tailward on the far side of the X-line.

A common way of stating what is supposed to happen is that 'dayside reconnection causes anti-Sunward convection over the polar cap followed by nightside reconnection producing Sunward return flow at lower latitudes'. It should be noted that this is a tautology: E_r is required for reconnection, and reconnection provides E_r! Since both dayside reconnection and nightside reconnection are electrical loads, somewhere there must be a dynamo, usually said to be on the lobe magnetopause. It thus becomes necessary to look to the flowing plasma over the lobe magnetosphere to power both dayside and nightside reconnection. There are problems with energy transfer over the long distances in the model shown in Figure 3, especially in time-varying situations, something that is rarely (if ever) discussed.

There are several important observations about dayside reconnection. (1) There is a lack of evidence of plasma dissipation associated with the reconnection process (Heikkila, 1975); this fact is generally ignored nowadays (note, however, Paschmann *et al.*, 1985). Even more decisive is the observation that the plasma loses energy (not gains) in general (Eastman and Hones, 1979) in going across the magnetopause. Jets of plasma have been observed by ISEE-1 (Paschmann *et al.*, 1979; Sonnerup *et al.*, 1981), but discussion of them requires the force equations; if the energy dissipation predicted by reconnection is not observed, then we must look to another theory or model to explain the momentum changes. Energy considerations can cast a veto on the reconnection model. (2) The jets predicted by reconnection would be on open field lines, whereas there are closed field lines in the cleft (McDiarmid *et al.*, 1976). We now know that a low latitude boundary layer (LLBL) is almost always present just inside the magnetopause, at least partly on closed field lines (Eastman and Hones, 1979; Mitchell *et al.*, 1987), as is the entry layer (Haerendel *et al.*, 1978). (3) No mechanism has as yet been found for the high value of resistance needed to maintain the electric field E_r along the X-line (Coroniti, 1985). Finally (4) quite often (even when the IMF is southward favouring the reconnection process) there is no evidence that it actually occurs (Sonnerup *et al.*, 1981).

There is not much doubt that the field lines in the central polar cap are open (see, e.g., Paulikas, 1973), implying that the magnetic field topology associated with reconnection is correct, including X-lines. What is questionable is the topology of the electric field; a drastic revision is needed here (Heikkila, 1975, 1984a). The usual cartoons depict models with an electrostatic field; thus, e.g., Cowley (1980) stated in the caption to his Figure 1 that the electric field is everywhere in the dawn-dusk direction. Figure 4(a) is based on this constant electric field usually assumed in reconnection theories, and the flow of plasma (and flux) toward open field lines in dayside reconnection; this model suffers from all the criticisms discussed in the last paragraph. Figure 4(b) shows a reversal of the tangential component (with an implied curl and, hence, time dependence), which could apply to a transient and local process as in Figure 1(b); this can lead to plasma flow across the magnetopause (Heikkila, 1982, 1984b), thus the possibility of plasma entry into the LLBL. This model does not suffer from the drawbacks of the reconnection model in Figure 4(a).

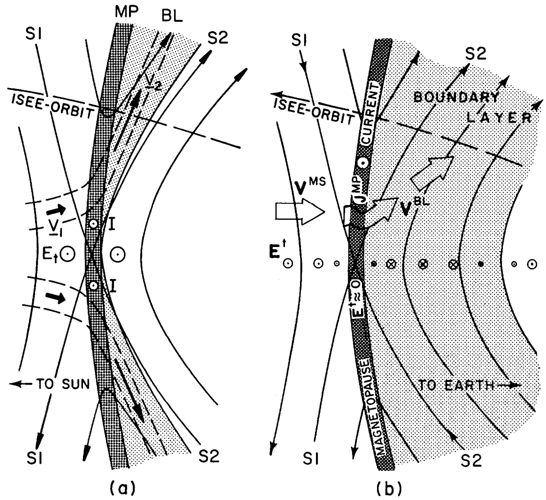


Fig. 4. Both parts show an open magnetosphere, with a normal component of the magnetic field in the magnetopause current layer, and a separatrix with two sheets S1 and S2. The reconnection model assumes an electric field along he X-line, as in (a) and, therefore, implies an outflow toward higher latitudes on open field lines. The model in (b) is based on the assumption that the plasma has sufficient momentum to enable it to create an ever-changing electric field to obey Newton's laws, perhaps being slightly deflected by the magnetic field. This model is much more complicated, since curl **E** may be non-zero, implying time-dependence in three dimensions.

There is also serious question about the applicability of the steady-state reconnection model of Figure 3 (neglect the magnetic island for the moment) to the distant magnetotail. Zwickl *et al.* (1984) reported that ISEE-3 showed only tailward flow beyond a distance of some 180 R_e over 97% of the time, and concluded that the X-line was rarely beyond 180 R_e. However, direct analysis of the magnetometer data was not as convincing, with the first publication showing a positive value for B_z out to a distance of 200 R_e, and slightly negative beyond that (Tsurutani *et al.*, 1984). Slavin *et al.* (1985) sorted the data in the y-direction as well, and found a positive sense in wide bands on either side; they concluded that the X-line was curved. However, they failed to point out that the flow was still tailward in the regions with positive B_z beyond 200 R_e, with the implication that the electric field was in the direction from dusk to dawn. Heikkila (1987) identified this as the low latitude boundary layer (LLBL), in a model of the magnetotail dominated by it (Heikkila, 1984a, b). The deduction from the ISEE-3 data is that the potential difference was some tens of kilovolts in the LLBL on both the morning and evening flanks at these distances. Similar values are found closer to the Earth (Foster, 1984; Heikkila, 1986), despite interpretations of good electric field data to the contrary (Mozer, 1984).

Recent observations have provided convincing evidence that the fourth topological feature is also present in the magnetotail as a transient feature. One event shown in Figure 5 is particularly outstanding; the polar angle shows a strong northward orientation of the magnetic field beginning at 08:20 UT, followed by a southward direction,

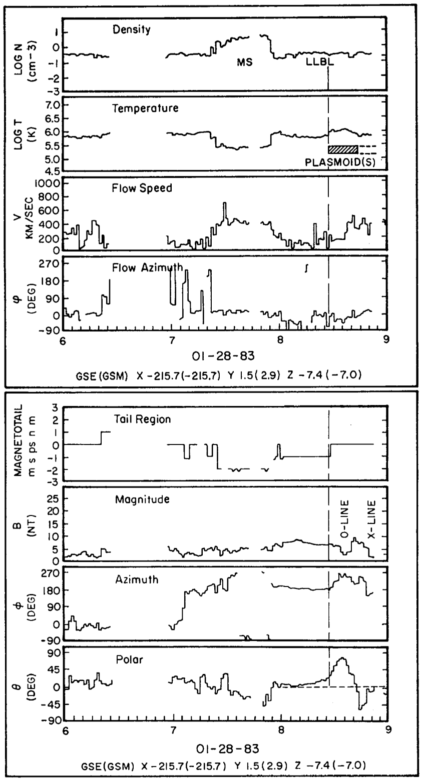


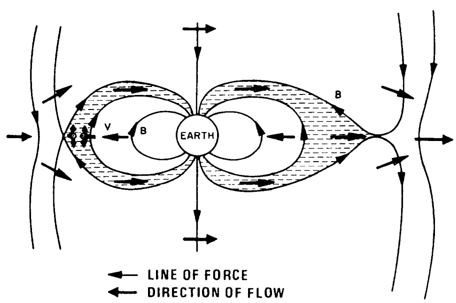
Fig. 5. ISEE-3 plasma and magnetic field data for a 3-hr period in the distant magnetotail. The record of the polar angle for the magnetic field indicates the presence of a plasmoid near 08 : 30 UT, verified by the plasma data. The magnitude of B indicates proximity to an O-line at 08 : 35, and an X-line at 08 : 50 UT. After leaving the magnetosheath at 07 : 55, the magnetic field had a northward component before the arrival of the plasmoid. This is consistent with the idea that the plasmoid had to burrow its way through closed field lines in its tailward escape from the magnetotail.

as would be consistent with a magnetic island passing by the spacecraft in the tailward direction. The plasma within the island is rather isotropic, somewhat heated, and flowing tailward. Such a feature is also known as a plasmoid. What is especially noteworthy is the northward orientation of the field in the interval before there was any indication of the approaching plasmoid (after the spacecraft left the magnetosheath at 07 : 55 UT). There is no evidence of the plasma jets in Figure 5 that would be associated with steady-state reconnection before the plasmoid appeared. This is fully consistent with the boundary-layer model of the magnetotail, with the X-line very far downstream (Heikkila, 1984b). The plasmoid must do work in its tailward flight through a region of closed field lines powered by means of a current connected to the polar cap (Iijima and Potemra, 1978), associated with recovery phase auroras in the polar cap.

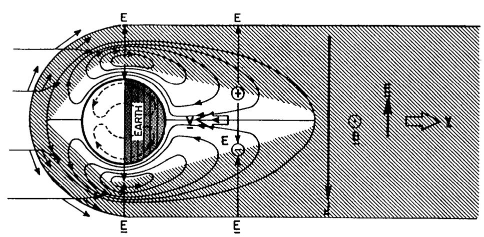
Thus there is a requirement for a process of transient reconnection (for the creation of a plasmoid); the proper theory must involve an electric field with a curl, since there are temporal changes in the magnetic field. A finite electromotive is necessary for access to energy stored in the magnetic field. It has been said that it is appropriate to begin theoretical work with a simple model at first, before we tackle the curl: we have to learn to walk first before we try to run. However, this comment overlooks the fact that $\text{curl}\,\mathbf{E} = -\partial\mathbf{B}/\partial t$ is an essential element of the physics. This omission is like trying to learn to swim with no water in the pool, so as to avoid the danger of drowning!

4. Discussion

The two ideas for the interaction between the solar wind and the magnetosphere were both proposed in 1961. One was the reconnection model (Figure 3) of Dungey (1961). The LLBL had not been discovered as yet; now that it has, we must modify Dungey's model as in Figure 6, showing tailward flow in the distant tail. By some process, largely unknown but perhaps as in Figure 1(b), plasma is able to cross the dayside magnetopause, and leave again far downstream. The other idea was proposed by Axford and Hines (1961), shown at the bottom of Figure 6. The LLBL has been added to their own figure also. With the addition of the LLBL, the two figures agree, but with the



(a)



(b)

Fig. 6. The magnetospheric boundary layer needs to be added to the classic models for the interaction of the solar wind with the magnetosphere to make each of them more physically realistic. This is done in (a) for the reconnection model introduced by Dungey, and in (b) for the viscous interaction model of Axford and Hines. This revision is required to take note of the tailward flow in the distant magnetotail, on closed field lines. The sense of this flow shows that the electric field is an agent of the plasma, allowing ti to obey Newton's laws. With this new emphasis, the two models are in fact complementary.

important conclusion that steady-state reconnection plays a very different role from what has come to be believed. This role apparently is its influence on the entry of solar wind plasma through the magnetopause; in this regard it should be noted that the processes of Lemaire (1977) and Heikkila (1982) do depend on the IMF. In the process depicted in Figure 1(b) the plasma particles enter by travelling parallel to the magnetic field B_n (normal to the current sheet), bringing their momentum and energy in *mechanical form*, described by *Newton's laws* rather than by Maxwell's equations. As such, Poynting's theorem is not applicable at the magnetopause to describe solar wind-magnetospheric coupling.

Figure 7 (Heikkila, 1979) is a diagram for the grand electric circuit for the magnetosphere, showing the dynamo in the boundary layer, and the various auroral loads. Circuit concepts such as this provide a clear distinction between cause and effect. Of course we still must look at force equations to understand the physics. In addition, open field lines in the central polar cap are associated with plasma convection across the polar cap (where there are no auroras). Plasma on open field lines can escape without doing much (or any) work. Nevertheless, it is these open flux tubes that the conventional wisdom insists does all the work of causing magnetospheric plasma convection and the creation of auroral phenomena ... *dubious principle frozen into dogma!*

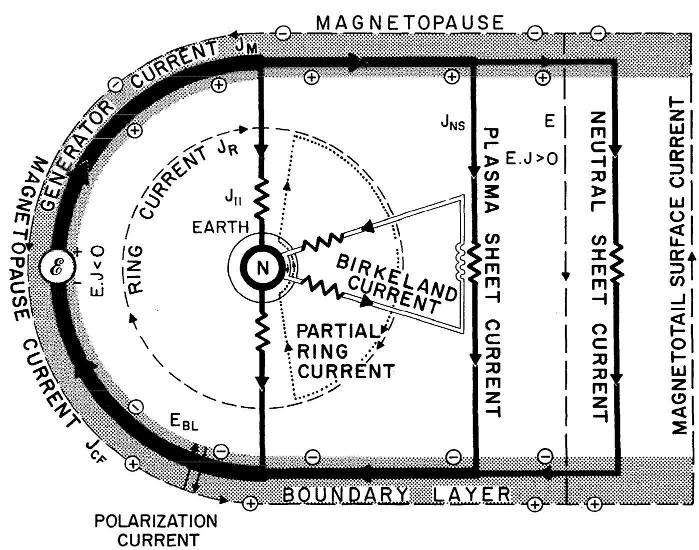


Fig. 7. The grand electric circuit of the magnetosphere. The tailward boundary layer flow forms a voltage generator, since the plasma can produce the necessary polarization current to deliver charge required for the correct electric field (to conserve plasma momentum). Inside the boundary-layer flow there is, in addition, a current generator, since there must be a $\mathbf{J} \times \mathbf{B}$ force so that the flow will be along the curved boundary parallel to the magnetopause. The electrical loads in the rest of circuit are related to various auroral features.

Acknowledgements

This work was supported by the National Science Foundation under grant ATM-8504269 and the Danish Meteorological Institute. I am thankful for the use of ISEE-3 data provided by Ed Smith and Sam Bame.

References

Alfvén, H.: 1950, *Cosmical Electrodynamics*, Clarendon Press, Oxford.
Alfvén, H.: 1981, *Cosmic Plasma*, D. Reidel Publ. Co., Dordrecht, Holland.
Alfvén, H.: 1986, TRITA-EPP-86–04, Royal Institute of Technology, Stockholm.
Alfvén, H. and Fälthammar, C.-G.: 1963, *Cosmical Electrodynamics*, Clarendon Press, Oxford.
Axford, I.: 1984, in E. W. Hones, Jr. (ed.), *Magnetic Reconnection in Space and Laboratory Plasmas*, AGU Monograph 30, p. 1.
Axford, I. and Hines, C. O.: 1961, *Can. J. Phys.* **39**, 1433.
Coroniti, F. W.: 1985, *Space Sci. Rev.* **42**, 399.
Cowley, S. W. H.: 1980, *Space Sci. Rev.* **26**, 217.
Dungey, J. W.: 1961, *Phys. Rev. Letters* **6**, 47.
Eastman, T. E. and Hones, E. W., Jr.: 1979, *J. Geophys. Res.* **84**, 2019.
Foster, J. C.: 1984, *J. Geophys. Res.* **89**, 855.
Haerendel, G., Paschmann, G., Sckopke, N., Rosenbauer, H., and Hedgecock, P. C.: 1978, *J. Geophys. Res.* **83**, 3195.
Heikkila, W. J.: 1975, *Geophys. Res. Letters* **2**, 154.
Heikkila, W. J.: 1979, in W. P. Olson (ed.), *Quantitative Magnetospheric Models*, AGU Geophysical Monograph 21, p. 389.
Heikkila, W. J.: 1982, *Geophys. Res. Letters* **9**, 159.
Heikkila, W. J.: 1984a, in E. W. Hones, Jr. (ed.), *Magnetic Reconnection in Space and Laboratory Plasmas*, AGU Geophysical Monograph 30, p. 39.
Heikkila, W. J.: 1984b, in T. Potemra (ed.), *Magnetospheric Currents*, AGU Geophysical Monograph 28, p. 208.
Heikkila, W. J.: 1986, *Geophys. Res. Letters* **131**, 233.
Heikkila, W. J.: 1987, in A. T. Y. Lui (ed.), *Magnetotail Physics*, Johns Hopkins University Press, p. 65.
Heikkila, W. J., Pellinen, R. J., Fälthammar, C.-G., and Block, L. P.: 1979, *Planetary Space Sci.* **27**, 1383.
Iijima and Potemra, T.: 1978, *J. Geophys. Res.* **83**, 599.
Jordan, E. C. and Balmain, K. G.: 1968, *Electromagnetic Waves and Radiating Systems*, Prentice Hall, New York.
Kulsrud, R. M.: 1983, *Phys. Today* **36**, 56.
Lemaire, J.: 1977, *Planetary Space Sci.* **25**, 887.
Mitchell, D. G., Kutchko, F., Williams, D. J., Eastman, T. E., Frank, L. A., and Russell, C. T.: 1987, *J. Geophys. Res.* **92**, 7394.
Mozer, F. S.: 1984, *Geophys. Res. Letters* **11**, 135.
Paschmann, G., Papamastorakis, I., Sckopke, N., Sonnerup, B. U. Ö., Bame, S. J., and Russell, C. T.: 1985, *J. Geophys. Res.* **90**, 12111.
Paschmann, G., Sonnerup, B. U. Ö., Papamastorakis, I., Sckopke, N., Haerendel, G., Bame, S. J., Asbridge, J. R., Gosling, J. T., Russell, C. T., and Elphic, R. C.: 1979, *Nature* **282**, 243.
Paulikas, G. A.: 1973, *Rev. Geophys. Space Phys.* **12**, 117.
Plonsey, R. and Collin, R. E.: 1961, *Principles and Applications of Electromagnetic Fields*, McGraw-Hill, New York.
Ramo, S. and Whinnery, J. R.: 1953, *Fields and Waves in Modern Radio,* John Wiley & Sons, New York.
Slavin, J. A., Smith, E. J., Sibeck, D. G., Baker, D. N., Zwickl, R. D., and Akasofu, S. I.: 1985, *J. Geophys. Res.* **90**, 10875.
Sonnerup, B. U. Ö.: 1985, in R. Kundu and G. Holman (eds.), *Unstable Current Systems and Plasma Instabilities in Astrophysics*, D. Reidel Publ. Co., Dordrecht, Holland.
Sonnerup, B. U. Ö., Paschmann, G., Papamastorakis, I., Sckopke, N., Haerendel, G., Bame, S. J., Asbridge, J. R., Gosling, J. T., and Russell, C. T.: 1981, *J. Geophys. Res.* **86**, 10049.

Tsurutani, B. T., Jones, D. E., Slavin, J. A., Sibeck, D. G., and Smith, J. E.: *Geophys. Res. Letters* **11**, 1062.
Vasyliunas, V. M.: 1968, *J. Geophys. Res.* **73**, 5805.
Vasyliunas, V. M.: 1975, *Rev. Geophys. Space Phys.* **13**, 303.
Vasyliunas, V. M.: 1983, in R. L. Carovillano and J. M. Forbes (eds.), *Solar-Terrestrial Physics*, D. Reidel Publ. Co., Dordrecht, Holland, p. 255.
Vasylianus, V. M.: 1984, in T. Potemra (ed.), *Magnetospheric Currents*, AGU Monograph 28, p. 63.
Zwickl, R. D., Baker, D. N., Bame, S. J., Feldman, W. C., Gosling, J. T., Hones, E. W., Jr., McComas, D. J., Tsurutani, B. T., and Slavin, J. A.: 1984, *J. Geophys. Res.* **89**, 11007.

THE DIELECTRIC ANALOGUE OF MAGNETOHYDRODYNAMICS*

BIBHAS R. DE

Chevron Oil Field Research Company, La Habra, California, U.S.A.

(Received 18 September, 1987)

Abstract. Almost half a century after Alfvén first conceived of the science of magnetohydrodynamics, it is still possible to trace his intuitive thinking to explore physical processes heretofore not considered. The ideas of magnetohydrodynamics (applicable to conducting fluids) can be transferred almost intact to purely dielectric fluids, such that we can arrive at a generalized concept applicable to *any* fluid – conducting or dielectric. In this sense, Alfvén's conception of magnetohydrodynamics may be ideationally even more profound than it has been thought to be so far.

1. Historical Background

The essential ideas of electrodynamic interaction in matter had been formulated by the turn of the century by André Marie Ampère, James Clerk Maxwell, Hendrik Antoon Lorentz, and others. A significant ramification of these, however, was proposed by Hannes Alfvén in 1942 when he predicted a form of wave behaviour in a magnetized conducting fluid (Alfvén, 1942; Alfvén and Fälthammar, 1963) that later came to be known by his name. The principle underlying this phenomenon later formed the basis of multifarious developments in plasma physics and space physics. It was Alfvén's now legendary scientific intuition that led him to combine Maxwell's curl equations with Lorentz' force law ('a current-carrying conductor in a magnetic field experiences a force') and Newton's second law ('force equals mass times the acceleration') to deduce the science of magnetohydrodynamics – interactions in a magnetized *conducting* fluid involving the interchange of magnetic and kinetic energies. In the light of this same intuition, it is instructive to explore the Ampère–Maxwell–Lorentz–Alfvén connection to seek its lessons. In so doing, however, let us also invoke a simple aesthetic criterion: that of completeness.

A prefatory comment is appropriate at the outset of this discussion which is concerned with physical processes in an idealized situation. We discuss Alfvén's concept of magnetohydrodynamics only in that context – in its 'textbook' sense. It is now well-known that Alfvén has for some time sought to de-emphasize the frozen flow concept and taught against its indiscriminate application – especially in the case of tenuous plasmas in space (see, e.g., Alfvén, 1981). The present discussion of another type of frozen flow in another, very different, medium is not intended to detract from that effort and that teaching.

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 99–104.

2. The Magnetohydroelectric Effect

As is widely recognized now, the greatness of Maxwell's intuition was rooted in his conception of displacement current (free-space displacement current + polarization current in dielectrics) that completed the troubled Ampère's law (conceived with only conductors in mind), and provided an understanding of electromagnetic waves. Lorentz, in adding his force law to these results, again considered only conductors and did not say anything about dielectrics. And Alfvén then confined himself entirely to the conductors, leaving unaddressed the dielectrics. We thus begin to sense a certain incompleteness in the story of development of ideas, starting with Lorentz. If we postulate a 'Lorentz force' in a dielectric, we can restore completeness to this development.

The basic suggestion of the magnetic force on a pure dielectric material carrying a pure polarization current follows from simple, straightforward arguments. Let us recall first the derivation of the Lorentz force in the case of a conductor. A *conduction* current consists of a flow of electrons (each having a velocity $\mathbf{v}_i$ and a charge q, say) under the influence of an applied electric field $\mathbf{E}$. The total current density is $\mathbf{J} = \Sigma q\mathbf{v}_i$, where the summation is taken over all the electrons in a unit volume. Each moving electron experiences a force $\mathbf{f}_i = q\mathbf{v}_i \times \mathbf{B}$ in the presence of a magnetic field $\mathbf{B}$. These individual forces on all the electrons are transmitted to the body of the conductor through collisions with atoms. Thus the total force on a unit volume is $\mathbf{F} = \Sigma q\mathbf{v}_i \times \mathbf{B} = (\Sigma q\mathbf{v}_i) \times \mathbf{B} = \mathbf{J} \times \mathbf{B}$.

No such flow of charges occurs in a dielectric carrying a pure *polarization* current, and one is thus apt not to think in terms of the presence of a similar force. We note, however, that there is here nevertheless a microscopic displacement of the positive and the negative charges bound in the atoms and molecules of the dielectric, and these individual charges are subject to the same Lorentz force as the free electrons in a conductor; the positive charges q^+ and the negative charges q^- move in opposite directions under the influence of a time-varying field $\mathbf{E}$, with velocities $\mathbf{v}_i^+$ and $\mathbf{v}_i^-$, respectively. Thus the net *polarization* current is $\mathbf{J} = \Sigma(q^+\mathbf{v}_i^+ + q^-\mathbf{v}_i^-)$. The force on an individual positive charge is $\mathbf{f}_i^+ = q^+\mathbf{v}_i^+ \times \mathbf{B}$ and on a negative charge, $\mathbf{f}_i^- = q^-\mathbf{v}_i^- \times \mathbf{B}$. Clearly, they both point in the same direction, giving a net force on each individual atom or molecule. The total volume force is again formally given by $\mathbf{F} = \Sigma(\mathbf{f}_i^+ + \mathbf{f}_i^-) = \Sigma(q^+\mathbf{v}_i^+ + q^-\mathbf{v}_i^-) \times \mathbf{B} = \mathbf{J} \times \mathbf{B}$. From these simple arguments, we are now able to make an important generalization: *All substances (conductors and dielectrics) experience the* $\mathbf{J} \times \mathbf{B}$ *force in a magnetic field.* This provides the full complement of the force law to the Maxwell's equations.

We will now take the next logical step, and attempt to provide the dielectric counterpart of Alfvén's ideas.

Consider for simplicity a pure dielectric fluid (nonconducting, lossless) with a polarizability χ and a dielectric constant $\varepsilon = (1 + \chi)\varepsilon_0$ placed in a magnetic field $\mathbf{B}$, and moving with a velocity $\mathbf{u}$ (assumed nonrelativistic). Then the electric field $\mathbf{E}'$ in the body of the moving fluid is related to the field $\mathbf{E}$ in the laboratory (or rest) frame by (see, e.g., Stratton, 1941; Section 1.23)

$$\mathbf{E}' = \mathbf{E} + \mathbf{u} \times \mathbf{B}, \tag{1}$$

so that the polarization current density in the body of the moving fluid is

$$\mathbf{J}' = \chi\varepsilon_0\dot{\mathbf{E}}'. \tag{2}$$

The current density in the rest frame may now be written as (*op. cit.*)

$$\mathbf{J} = \mathbf{J}' + \rho_p\mathbf{u}, \tag{3}$$

where ρ_p is the polarization space charge density in the fluid. This is the density of the induced polarization charges that arise in the body of the moving fluid, any free charges arising on a rigid bounding conductor. We now have

$$\mathbf{J} = \chi\varepsilon_0\left[\dot{\mathbf{E}} + \frac{\partial}{\partial t}(\mathbf{u} \times \mathbf{B})\right] + \rho_p\mathbf{u}. \tag{4}$$

This may be recognized as the dielectric equivalent of the 'Generalized Ohm's Law' of magnetohydrodynamics.

The volume force on the fluid consists of the electromagnetic $\mathbf{J} \times \mathbf{B}$ force, and the electrostatic force $\rho_p\mathbf{E}$ (cf. Stratton, 1941; Section 2.21; we assume here that the fluid is incompressible). Hence, the force balance equation (Newton's second law) for the fluid is

$$\frac{\partial}{\partial t}(\rho\mathbf{u}) = \mathbf{J} \times \mathbf{B} + \rho_p\mathbf{E}, \tag{5}$$

where ρ is the mass density of the fluid. Other force terms due to gravity, pressure gradient, etc., are possible. The above relation is the fundamental force equation of magnetohydrodynamic interaction *in a dielectric fluid*. It involves an interchange of magnetic, fluid-kinetic *and* electrostatic energies. For this reason, the term 'magneto-hydroelectric interaction' was proposed to describe the phenomenon (De, 1979a,b, 1980).

3. The Equation of Magnetohydroelectrics: Magnetic Field Freezing and Dielectric 'Alfvén Waves'

In order to develop our discussion further, it is necessary to obtain a relationship between the bound polarization charge density ρ_p and the electric field $\mathbf{E}$. Consideration of the charge build-up on an elementary capacitor will show that

$$\nabla \cdot \chi\varepsilon_0\mathbf{E} = -\rho_p. \tag{6}$$

Recall now the Maxwell's equations

$$\nabla \cdot \mathbf{B} = 0, \tag{7}$$

$$\nabla \times \mathbf{B} = \mu\mathbf{J} + \mu\varepsilon_0\dot{\mathbf{E}}, \tag{8}$$

$$\nabla \times \mathbf{E} = -\dot{\mathbf{B}}; \tag{9}$$

where μ is the magnetic permeability of the fluid. Using Equations (7), (8), and (9) in Equation (4), we are now able to derive

$$\ddot{\mathbf{B}} = \frac{1}{\mu\varepsilon}\nabla\cdot\mathbf{B} + \frac{\chi}{1+\chi}\frac{\partial}{\partial t}\nabla\times(\mathbf{u}\times\mathbf{B}) + \frac{1}{\varepsilon}\nabla\times(\rho_p\mathbf{u}). \tag{10}$$

This equations contains all the essence of the magnetohydroelectric interaction when taken in conjunction with Equation (5). While this latter equation does not enter directly into the derivation of Equation (10), it determines whether or not the magnetohydroelectric effect is significant (Equation (10) by itself could be satisfied for arbitrarily small values of **B**; see De, 1979b).

When the first term on the right-hand side of Equation (10) dominates, we have

$$\ddot{\mathbf{B}} = \frac{1}{\mu\varepsilon}\nabla^2\mathbf{B}, \tag{11}$$

the familiar equation for three-dimensional electromagnetic wave propagation in a dielectric. When the second term dominates and $\chi \gg 1$, we arrive at the well-known condition for frozen flow (cf. Alfvén and Fälthammar, 1963)

$$\ddot{\mathbf{B}} = \frac{\partial}{\partial t}\nabla\times(\mathbf{u}\times\mathbf{B}). \tag{12}$$

Thus, in a medium radically different from what Alfvén was concerned with, we have arrived at the same physical condition he had envisioned. In this state of frozen flow, it is also possible to deduce an Alfvén wave-like wave behaviour (the 'magnetohydroelectric wave'; see De, 1979a). Such waves propagate along magnetic field lines with a velocity v given by

$$v^2 = c^2\left(\frac{1}{\chi} + \frac{v_A^2}{c^2}\right)\left(1 + \frac{1}{\chi} + \frac{v_A^2}{c^2}\right)^{-1}, \tag{13}$$

where c is the velocity of light in free space, and $v_A = B/(\mu\rho)^{1/2}$ is again a familiar parameter that makes its appearance: the Alfvén velocity. It has further been shown (*op. cit.*) that a fully generalized wave behaviour can be derived in an arbitrary medium which is partly conducting and partly dielectric, and that in various appropriate limits this wave reduces to the ordinary electromagnetic wave, the Alfvén wave and the magnetohydroelectric wave.

4. Magnetic Flux Amplification and Electric Field Freezing

Our discussion so far has developed in close parallelism with conventional magnetohydrodynamics. We now wish to venture somewhat far afield to explore if anything more can be gleaned from our formulation thus far. When the last term on the right-hand side

Equation (10) dominates, there arises a state described by

$$\ddot{\mathbf{B}} = \frac{1}{\varepsilon} \nabla \times (\rho_p \mathbf{u}). \tag{14}$$

The induced magnetic field may now be perpendicular to the fluid motion, and parallel to the original static magnetic field. This is *not* the state of frozen-in magnetic field lines; rather, it indicates an amplification of the magnetic flux resulting from an exchange of energy among the three fields (magnetic, electric, and velocity). A magnetic flux tube here may be imagined to be constricted. As we shall see below, this is in fact a state of frozen-in electric field lines (a concept that would be meaningless in the case of a perfectly conducting fluid).

Upon using Equations (6) and (9) in the above equation, we obtain

$$\dot{\mathbf{E}} = -(\nabla \cdot \mathbf{E})\mathbf{u}, \tag{15}$$

$$\dot{\mathbf{E}} = \frac{1}{\varepsilon} \rho_p \mathbf{u}; \tag{16}$$

the implications of which are immediately obvious: the change in the electric field $\mathbf{E}$ equals the divergence of the field times the fluid velocity – i.e., it owes itself to a movement of the polarization space charges along with the fluid. The electric field lines are ‘tied’ to these space charges and move along with them.

5. Remarks

From the above discussion, it follows simply that one could erect a generalized formalism for magnethydrodynamic interaction in a fluid of generalized property (conducting *and* dielectric). In the limit of infinite conductivity, such a formalism would lead to the state of frozen-in magnetic field lines; in the other limit, that of infinitely high dielectric constant, a state of frozen-in electric field lines is possible. In this sense, Alfvén’s conception of magnetohydrodynamics may be ideationally even more profound in its scope than it has been thought to be so far. The physical realms of manifestation of the conducting and the dielectric effects, however, differ greatly. Frozen flow in conducting fluids is favoured for low frequencies and large length scales; in dielectric fluids quite the opposite is true. Thus the realms of applicability are also very different. Whereas the former effect has found application in space science and in large scale devices in the industry, the latter effect – whose applications, if any, may well lie far into the future – will conceivably apply to experiments and devices involving motions at microscopically small physical length scales.

Acknowledgement

I wish to thank Professor William B. Thompson of the University of California at San Diego for helpful comments leading to a clearer presentation of the ideas in this paper.

References

Alfvén, H.: 1942, *Nature* **150**, 405.
Alfvén, H.: 1981, *Cosmic Plasma*, D. Reidel Publ. Co., Dordrecht, Holland.
Alfvén, H. and Fälthammar, C.-G.: 1963, *Cosmical Electrodynamics*, Oxford University Press, London.
De, B. R.: 1979a, *Phys. Fluids* **22** (**1**), 189.
De, B. R.: 1979b, *Astrophys. Space Sci.* **62**, 255.
De, B. R.: 1980, *Phys. Fluids* **23** (**2**), 408.
Stratton, J. A.: 1941, *Electromagnetic Theory*, McGraw-Hill, New York.

MAGNETOSPHERE-IONOSPHERE INTERACTIONS – NEAR-EARTH MANIFESTATIONS OF THE PLASMA UNIVERSE*†

CARL-GUNNE FÄLTHAMMAR

Department of Plasma Physics, The Royal Institute of Technology, Stockholm, Sweden

(Received 2 July, 1987)

Abstract. As the Universe consists almost entirely of plasma, the understanding of astrophysical phenomena must depend critically on our understanding of how matter behaves in the plasma state. *In situ* observations in the near-Earth cosmical plasma offer an excellent opportunity of gaining such understanding. The near-Earth cosmical plasma not only covers vast ranges of density and temperature, but is the site of a rich variety of complex plasma physical processes which are activated as a result of the interactions between the magnetosphere and the ionosphere.

The geomagnetic field connects the ionosphere, tied by friction to the Earth, and the magnetosphere, dynamically coupled to the solar wind. This causes an exchange of energy and momentum between the two regions. The exchange is executed by magnetic-field aligned electric currents, the so-called Birkeland currents. Both directly and indirectly (through instabilities and particle acceleration) these also lead to an exchange of plasma, which is selective and therefore causes chemical separation. Another essential aspect of the coupling is the role of electric fields, especially magnetic-field aligned ('parallel') electric fields, which have important consequences both for the dynamics of the coupling and, especially, for energization of charged particles.

1. Introduction

Ionized matter, plasma, is the overwhelmingly dominating constituent of the Universe as a whole – the plasma universe (Alfvén 1986a, b). Matter in the plasma state is characterized by a complexity that vastly exceeds that exhibited in the solid, liquid and gaseous states. In fact, the physical, and especially the electrodynamical, properties of plasma are still far from well understood.

These properties are still subject to basic research, and many fundamental questions remain to be answered. However, important progress has been made recently as a result of experiments in the *laboratory* and in those regions of *space* accessible to *in situ* observations and experimentation.

While it is universally acknowledged that our Universe is a plasma universe, it seems to be far from fully realized that the physical understanding of this Universe depends critically on our understanding of matter in the plasma state. In fact, the recent progress

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 105–133.

in plasma physics should provide a much improved foundation for understanding astrophysical processes in the Universe of the present – as well as cosmogonic processes of the past (Alfvén 1942, 1981, 1984; Alfvén and Arrhenius, 1976). So far the increasing insight into the behaviour of matter in the plasma state has not been widely applied to astrophysics. To make full use of this insight should be a very important step toward a better understanding of our plasma universe.

A brief look at the evolution of plasma physics is useful in establishing an appropriate perspective.

Early plasma experiments were limited essentially to cool and/or weakly ionized plasmas. They formed the limited empirical basis on which the *classical plasma theory* was built. This theory was developed into a high degree of mathematical sophistication and was believed to have general validity. One of the predictions based on it was that magnetic confinement of plasma should be rather easy, and thermonuclear fusion possible within 15 years.

When the thermonuclear effort made it possible to produce and study hot and highly ionized plasma in the laboratory, it was found that the plasma exhibited many kinds of unpredicted, 'anomalous' behaviour. The 'thermonuclear crisis' that resulted led to the start of a new epoch in thermonuclear research, characterized by a close interplay between experimental and theoretical research. This has led to impressive progress in solving plasma physical problems that are vastly more complex than envisaged by classical plasma theory.

Similarly, in space research it was widely believed that the cosmical plasma would have negligible resistivity, as predicted by classical formulas, and behave essentially as an ideal MHD medium. If so, the electric field would be a secondary parameter of little importance, and magnetic-field-aligned ('parallel') electric fields out of the question. As a consequence, the electric field and especially the magnetic-field-aligned electric field, which we now know to be of crucial importance, were long disregarded. Even to this day, only very few direct measurements of electric fields have been made in the outer magnetosphere.

The magnetosphere was universally assumed to be populated by a hydrogen plasma from the solar wind, whereas we now know that it is sometimes dominated by oxygen plasma originating in the Earth's own atmosphere. As a result of generally accepted theories, one did not until relatively recently make the appropriate measurements of outflowing ions and of magnetospheric plasma composition. Much of this delay could have been avoided, if results already known from laboratory plasma experiments had been applied to the space plasma. In fact, on this basis Hannes Alfvén proposed parallel electric fields as an accelerating mechanism for auroral primaries already in 1958, but the idea was almost universally refuted as incompatible with classical theory.

It is a sobering fact that even after hundreds of satellites had circled the Earth, the generally accepted picture of our space environment was still fundamentally wrong in aspects as basic as the existence and role of electric fields and even the origin and chemical composition of the near-Earth plasma itself. In the light of this, how can we believe in detailed theoretical models of distant astrophysical objects, until we have learned – and applied to astrophysics – the lessons of how the real plasma behaves in the Earth's own magnetosphere.

2. The Magnetosphere-Ionosphere System

The Earth's ionosphere and magnetosphere constitute a cosmical plasma system that is readily available for extensive and detailed *in situ* observation and even active experimentation. Its usefulness as a source of understanding of cosmical plasmas is enhanced by the fact that it contains a rich variety of plasma populations with densities ranging from more than 10^{12} m^{-3} to less than 10^{4} m^{-3} and temperatures from about 10^{3} K to more than 10^{7} K (equivalent temperature). Even more importantly, this neighbourhood cosmical plasma is the site of numerous and complex plasma physical processes which for example lead to particle acceleration and chemical separation. The understanding of these processes should be essential also to the understanding of remotely observed astrophysical phenomena that take place in plasmas that will remain out of reach of *in situ* observation (Fälthammar *et al.*, 1978; Haerendel, 1980, 1981; Fälthammar, 1985). For example, one of the outstanding characteristics of cosmical plasmas is their ability to efficiently accelerate charged particles. Many kinds of particle acceleration take place in the near-Earth plasmas, and this allows us to study in detail the mechanisms responsible.

A basic reason why the near-Earth plasmas are so active in terms of plasma physical processes is the coupling that the geomagnetic field imposes between the hot thin magnetospheric plasma, which is dynamically coupled to the solar wind and the cool, dense ionospheric plasma, which is tied by friction to the Earth (Vasyliunas, 1972; Greenwald, 1982).

This situation causes an exchange of momentum and energy between the two regions. The exchange is executed through electric currents – the Birkeland currents – that flow between them. Both directly and indirectly (through the instabilities and acceleration that they cause) the Birkeland currents also lead to an exchange of matter between the magnetosphere and the ionosphere. The exchange of matter is selective, so that the chemical composition of the ionospheric plasma that populates the magnetosphere is very different from that in its region of origin. The very efficient chemical separation that has unexpectedly been discovered in the near-Earth plasma, and is accessible to *in situ* investigation there, should also be of considerable astrophysical interest.

The present paper will concentrate on some crucial aspects of the magnetosphere-ionosphere system, namely the electric fields and currents and their role in particle acceleration, plasma transport and chemical separation.

3. Birkeland Currents

A phenomenon of paramount importance for the coupling between the magnetosphere and the ionosphere is that of Birkeland currents. In addition to being prime agents for exchange of *momentum* and *energy* between the two regions they also play an important role in redistributing *matter* between them. The energy coupling between the magnetosphere and ionosphere by means of the Birkeland currents has been discussed by Sugiura (1984).

A reason why the Birkeland currents are particularly interesting is that, in the plasma forced to carry them, they cause a number of plasma physical processes to occur (waves, instabilities, double-layer formation). These in turn lead to consequences such as acceleration of charged particles, both positive and negative, and chemical separation (such as the preferential ejection of oxygen ions). Both these classes of phenomena should have a general astrophysical interest far beyond that of understanding the space environment of our own Earth (Alfvén 1977, 1986c).

3.1. The distribution of Birkeland currents

Although predicted by pioneers like Birkeland and Alfvén the existence of electric currents connecting the magnetosphere and ionosphere apparently came as a surprise to many. In fact the first measurements revealing the magnetic effects of what we now call Birkeland currents were initially interpreted as standing Alfvén waves above the auroral zone before a correct interpretation was reached (see Dessler, 1984).

Since then the Birkeland currents have been investigated by means of many satellites. Their general large scale distribution at low altitude (Figure 1) is well established and described in an extensive literature. Much of the knowledge as of 1983 is summarized in the AGU Geophysical Monograph 28 edited by Potemra (1984). A concise review of field-aligned as well as ionospheric current systems was given by Baumjohann (1983). More recent papers on the distribution of Birkeland currents are those of Potemra *et al.* (1984, 1987); Zanetti *et al.* (1984); Araki *et al.* (1984); Iijima *et al.* (1984); Potemra and Zanetti (1985); Hruška (1986); Smits *et al.* (1986); Bythrow *et al.* (1986, 1987); Barfield *et al.* (1986); Kelly *et al.* (1986); Primdahl and Marklund (1986); Smits *et al.* (1986); Erlandson *et al.* (1987). A summary of magnetospheric currents as well as electric fields with comprehensive references to recent literature has been given by Mauk and Zanetti (1986). A new Birkeland current system, flowing in and out of the polar cap and intensifying during periods of northward interplanetary magnetic field has been described by Iijima *et al.* (1984). It is referred to as the NBZ system (for Northward B_z). Birkeland currents in the polar cusp have a pronounced dependence on the y-component of the interplanetary magnetic field. These currents may reflect the most direct coupling between the solar wind generator into the ionosphere (Potemra *et al.*, 1984; Potemra and Zanetti, 1985; Clauer *et al.*, 1984; Clauer and Kamide, 1985; Zanetti and Potemra, 1986).

The dependence of Birkeland currents and plasma convection patterns on B_y have been investigated by means of Dynamics Explorer data both for southward and northward B_z (Burch *et al.*, 1985; Reiff and Burch, 1985).

In theoretical studies Kan *et al.* (1984) and Marklund *et al.* (1985) have investigated the role of partial blocking of secondary Birkeland currents in causing the rotation of the ionospheric electric field pattern observed during substorms. The degree of closure of secondary Birkeland currents (associated with gradients in the height integrated

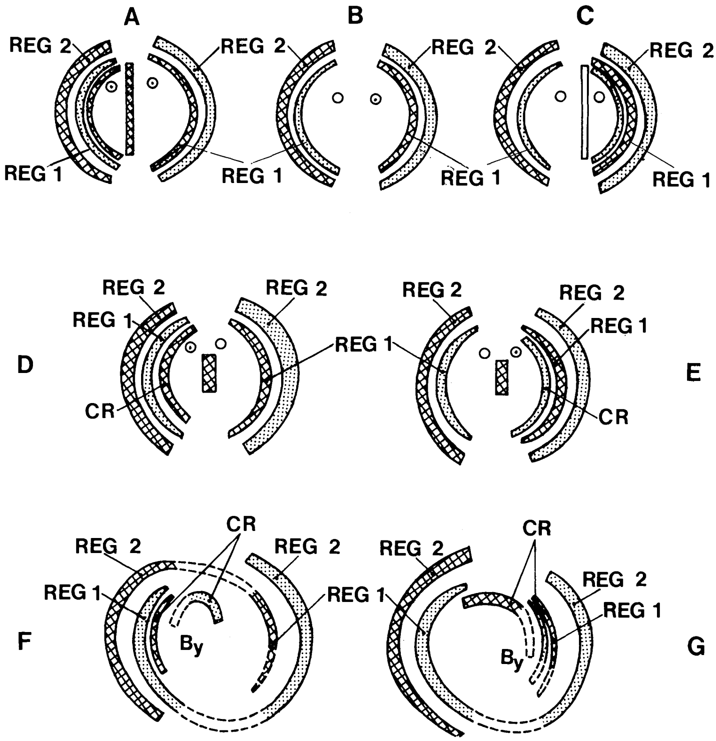


Fig. 1. Schematic Birkeland current patterns according to Reiff and Burch (1985) for various orientations of the interplanetary magnetic field. The upper row is for strongly northward B_z and B_y going from positive (A) through zero (B) to negative (C). The lower rows are for B_z weakly northward (middle row) and southward (bottom row). Patterns to the left are for positive B_y and to the right for negative B_y.

ionospheric Pedersen conductivity) also plays a key role in a new model of the westward travelling surge developed by Rothwell *et al.* (1984).

Within the large-scale Birkeland currents there exist fine structures in the form of thin current sheets with extreme current densities. A case reported by Burke *et al.* (1983) and Burke (1984) is shown in Figure 2. The sharp downward and upward slopes of the narrow dip in the magnetic field component B_y correspond to a pair of thin upward and downward Birkeland current sheets. The upward current sheet had a latitudinal extent of less than 2 km and an average current density of 135×10^{-6} A m^{-2}. In the downward current sheet of the same event the current density was 15×10^{-6} A m^{-2}. The authors note that the upward currents were carried by electrons that appeared to have fallen through a potential drop of a few kV. Also, the observed electron population, the relation between current density and accelerating voltage nearly (but not quite) agreed with adiabatic motion in the mirror field. (The voltage or the source plasma density or both would need to be a little higher than estimated.) The measured electron temperature, about 200 eV, did not indicate any substantial heating, which would be expected

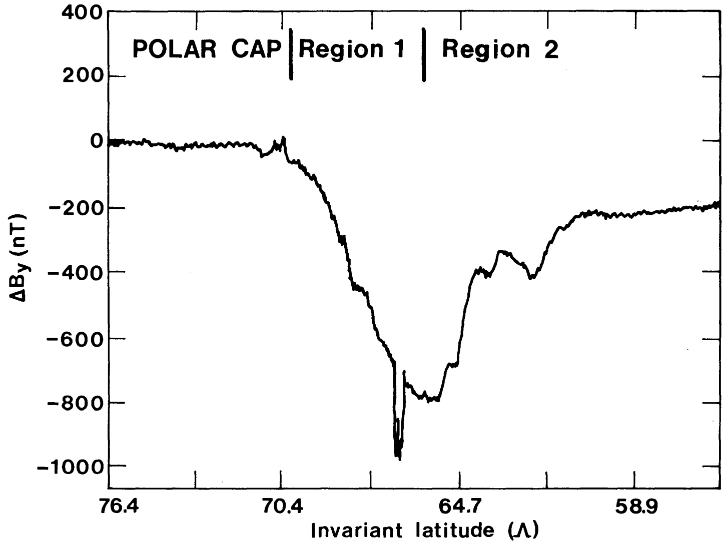


Fig. 2. Magnetic field signature of a pair of thin, high-density current sheets within the evening region 1, Birkeland current (Burke, 1984). The steep gradients on either side of the narrow dip beginning at 11 : 48 : 16 UT correspond to outward and inward Birkeland current sheets of 135×10^{-6} A m^{-2} and 15×10^{-6} A m^{-2}, respectively.

if anomalous resistivity played a major role. For the downward currents, which could readily be carried by upflowing cold ionospheric electrons, conditions at the 1000 km satellite altitude are close to the limit for ion-cyclotron instability.

Bythrow *et al.* (1984) have reported very high current densities – up to 94×10^{-6} A m^{-2} – also in earthward currents (measured by HILAT). From the current to the plates of the ion drift meter the authors estimated an ion number density of 2×10^{10} m^{-3} and, hence, a magnetic-field-aligned drift velocity of 30 km s^{-1} for the electrons carrying the Birkeland current. They concluded that this should be enough to destabilize electrostatic ion-acoustic waves as well as electrostatic ion-cyclotron waves. Simultaneous measurements of electron fluxes indicated that 2–4 km equatorward of this Birkeland current the height-integrated Pedersen conductivity had a sharp gradient (2 Ω^{-1} km^{-1}), which in combination with a prevailing northward electric field could be the cause of the observed Birkeland current.

Small-scale current structures have been observed not only near the ionosphere but even in the equatorial region. Thus it has been shown by Robert *et al.* (1984) that most of the SIP's (short irregular pulsations) observed at GEOS-2 are in fact the magnetic signatures of localized current structures passing by the spacecraft at a high velocity. The structures are estimated to have a current density of 6×10^{-9}–3×10^{-7} A m^{-2}, a size of 20–900 km and to move at a velocity of 15–170 km s^{-1}. They are associated with large electric field spikes (3–25 mV m^{-1}).

Recent observations of substorm-associated magnetic-field-aligned currents at the geosynchronous orbit have been reported by Nagai *et al.* (1987) and Nagai (1987).

3.2. Driving electromotive forces

One may distinguish between two kinds of electromotive forces that can drive Birkeland current. One is the MHD dynamo action of the bulk motion of plasma in the solar wind, plasma sheath and outer magnetosphere. The other is due to charge separation generated by differential drift of charged particles (gradient and curvature drift). This kind of generators draws on the kinetic or thermal energy of individual particles and may be characterized as *thermoelectric*. Both these kinds of generators are likely to be important in the magnetosphere (see, e.g., Block 1984; Vasyliunas, 1984).

An *internal* source of MHD dynamo action is the forced rotation of the ionosphere. (To this average contribution is added the dynamo action of ionospheric winds.) The corotational dynamo has an e.m.f. of nearly 100 kV (the equator being negative and both poles positive). However, most of this e.m.f. connects to low- and mid-latitude plasmas. These have a low ohmic resistance and a small enough moment of inertia that they are easily forced to corotate. Thus the net e.m.f. of the circuit and, hence, the Birkeland currents, stay nearly zero.

The high latitude part, from the auroral ovals to the poles, still accounts for about 10 kV of the corotational e.m.f. This is small, but not negligible, compared to the externally applied polar cap potential.

Although in the case of the Earth the internal dynamo plays a minor role, the situation can be different in other magnetospheres (for a review see, e.g., Hill, 1984). Thus Jupiter's magnetospheric processes seem to be dominantly powered by the rotation of the planet. In this case the Jovian satellite Io and its plasma torus are important as an external load (Shawhan, 1976; Eviatar and Siscoe, 1980). Rotational dynamo action has also been proposed to be important at Uranus (Hill *et al.*, 1983).

Of *external* sources there are both (1) MHD-type dynamos (the solar wind, the plasma sheet, and regions of the magnetosphere where convection field is externally enforced, e.g., by a viscous-like interaction) and (2) thermoelectric generators (regions where the gradient and curvature drifts produce charge separation (see, e.g., Block, 1984; Atkinson, 1984a; Vasyliunas, 1984).

It is outside the scope of the present paper to discuss the dynamos themselves. These are well described in the literature and for reviews the reader is referred to, e.g., Stern (1983, 1984). Only one aspect will be briefly discussed, namely the possible role of spatially small-scale dynamo regions and corresponding fine-structure in the ionosphere-magnetosphere coupling.

It has been suggested by Heikkila (1982), Lemaire (1977), and Lemaire *et al.* (1979) that plasma from the magnetosheath is injected in the form of clouds into the magnetosphere. Until they lose their momentum these clouds would form localized and temporary MHD-dynamo regions on closed field lines. In addition they would create regions where, due to curvature and gradient drifts, the plasma would contain both protons of solar wind origin and magnetospheric O^+ ions. It has been suggested by Lundin (1984), Lundin and Dubinin (1985), and Hultqvist *et al.* (1986) that such clouds would form dynamo regions by polarization due to the differential motions of the different types of ions.

A general expression for the differential flow vector of two ion species has been derived by Hultqvist *et al.* (1986). From measured values of particles and fields it was estimated that terms containing pressure gradients and transverse electric currents could easily reach values of some hundreds of km s^{-1}, and that also inertia terms and the magnetic gradient terms could approach 100 km s^{-1} with quite resonable assumptions about characteristic times and characteristic lengths.

Thus determination of electric field from particle fluxes could be uncertain by tens of mV m^{-1} even if local particle distribution functions were known exactly (but not their gradients or whether variations were temporal or spatial). This underscores the need for direct measurement of the electric field itself, e.g., by double probes.

The concept of intruding plasma clouds as localized generator regions for auroral arc structures has been further elaborated by Stasiewicz (1985b). As the localized cloud dynamo drives Birkeland currents to the ionosphere and back, magnetic-field aligned potential drops may develop in the upward current branch. As a necessary but not sufficient condition for this to happen Stasiewicz (1985b) concludes that the scale has to be so small that the characteristic dimension is less than $3(B_i/B_m)r_{ge}$, where B_i/B_m is the ionospheric to equatorial magnetic field strength ratio and r_{ge} is the electron gyroradius in the equatorial plane. These considerations are applied also to nightside plasma ion clouds whose existence has been known for a long time (DeForest and McIlwain, 1971).

4. Redistribution of Plasma

We now know that ionospheric ions contribute significantly to populating many regions of the magnetosphere (in addition to the obvious one, the plasmasphere). They are also present in all energy ranges from thermal to high energy. It started with the discovery by Shelley *et al.* (1972) of precipitating O^+ with energies up to 12 keV and was later followed by the first direct observations of the O^+ ions leaving the ionosphere (Shelley *et al.*, 1976). Consequences for magnetosphere ionosphere coupling were discussed by Sharp and Shelley (1981). Reviews of the ionosphere as a source of magnetospheric ions were given by Shelley *et al.* (1982, 1985); Horwitz (1982); Yau *et al.* (1985); Chappell *et al.* (1987), and Shelley (1987). For further results related to this topic, see, e.g., Lennartsson *et al.* (1985); Stokholm *et al.* (1985); Ipavich *et al.* (1985); Waite *et al.* (1985); Green *et al.* (1986); Horwitz *et al.* (1986); Moore *et al.* (1986); Kremser *et al.* (1987). Only rather recently has the composition of the bulk of the storm time ring current been measured (Gloeckler *et al.*, 1985). There, too, it is found that injection of ionospheric ions is important.

The presence of ionospheric ions in the magnetospheric plasma has of course important consequences both macroscopically (e.g., by local-dynamo effects, as mentioned above) and microscopically (by their influence on wave propagation, instabilities and wave particle interaction). Heavy ions of ionospheric origin may also influence the localization and initiation of plasma sheet instabilities during substorms (Baker *et al.*, 1985).

A comprehensive collection of papers on the distribution of hot energetic ions in the magnetosphere is found in a book edited by Johnson (1983), see also Hultqvist (1983a, b). Even very energetic (112–157 keV) O^+ ions have been observed in the plasma sheet (Ipavich *et al.*, 1984). In a review, Hultqvist (1985) emphasizes that present knowledge of low-energy plasma in the magnetosphere is very far from complete and improving this knowledge is greatly needed.

Only recently have measurements been made in the entire range of the ring current population (Stüdemann *et al.*, 1986). The importance of oxygen ions during magnetic storms was confirmed and in one case the oxygen-ion energy density equalled that of proton energy density even up to the highest energy (300 keV) observed by the instrument. Computations by Cladis and Francis (1985) indicate that oxygen ions in the

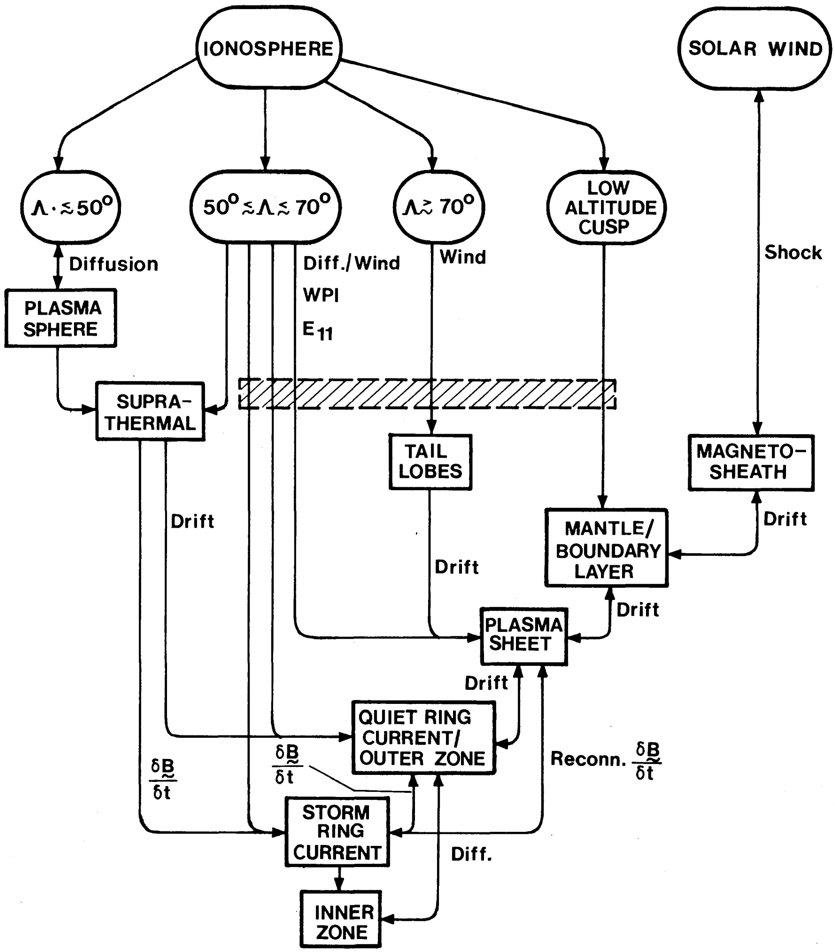


Fig. 3. Schematic overview of sources, transport, and acceleration of plasma in the magnetosphere accoridng to Collin *et al.* (1984).

storm-time ring current may go through a closed-loop ionospheric-magnetospheric circulation (cf. Block and Fälthammar, 1969). It is interesting to note that, in an early paper on ring-current energization, Dessler *et al.* (1961) argued that the ring-current particles could be locally heated ions from the normal protonosphere.

We may note that in the auroral acceleration region the incomplete knowledge of the low-energy plasma introduces a considerable uncertainty in stability analyses, as discussed for example in the review by Kaufmann (1984). See also Lennartsson *et al.* (1985). A schematic overview of sources, transport and acceleration of plasma in the magnetosphere according to Collin *et al.* (1984) is shown in Figure 3.

At low and middle latitudes storm-time depletion of the plasmasphere are followed by a diffusive refilling process that takes 7–22 hours (Horwitz *et al.*, 1984).

On polar cap field lines a supersonic streaming of ionospheric plasma – the polar wind – has long been predicted for theoretical reasons. These were initially discussed by Hanson and Pattersson (1963) and Dessler and Michel (1966), later formalized by Axford (1968), Banks and Holzer (1968), and others, for a review see Cowley (1980). The polar wind has been observed by the Dynamics Explorer spacecraft (Gurgioli and Burch, 1982, 1985; Nagai *et al.*, 1984; Waite *et al.*, 1985; Lockwood *et al.*, 1985). In addition to the theoretically expected cold polar wind there are also substantial fluxes of suprathermal (above 100 eV) field-aligned O^+ ions that seem to have been subject to some other acceleration processes. Topside ionospheric electron density profiles can be used to deduce the magnetic field-aligned outflow of ions (Lockwood and Titheridge, 1981, 1982; Lockwood, 1982, 1983). Persoon *et al.* (1983) concluded that in addition to a subsonic to supersonic transition at about 1000 km altitude there is a transition from collision dominated to collision free outflow at about 1.5–2R_E. Over the polar cap, DE-1 observations reported by Waite *et al.* (1985) show outward flow of suprathermal low-energy (less than 10 eV) O^+ ions with fluxes exceeding $2 \times 10^{12}\ \mathrm{m}^{-2}\ \mathrm{s}^{-1}$, mainly from a region near the dayside polar cap boundary. The integrated source strength is estimated to be $7 \times 10^{24}\ \mathrm{s}^{-1}$ for quiet (K_p less than 3) and $2 \times 10^{25}\ \mathrm{s}^{-1}$. The distinction between classical polar wind outflow and O^+ enhanced suprathermal flow has been analysed by Moore *et al.* (1985).

The extraction of ionospheric ions is related to the Birkeland currents both directly through exchange of charge carriers and indirectly by parallel and perpendicular ion-acceleration mechanisms driven by the Birkeland currents.

Observations of upflowing ions from the auroral zone and polar cap were reviewed by Yau *et al.* (1984). The total outflow of ionospheric ions into the magnetosphere, according to Collin *et al.* (1984) from S3–3 data, is given in Table I. Locally, upward oxygen ion fluxes exceeding $10^{14}\ \mathrm{m}^{-2}\ \mathrm{s}^{-1}$ have been observed with DE-2 at 900 km altitude and account for a substantial fraction of the simultaneously observed Birkeland currents (Heelis *et al.*, 1984).

A direct effect of Birkeland currents is due to the fact that the closed-loop current, of which the Birkeland currents are a part, is carried by different particle species in different parts of the current loop.

Possible consequences at *ionospheric* levels were discussed by Block and Fälthammar

TABLE I

The estimated terrestrial ion outflow in the energy range 0.5 to 16 keV for O^+ and H^+ during magnetic storms and quiet times

	Range	Mean
Quiet time		
H^+	$0.7-1.4 \times 10^{25}\ s^{-1}$	$1.1 \times 10^{25}\ s^{-1}$
O^+	$0.15-0.4 \times 10^{25}\ s^{-1}$	$0.27 \times 10^{25}\ s^{-1}$
Total	$0.85-1.8 \times 10^{25}\ s^{-1}$	$1.3 \times 10^{25}\ s^{-1}$
O^+/H^+	0.1 –0.4	0.25
Storms time		
H^+	$1.5-4.5 \times 10^{25}\ s^{-1}$	$3.0 \times 10^{25}\ s^{-1}$
O^+	$3.5-5.0 \times 10^{25}\ s^{-1}$	$4.2 \times 10^{25}\ s^{-1}$
Total	$5.0-9.5 \times 10^{25}\ s^{-1}$	$7.2 \times 10^{25}\ s^{-1}$
O^+/H^+	0.7–2.1	1.4

The range indicates the uncertainty of the estimate resulting from both counting statistics and uncertainties in the identification of the newly outflowing ions.

(1968, 1969) who showed that this effect can modify the F-region density distribution and contribute to the formation of F-region troughs. As the density depletions are associated with loss of ionospheric ions to the magnetosphere, they also suggested that 'the ionosphere and magnetosphere might form a more or less closed loop for the plasma' (Block and Fälthammar, 1969).

At the *magnetospheric* end the consequences of the transition of current carriers has been analyzed in a series of papers by Atkinson (see Atkinson, 1984a, and references given therein). If Birkeland currents carried predominantly by electrons connect to transverse magnetospheric currents carried largely by ions, depletions or enhancements should occur depending on the direction of the Birkeland currents. Thus, inward Birkeland currents would cause enhancements of magnetospheric plasma and outward currents would cause depletions. Another example: if part of the cross-tail electric current is deviated through the ionosphere, plasma accumulates in the morning side and evacuates in the evening side of the magnetosphere. According to Atkinson (1984b) the distribution of Birkeland currents at ionospheric altitudes can thus be used to diagnose plasma redistribution in the outer magnetosphere. A steady-state model developed on this basis has been further extended to include thick adjacent current sheets mapping to the whole plasma sheet (Atkinson, 1984c).

A particularly interesting exchange of mass between the ionosphere and the magnetosphere is the outflow of heavy ionospheric ions in the form of 'beams' and 'conics' (see, e.g., Gorney *et al.*, 1981), both because of the acceleration mechanisms of which they bear witness and because they show that very effective chemical separation can take place in a cosmical plasma. The latter could have important consequences in the context of astrophysical abundance considerations.

In the *beams* the distribution function has its maximum along the magnetic field.

However, considering the limits set by instrument resolution some of them may be post-accelerated conics masquerading as beams. They appear to be accelerated by an electric field, and contain information about the potential drop between their source and the observation point. However, they do not give a simple quantitative measure of this potential, because it is obvious that they have also been subject to non-adiabatic processes. For example, their energy spread (50–150 eV) is much greater than would be expected (0.2 eV) if ionospheric O^+ had only fallen adiabatically through a potential drop. It is also well established (Kintner *et al.*, 1979; Cattell, 1981; Kaufmann and Kintner, 1982, 1984) that there is very close correlation between ion beams and electrostatic hydrogen-cyclotron waves. As described in a review by Kaufmann (1984) the typical observed distribution functions of the ion beams may be explained if it is assumed that these waves are the result of the instability of the beams. The typical beam temperatures of 50–150 eV are approximately such that the beams should be marginally stable to generation of hydrogen-cyclotron waves. As the ascent through the mirror field tends to narrow the distribution, the beams would, in this scenario, be kept near the limit of instability. This would also mean that at high altitude the beams would have such a width as to efficiently prevent them from reaching the opposite hemisphere. Hence the non-observation of downgoing beams.

The *conics* have a maximum in their distribution function at a non-zero value of transverse to parallel velocity ratio and are apparently the result of a transverse acceleration followed by expulsion by the magnetic mirror force. Two main explanations of the transverse acceleration have been proposed. One main explanation invokes waves, either electrostatic ion-cyclotron waves (Ungstrup *et al.*, 1979; Lysak *et al.*, 1980; Papadopoulos *et al.*, 1980; Ashour-Abdalla *et al.*, 1981; Singh *et al.*, 1981, 1983; Dusenberry and Lyons, 1981; Okuda and Ashour-Abdalla, 1981, 1983; Lockwood, 1982; Ashour-Abdalla and Okuda, 1983, 1984; Okuda, 1984; Gurnett *et al.*, 1984; Retterer *et al.*, 1987) or lower hybrid waves (Chang and Coppi, 1981; Singh and Schunk, 1984; Kintner *et al.*, 1986; cf. also Temerin, 1986). In a two-component plasma of stationary hydrogen and oxygen ions and drifting electrons preferential heating of either hydrogen or oxygen can take place depending on the ratio of electron drift speed and the ratio of hydrogen to oxygen concentration (Ashour-Abdalla and Okuda, 1983). For a given critical drift the maximum perpendicular heating is generally larger for the oxygen ions than for the hydrogen ions (Ashour-Abdalla and Okuda, 1984). Both theoretical analysis and numerical simulation were used and gave results in good agreement. Nishikawa *et al.* (1985) studied ion heating by hydrogen-cyclotron waves, such as are often observed on auroral field lines, using analytical methods as well as numerical simulations. Much stronger heating resulted for oxygen ions than for hydrogen ions. Preferential heating of oxygen ions has been reported from rocket-borne measurements (Moore *et al.*, 1986). As pointed out by Horwitz (1984) transverse acceleration of O^+ ions is also favoured by the fact that due to greater inertia they have a longer residence time in the acceleration region. Results reported by Gorney *et al.* (1985) indicate that the residence time and, hence, the heating of upflowing ions may be much enhanced by downward pointing parallel electric fields. Recently, Ashour-Abdalla *et al.*

(1987) have used numerical simulation to study ion heating in multicomponent plasmas in the presence of an auroral electron beam. The result was used for a clarification of the various types of heating that can occur at different altitudes.

Kintner and Gorney (1984) searching the S3–3 data found only one case of perpendicular ion acceleration and broadband plasma waves at the satellite. The wave mode could not be identified, but the electric field of the waves was, in all cases, smaller than required by present theories.

Several authors (Mozer *et al.*, 1980; Lennartsson, 1980; Kletzig *et al.*, 1983; Yang and Kan, 1983; Greenspan, 1984; Ohsawa, 1986; Borovsky, 1984; Borovsky and Joyce, 1986) have considered electrostatic shocks or similar electrostatic structures as an alternative or complementary explanation of the transverse acceleration. In this case, too, it is found that the heavier ion species are preferentially accelerated and tend to become less field-aligned. Whereas Yang and Kan (1983) consider this an auxiliary mechanism (to cyclotron heating), Borovsky (1984) finds in his simulations that the particles become more field-aligned as the ion-cyclotron waves grow and, therefore, suggests that the ion conics produce the waves rather than *vice versa.*

According to Cattell (1984) the S3–3 data indicate that both electrostatic ion-cyclotron waves and electric field gradients contribute to the energizing of the ions but that neither is sufficient to account for the observed energy. According to Cladis (1986) even the large-scale convection electric field can contribute to accelerating escaping ions.

A major difficulty in clarifying the processes leading to formation of beams and conics is the limited knowledge of cold background electrons and ions (Kaufman, 1984). The present state observational knowledge of low-energy plasma outside the plasmapause has been reviewed by Hultqvist (1985; cf. also Stokholm *et al.*, 1985).

The literature on ion acceleration is now quite extensive. A survey is provided by AGU Monograph No. 38 (Chang 1986).

5. Magnetic-Field-Aligned Electric Fields

One of the crucial questions in magnetosphere-ionosphere coupling is the ability of the plasma to support magnetic-field-aligned ('parallel') electric potential drops and thus electrically and dynamically decouple magnetically-connected plasma regions. This property is also intimately related to the ability to carry Birkeland currents and to energize charged particles.

There has now for some time been an almost complete consensus that such electric fields do exist and that they play an important role in energizing auroral particles. It is, however, also clear that parallel electric fields alone cannot account for all the observed features of the accelerated particle populations. A review of parallel electric fields, with extensive references to the literature, has been given by Fälthammar (1983). The present discussion will be limited to general outlines and some comments.

5.1. Possible types of parallel fields

A central problem concerning parallel electric fields is what forces are responsible for balancing the *dc* electric force on the charged particles. On the basis of the kind of force involved, the parallel electric fields can be divided into three categories (Fälthammar, 1977, 1978). The three forces are:

(1) The net force from wave fields. In the magnetosphere only the electric part of the wave field, and only its component parallel to the magnetic field, can contribute appreciably to the momentum balance. The prime example of this case is that of anomalous resistivity. The collisionless thermoelectric effect proposed by Hultqvist (1971) would also belong to this category.

(2) The magnetic mirror force. Magnetic mirror supported parallel fields have been extensively invoked in explaining observed particle distributions above the auroral zone.

(3) Inertia forces. A final possibility is that the force from the electric field is balanced by the inertia of the charged particles themselves. This is the situation in electric double layers. Such are well known from the laboratory and are often considered to be important in space, too.

It is very likely that these categories occur in combinations. E.g., in the presence of a parallel field supported by the magnetic mirror force, strong wave activity may still substantially change particle distributions. Or, numerous weak electric double layers may appear and disappear at random with a result very much resembling a state of anomalous resistivity (Block, 1972, 1981).

Each of these categories of electric fields has its own peculiarities. A couple of these will be mentioned here.

Anomalous resistivity requires that the wave electric field along the magnetic field (and *a fortiori* the total wave field) has an r.m.s. value well exceeding (probably by a factor of 10 or more) the *d.c.* field that it supports: $E_{\mathrm{r.m.s.}} \gg E_{d.c.}$ (Fälthammar, 1977; Shawhan *et al.*, 1978). Although the existing wave fields are not known in enough detail for an accurate evaluation, it can be estimated that anomalous resistivity might account for parallel *d.c.* fields of the order of mV m^{-1}. This could still be enough for supporting several kV, but the potential drop would have to be distributed over distances of the order of one or more Earth radii.

Another feature of the anomalous resistivity is that the power is dissipated locally. For the current densities known to prevail above the aurora, any appreciable parallel field supported by anomalous resistivity would imply extremely rapid heating of the local plasma of the order of eV s^{-1} (Block and Fälthammar, 1976; Block, 1984).

In a plasma with two ion species, anomalous resistivity may also lead to selective ion acceleration. A numerical simulation by Mitchell and Palmadesso (1984) showed that the momentum transferred from the waves mainly affected one ion species (H^+) leaving the other (O^+) to be freely accelerated in the d.c. electric field. From numerical simulations Rowland and Palmadesso (1983) concluded that low-frequency ion-cyclotron turbulence can limit the high-velocity runaways via pitch angle scattering. From comparisons between simulations and dynamics explorer observations electron precipitation bursts Lin and Rowland (1985) suggest that anomalous resistivity does play an important role in connection with particle energization.

Parallel electric fields supported by the *magnetic mirror force* could in principle exist even in the absence of a current, as shown by Alfvén and Fälthammar (1963; cf. also

Persson, 1963, 1966; Whipple, 1977; Serizawa and Sato, 1984). Mostly, however, the upward mirror force is invoked in the context of upward Birkeland currents, where the principal current carriers are impeded by the magnetic mirror force.

In this case, too, the maximum field strength that can be supported should typically be of the order of a few mV m^{-1}. Thus, any large potential drops would have to involve large distances along the magnetic field. Unlike anomalous resistivity, magnetic-mirror supported electric fields need not involve strong local heating.

As shown by, e.g., Knight (1973); Lemaire and Scherer (1974, 1983); Friedman and Lemaire (1980) and others, the current voltage relation for mirror-supported fields is, under certain assumptions, linear over 2 or 3 powers of ten in voltage and current. For a mirror-supported field this is a relation between the current density and the *total* voltage drop, *not* a local relation between current density and electric field at any given part of the flux tube. Thus there does not exist a *conductivity*, only a *conductance*. For typical plasma sheet parameters this conductance is 3×10^{-6} A $(\mathrm{mV/m})^{-1}$ (Fälthammar, 1978). This holds, however, only *provided* the loss cone of the source plasma is continually replenished. Otherwise the conductance can be reduced to arbitrarily low values. Furthermore, it has been shown by Yamamoto and Kan (1985) that the current density can be substantially reduced relative to that given by the Knight–Lemaire formula, if the potential drop is concentrated at altitudes as low as 4000 km. According to simulations by Yamamoto and Kan (1986a, b) wave-particle interaction may further reduce the electric current at given potential below the value that would apply in the adiabatic case.

Corresponding to the current-voltage relation there should be a relation between energy flux and accelerating voltage such that in a certain energy range the former is proportional to the square of the accelerating voltage. Such a relation has been confirmed by Lyons *et al.* (1979) and Menietti and Burch (1981). A linear rather than quadratic relation was reported by Wilhelm (1980) but this seems to be explainable in terms of a spatial variation of the source plasma.

Electric double layers represent the opposite extreme in terms of spatial distribution. The thickness L of a double layer with voltage drop V in a plasma of electron temperature T_e is of the order of

$$L = C(\mathrm{e}V/kT_e)^{1/2}\,\lambda_{\mathrm{D}}\,,$$

where λ_{D} is the Debye length and the factor C is of the range 10–100 (Shawhan *et al.*, 1978).

An important feature of the double layer is that the power released goes into energetic particles that deposit their energy elsewhere (for example in the denser ionosphere below), and therefore there is no problem of excessive local heating.

5.2. Observational Evidence

The earliest observations interpreted as evidence of parallel electric fields were those of field alignment and narrow energy peaks in precipitating electron fluxes (McIlwain, 1960; Albert, 1967; Evans, 1968). Gradually the evidence accumulated and now includes both observation of (1) natural particle populations, (2) motion of artificially particles injected by means of active experiments, and (3) direct measurements of electric fields. At the same time it has become abundantly clear that the situation is not simple, and that d.c. electric fields alone cannot explain all the characteristics of observed particle spectra. In the following will be given a very brief summary of observations interpreted as evidence for parallel electric fields as well as objections against this interpretation. For a more complete review and references to the original papers, see e.g. Fälthammar (1983) and, for particle observations, Kaufmann (1984).

5.2.1. *Observations of Natural Particle Populations*

In addition to the field alignment and narrow energy peaks already mentioned, there is a large amount of satellite data showing characteristic structures in the particle distribution functions (see, e.g., Kaufmann, 1984). This include occurrence of an *acceleration boundary* in downcoming electron distributions, similar to what would be expected if the particles had been accelerated in a potential drop of a few kV. Another important feature is a *widened loss cone*, as would be expected from a potential drop below the satellite (again in the kV range). *Upstreaming ion beams*, if assumed to have passed an electrostatic accelerated ion region, indicate a potential drop below the satellite that is in rough agreement with the widened electron loss cone.

While most observations of accelerated particle populations have been made at altitudes of a few hundred to a few thousand kilometers, an inverted V event observed at 13 R_E has been reported (Huang *et al.*, 1984) showing that at least sometimes the acceleration region can be very far away.

Low voltage (tens of V) upward pointing electric fields – perhaps analogous to the wall sheath in a laboratory plasma – have been reported by Winningham and Gurgiolo (1982). Equatorward of the morning side polar cusp the electrons that carry the downward region 1 Birkeland currents appear to be accelerated by potential drops of tens of V at altitudes of several thousands kilometers (Burch *et al.*, 1983).

Although the magnitude of the potentials of parallel electric fields can be estimated from existing particle data, the determination of their spatial distribution is much more difficult. As shown by Greenspan *et al.* (1981) even distinguishing between double layers and smoothly distributed electric fields is difficult with existing data and would require accurate high-resolution measurements of low energy electrons (around 100 eV and less). As the distribution functions often vary appreciably within a satellite spin period multiple detectors with high time-resolution would be needed. To extract the information carried by the upstreaming ions a wider energy coverage would also be desirable.

A number of objections were already long ago raised against the interpretation of auroral particle distributions in terms of parallel electric fields (O'Brien, 1970; Whalen and McDiarmid, 1972; Whalen and Daly, 1979). For example, Hall and Bryant (1974) considered that the shape of the angular distribution of electrons and of the width of the energy peak were indicative of a stochastic acceleration process. Wave particle interaction was also invoked by many authors to explain the width of the energy peak and the occurrence of multiple peaks (for references, see Hall *et al.*, 1984). The velocity space features (acceleration boundary, widened loss cone) are diffuse and the velocity space region that corresponds to trapping between the electric field above and the magnetic mirror below is populated. The upcoming ion beams are much wider than could be explained by electrostatic acceleration alone. These and other difficulties with purely electrostatic acceleration were summarized by Bryant (1987). Although some of these objections can be eliminated even within adiabatic models (cf., e.g., Block, 1984; Brüning and Goertz, 1985; Lotko, 1986) it is of course not surprising that non-adiabatic processes play a role, too (cf. Section 5.3).

5.2.2. *Active Experiments*

The first active experiments to indicate the existence of parallel electric field were shaped charge Ba releases (Haerendel *et al.*, 1976), where a clear acceleration of the Ba ions could be seen (in one case corresponding to a voltage drop of 7.4 kV at an altitude of 7500 km). This seems to be one of the most conclusive observations of the existence of a parallel electric field. By now a total of half a dozen such experiments have been made. The main results of these have been compiled by Stenbaek-Nielsen *et al.* (1984).

In a Ba jet experiment Stenbaek-Nielsen *et al.* (1984) noticed a sudden decrease in the speed of progression of the tip of the jet as it reached 8100 km altitude. Their interpretation was that at this altitude the barium was accelerated rapidly upward to a sufficient speed that the density decreased below detectability. For this to happen through energization by wave fields and subsequent magnetic-mirror expulsion, the authors estimate that gyroresonant waves in excess of 25 mV m^{-1} would have been required. They, therefore, favour a d.c. electric field as the only plausible explanation. From a detailed study of the brightness distribution they estimate a lower limit to the strength of the d.c. field. The result is that the potential must have been in excess of 1 kV. Because of the limited resolution of the TV images only an upper limit (200 km) could be set to the distance over which the potential drop occurred. Hence, the field strength must have been at least 5 mV m^{-1}.

One important observation in this case was that as the Ba jet itself drifted (westward), and auroral arc segments drifted through it (from south to north), no apparent corresponding changes were seen in the behaviour of the barium. The situation persisted for at least 10 min. Thus the observed electric field appears to have been a large-scale horizontal structure and not associated with individual arc structures. Of course this does not exclude arc-related parallel electric fields still higher up. The simultaneously observed background luminosity at 6300–4278 Å corresponded to a characteristic energy of the precipitating electrons of about one keV.

Active experiments have also been made using electron beams ejected from a rocket to probe the parallel electric fields. Reflexions of the electrons were observed, which are compatible with the existence of parallel electric fields above the rocket but do not constitute a proof (Wilhelm *et al.*, 1984, 1985). If interpreted in terms of parallel electric fields they indicate field strengths of 1–2 mV m^{-1} above about 2500 km and potentials of at least a few kV or more.

5.2.3. *Electric Field Measurements*

Direct electric field measurements at high and low altitudes have shown different latitude distributions that imply the existence of parallel electric fields at intermediate altitudes (Mozer and Torbert, 1980). The discovery by direct measurements of the so-called electrostatic shocks (Mozer *et al.*, 1977), i.e., regions of strong (hundreds of $\mathrm{mV\,m^{-1}}$) over short distances (a few km) imply that electric field mapping along electrically equipotential magnetic field lines does not apply. These results have been confirmed and extended by the VIKING satellite (Fälthammar *et al.*, 1987; Block *et al.*, 1987; Block, 1987). In a static situation this would irrevocably imply the existence of parallel electric fields somewhere between the ionosphere and the satellite. However, it cannot be excluded that this lack of mapping may instead be due to the induction field of an oblique Alfvén wave as proposed by Haerendel (1983).

Because of their limited spatial extent, *strong* electric double layers are difficult to detect (Boehm and Mozer, 1981). However, certain electric field structures observed with the VIKING satellite have been identified as strong electric double layers (Block *et al.*, 1987). Occurrence of numerous *weak* double layers have been discovered (Temerin *et al.*, 1982; Hudson *et al.*, 1983) and seem to be able to account for integrated potential drops of the order of several kV.

As already mentioned the occurrence of numerous weak double layers and solitons have been known from electric field measurement at high altitude on the satellite S3-3. According to results reported by Boehm *et al.* (1984) and Kellog *et al.* (1984) similar structures exist even at rocket altitudes (above 200 km). In both cases the observations were made by means of double probe electric field experiments. But in neither case was the experiment designed for this unexpected discovery. Therefore, the information on the size and motion of the structures is still incomplete.

In the flight reported by Kellog *et al.* (1984) a lower limit to the typical voltage drop of the double-layer-like structures was determined to be 0.4 V. In the corresponding structures observed by Boehm *et al.* (1984) the electric fields, mostly parallel to the magnetic field, were typically $50\ \mathrm{mV\,m^{-1}}$ and the corresponding potentials at least 0.1 V. However, the lower limit of the potentials observed varied up to 2 V. No limit could be set on the size of the structures but a lower limit of their velocity was estimated to be $15\ \mathrm{km\,s^{-1}}$. In addition closely spaced soliton-like structures were observed with electric fields greater than $1\ \mathrm{mV\,m^{-1}}$. Further measurements with dedicated instrumentation are necessary to clarify the nature of these phenomena.

5.3. Some further comments

Parallel electric fields have usually been considered to be important mainly in regions of upward electric currents, because outflowing ionospheric electrons were thought to provide copious current carrying capability for downward currents. However, theoretical work by Newman (1985) has shown that also downward parallel electric fields may exist. The key to this is the existence, at the low altitude end of the field line, of an ambipolar diffusion region with an upward directed electric field with a potential of a few V. This is found to be sufficient to exponentially reduce the densities of electrons and ions enough that a net downward electric field above would not extract excessive electron current. Observations have been reported by Gorney *et al.* (1985), where the phase space distributions indicate acceleration by a downward-pointing electric field with a potential of a few tens to a few hundreds of V over the altitude range of 1000–6000 km.

Not unexpectedly, parallel electric fields alone are insufficient to account for all features of auroral particle distributions. Features that seem to require other explanations have been pointed out by several authors. These features were summarized by Bryant (1983), who favours an approach aimed at explaining the auroral acceleration entirely by wave-particle interactions without recourse to a *d.c.* electric field. This approach has been expounded in a series of papers (Bingham *et al.*, 1984; Hall *et al.*, 1984; Bryant, 1987, and references given therein). The acceleration mechanism favoured by these authors is lower hybrid waves driven by ion beams streaming toward the Earth in the plasmasheet boundary. Referring to the ion beams reported, e.g., by DeCoster and Frank (1979) and the wave observations of Gurnett and Frank (1977), and Mozer *et al.* (1979) the authors find that (1) the energy of the beams is easily sufficient to power the auroral acceleration and (2) the normalized energy density of lower hybrid waves on auroral field lines is high.

However, the electric field of these waves is very nearly transverse to the magnetic field, and it is only the parallel component of the wave electric field that can contribute to the field-aligned acceleration. Therefore, more detailed knowledge of the wave fields seems to be necessary to assess with certainty what effect the waves have on the particles. Furthermore, as some of the evidence of field-aligned parallel fields remains intact, attempts to explain auroral acceleration entirely without parallel *d.c.* fields appear problematic, and a combined approach more promising.

The combined effect of (ion-acoustic) wave turbulence and a d.c. parallel electric field has been analyzed by Stasiewicz (1984), using the quasi-linear Vlasov equation to estimate the runaway region in velocity space. One of the results is a new interpretation of the classical type of peaked auroral spectrum (see Figure 4). The accelerating potential U is not given by the energy at the peak, but by the difference between this energy and the energy at the minimum of the spectrum. The latter energy is in typical cases about 1 keV and is related to the energy required for electron runaway in the presence of the wave field. A theoretical prediction of its value is, however, not possible without much better knowledge of the actual wave spectra in the interaction region than is now available.

For the low-energy electrons, region I in Figure 4, the spectral form E^{-1} is ascribed to the heating of the trapped electrons and not to atmospheric backscatter (cf. Evans, 1974). In this interpretation region II in Figure 4 corresponds to runaway electrons that have fallen freely through the d.c. electric field of the acceleration region. Hot magnetospheric electrons with velocities exceeding the runaway velocity will pass the acceleration region unimpeded and form region III of the spectrum.

Stasiewicz (1985a) also derived a relation between the voltage and current density

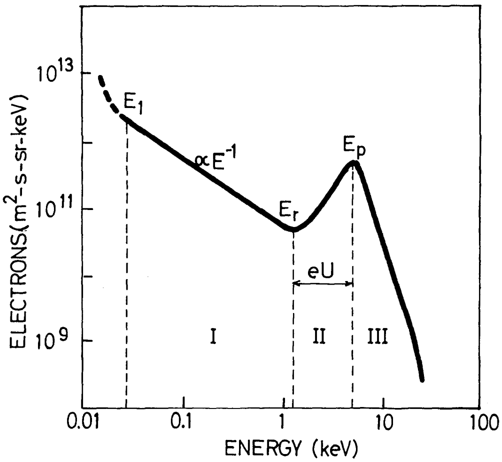


Fig. 4. New interpretation of inverted-V electron spectra according to Stasiewicz (1984). The accelerated voltage v is now related to the difference of the energies E_p at the peak and E_r at the minimum.

that is a generalization of that of Knight (1973) and a corresponding relation between energy at the spectral peak and the precipitating energy flux. When the energy at the peak is much larger than the source plasma temperature, this relation reduces to the quadratic form that applies in the adiabatic case (Lundin and Sandahl 1978).

6. Concluding Remarks

For a proper understanding of astrophysical phenomena, theoretical analysis, which by necessity must rely on simplifying assumptions, must be guided by empirical knowledge of how real plasmas behave. Laboratory experiments have an essential role to play in this context, but *in situ* observations of the magnetospheric plasma can, in some respects, provide us with an even better knowledge of plasma behaviour in natural conditions. The plasma in the magnetosphere (including the ionosphere) and the solar wind are the only cosmic plasmas accessible to extensive *in situ* observations and experimentation.

Observations of magnetospheric plasmas extend our empirical knowledge to a new range of plasma parameters by many powers of ten. It is also fortunate that plasmas in the Earth's neighbourhood cover such wide ranges of density and temperature and, most importantly, that magnetosphere-ionosphere interactions cause such a rich variety of plasma processes to take place.

The truly fundamental progress in the understanding of the magnetosphere has only begun. Important observational discoveries have opened a new epoch in magnetospheric research. The knowledge already obtained, and the insights still to be gained, in magnetospheric research should be of great value in understanding astrophysical plasmas and may have profound impacts on astrophysics.

References

Albert, R. D.: 1967, 'Nearly Monoenergetic Electron Fluxes Detected During a Visible Aurora', *Phys. Rev. Letters* **18**, 369.

Alfvén, H.: 1942, *On the Cosmogony of the Solar System*, Stockholms Observatoriums Annaler, I, Vol. 14, No. 2.

Alfvén, H.: 1958, 'On the Theory of Magnetic Storms and Aurorae', *Tellus* **10**, 104.

Alfvén, H.: 1977, 'Electric Currents in Cosmic Plasmas', *Rev. Geophys. Space Phys.* **15**, 271.

Alfvén, H.: 1981, *Cosmic Plasma*, Monograph, D. Reidel Publ. Co., Dordrecht, Holland.

Alfvén, H.: 1984, 'Cosmogony as an Extrapolation of Magnetospheric Research', *Space Sci. Rev.* **39**, 65.

Alfvén, H.: 1986a, 'The Plasma Universe', *Phys. Today* **39**, 22.

Alfvén, H.: 1986b, 'Model of the Plasma Universe', *IEEE Transaction on Plasma Science* **PS-14**, 629.

Alfvén, H.: 1986c, 'Double Layers and Circuits in Astrophysics', *IEEE Transactions on Plasma Science* **PS-14**, 779.

Alfvén, H. and Arrhenius, G.: 1976, *Evolution of the Solar System*, NASA Scientific Publ. 345, US Government Printing Office, Washington, D.C.

Alfvén, H. and Fälthammar, C.-G.: 1963, *Cosmical Electrodynamics, Fundamental Principles*, Clarendon Press, Oxford.

Araki, T., Kamei, T., and Tyemori, T.: 1984, 'Polar Cap Vertical Currents Associated with Northward Interplanetary Magnetic Field', *Geophys. Res. Letters* **11**, 23.

Ashour-Abdalla, M. and Okuda, H.: 1983, 'Transverse Acceleration of Ions on Auroral Field Lines', in R. G. Johnson (ed.), *Energetic Ions Composition in the Earth's Magnetosphere*, Terra Scientific Publ. Co., Tokyo, p. 43.

Ashour-Abdalla, M. and Okuda, H.: 1984, 'Turbulent Heating of Heavy Ions on Auroral Field Lines', *J. Geophys. Res.* **89**, 2235.

Ashour-Abdalla, M., Okuda, H., and Cheng, C. Z.: 1981, 'Acceleration of Heavy Ions on Auroral Field Lines', *Geophys. Res. Letters* **8**, 795.

Ashour-Abdalla, M., Schriver, D., and Okuda, H.: 1987, *Transverse Ion Heating in Multicomponent Plasmas Along Auroral Zone Field Lines*, Report PPG-1090, Center for Plasma Physics and Fusion Engineering, University of California, Los Angeles.

Atkinson, G.: 1984a, 'The Role of Currents in Plasma Redistribution, Magnetospheric Currents', in T. Potemra (ed.), *AGU Geophysical Monograph 28*, Proc. Chapman Conf. on Magnetosphere Current Systems, Irvington, Virginia, April 5–8, 1983, Vol. 28, 325.

Atkinson, G.: 1984b, 'Field-Aligned Currents as a Diagnostic Tool: Result, a Renovated Model of the Magnetosphere', *J. Geophys. Res.* **89**, 217.

Atkinson, G.: 1984c, 'Thick Current Sheets in the Renovated Model of the Magnetosphere', *J. Geophys. Res.* **89**, 8949.

Axford, W. I.: 1968, 'The Polar Wind and the Terrestrial Helium Budget', *J. Geophys. Res.* **73**, 6855.

Baker, D. N., Fritz, T. A., Lennartsson, W., Wilken, B., Kroehl, H. W., and Birn, J.: 1985, 'The Role of Heavy Ionospheric Ions in the Localizatin of Substorm Disturbances on March 22, 1979: CDAW 6', *J. Geophys. Res.* **90**, 1273.

Banks, P. M. and Holzer, T. E.: 1968, 'The Polar Wind', *J. Geophys. Res.* **73**, 6846.

Barfield, J. N., Saflekos, N. A., Sheehan, R. E., Carovillano, R. L., Potemra, T. A., and Knecht, D., 1986, 'Three-Dimensional Observations of Birkeland Currents', *J. Geophys. Res.* **91**, 4393.

Baumjohann, W.: 1983, 'Ionospheric and Field-Aligned Current Systems in the Auroral Zone: A Concise Review', *Adv. Space Res.* **2**, 55.

Bingham, R., Bryant, D. A., and Hall, D. S.: 1984, 'A Wave Model for the Aurora', *Geophys. Res. Letters* **11**, 327.

Block, L. P.: 1972, 'Potential Double Layers in the Ionosphere', *Cosmic Electrodyn.* **3**, 349.

Block, L. P.: 1981, *Physics of Auroral Arc Formation*, AGU Geophysical Monograph 25, Am. Geophys. Union, Washington D.C., p. 218.

Block, L. P.: 1984, 'Three-Dimensional Potential Structure Associated With Birkeland Currents, Magnetospheric Currents', in T. Potemra (ed.), *AGU Geophysical Monograph 28*, Proc. Chapman Conf. on Magnetosphere Current Systems, Irvington, Virginia, April 5–8, 1983, p. 315.

Block, L. P.: 1987, 'Acceleration of Auroral Particles by Magnetic-Field Aligned Electric Fields', *Proc. 8th ESA Symposium on European Rocket and Balloon Programmes and Relates Research,* Sunne, Sweden, 17–23 May, 1987, ESA SP-270, p. 281.

Block, L. P. and Fälthammar, C.-G.: 1968, 'Effects of Field-Aligned Currents on the Structure of the Ionosphere', *J. Geophys. Res.* **73**, 4807.

Block, L. P. and Fälthammar, C.-G.: 1969, in B. M. McCormac and A. Omholt (eds.), *Field-Aligned Currents and Auroral Precipitation in Atmospheric Emissions*, Proc. NATO Adv. Study Inst., Ås, Norway, July 29–August 9, 1968, p. 285.

Block, L. P. and Fälthammar, C.-G.: 1976, 'Mechanisms that May Support Magnetic Field-Aligned Electric Fields in the Magnetosphere', *Ann. Geophys.* **32**, 161.

Block, L. P., Fälthammar, C.-G., Lindqvist, P.-A., Marklund, G. T., Mozer, F. S., and Pedersen, A.: 1987, 'Electric Field Measurements on Viking: First Results', *Geophys. Res. Letters* **14**, 435.

Boehm, M. H. and Mozer, F. S.: 1981, 'An S3-3 Search for Confined Regions of Large Parallel Electric Fields', *Geophys. Res. Letters* **8**, 607.

Boehm, M. H., Carlson, C. W., McFadden, J. and Mozer, F. S.: 1984, 'Observations of Double-Layer-Like and Soliton-Like Structures in the Ionosphere', *Geophys. Res. Letters* **11**, 511.

Borovsky, J. E.: 1984, 'The Production of Ion Conics by Oblique Double Layers', *J. Geophys. Res.* **89**, 2251.

Borovsky, J. E. and Joyce, G.: 1986, 'The Direct Production of Ion Conics by Plasma Double Layers', in T. Chang (ed.), *Ion Acceleration in the Magnetosphere and Ionosphere*, AGU Geophysical Monograph Vol. 38, p. 317.

Bryant, D. A.: 1983, 'The Hot Electrons in and Above the Auroral Ionosphere: Observations and Physical Implications', in *High Latitude Space Plasma Physics*, Plenum, New York, 295, 1983.

Bryant, D. A.: 1987, 'Wave Acceleration of Auroral Electrons', *8th ESA Symposium on European Rocket and Balloon Programmes and Related Research,* Sunne, Sweden, 17–23 May, 1987, ESA SP-270, p. 273.

Brüning, K. and Goertz, C.: 1985, 'Influence of the Electrons Source Distribution on Field-Aligned Currents', *Geophys. Res. Letters* **12**, 53.

Burch, J. L., Reiff, P. H., and Sugiura, M.: 1983, 'Upward Electron Beams Measured by DE-1: A Primary Source of Dayside Region-1 Birkeland Currents', *Geophys. Res. Letters* **10**, 753.

Burch, J. L., Reiff, P. H., Menietti, J. D., Heelis, R. A., Hanson, W. B., Shawhan, S. D., Shelley, E. G., Sugiura, M., Weimer, D. R., and Winningham, J. D.: 1985, 'IMF By-Dependent Plasma Flow and Birkeland Currents in the Dayside Magnetosphere 1. Dynamics Explorer Observations', *J. Geophys. Res.* **90**, 1577.

Burke, W. J.: 1984, Electric Fields and Currents Observed by S3–2 in the Vicinity of Discrete Arcs, Magnetospheric Currents', in T. Potemra (ed.), *AGU Geophysical Monograph 28*, Proc. Chapman Conf. on Magnetospheric Current Systems, Irvington, Virginia, April 5–8, 1983, Vol. 28, 294.

Bythrow, P. F., Doyle, M. A., Potemra, T. A., Zanetti, L. J., Huffman, R. E., Meng, C. I., Hardy, D. A., Rich, F. J., and Heelis, R. A.: 1986, 'Multiple Auroral Arcs and Birkeland Currents: Evidence for Plasma Sheet Boundary Waves', *Geophys. Res. Letters* **13**, 805.

Bythrow, P. F., Potemra, T. A., Hanson, W. B., Zanetti, L. J., Meng, C.-I., Huffman, R. E., Rich, F. J. and Hardy, D. A.: 1984, 'Earthward Directed High-Density Birkeland Currents Observed by HILAT', *J. Geophys. Res.* **89**, 9114.

Bythrow, P. F., Potemra, T. A., Zanetti, L. J., Erlandson, R. A., Hardy, D. A., Rich, F. J., and Acuna, M. H.: 1987, 'High Latitude Currents in the 0600 to 0900 MLT Sector: Observatins from Viking and DMSP-F7', *Geophys. Res. Letters* **14**, 423.

Cattell, C.: 1981, 'The Relationship of Field-Aligned Currents to Electrostatic Ion-Cyclotron Waves', *J. Geophys. Res.* **86**, 3641.

Cattell, C.: 1984, 'Association of Field-Aligned Currents with Small-Scale Auroral Phenomena, Magnetospheric Currents', in T. Potemra (ed.), *AGU Geophysical Monograph 28*, Proc. Chapman Conf. on Magnetosphere Current Systems, Irvington, Virginia, April 5–8, 1983.

Chang, T. (editor-in-chief): 1986, *Ion Acceleration in the Magnetosphere and Ionosphere*, Geophysical Monograph 38, American Geophysical Union, Washington, D.C.

Chang, T. and Coppi, B.: 1981, 'Lower Hybrid Acceleration and Ion Evolution in the Supraauroral Region', *Geophys. Res. Letters* **8**, 1253.

Chappell, C. R., Moore, T. E., and Waite, J. G., Jr.: 1987, 'The Ionosphere as a Fully Adequate Source of Plasma for the Earth's Magnetosphere', *J. Geophys. Res.* **92**, 5896.

Cladis, J. B.: 1986, 'Parallel Acceleration and Transport of Ions from Polar Ionosphere to Plasma Sheet', *Geophys. Res. Letters* **13**, 893.

Cladis, J. G. and Francis, W. E.: 1985, 'The Polar Ionosphere as a Source of the Storm Time Ring Current', *J. Geophys. Res.* **90**, 3465.

Clauer, C. R. and Kamide, Y.: 1985, 'DP 1 and DP 2 Current Systems for the March 22, 1979 Substorms', *J. Geophys. Res.* **90**, 1343.

Clauer, C. R., Banks, P. M., Smith, A. Q., Jørgensen, T. S. Friis-Christensen, E., Vennerstrom, S., Wickwar, V. B. Kelly, J. D. and Doupnik, J.: 1984, 'Observation of Interplanetary Magnetic Feld and of Ionospheric Plasma Convection in the Vicinity of the Dayside Polar Cleft', *Geophys. Res. Letters* **11**, 891.

Collin, H. L., Sharp, R. D., and Shelley, E. G.: 1984, 'The Magnitude and Composition of the Outflow of Energetic Ions from the Ionsophere', *J. Geophys. Res.* **89**, 2185.

Cowley, S. W. H.: 1980, 'Plasma Populations in a Simple Open Model Magnetosphere', *Space Sci. Rev.* **26**, 217.

DeCoster, R. J. and Frank, L. A.: 1979, 'Observations Pertaining to the Dynamics of the Plasma Sheet', *J. Geophys. Res.* **84**, 5099.

DeForest, S. E. and McIlwain, C. E.: 1971, 'Plasma Clouds in the Magnetosphere', *J. Geophys. Res.* **76**, 3587.

Dessler, A. J.: 1984, 'The Evolution of Arguments Regarding the Existence of Field-Aligned Currents, Magnetospheric Currents', in T. Potemra (ed.), *AGU Geophysical Monograph 28*, Proc. Chapman Conf. on Magnetosphere Current System, Irvington, Virginia, April 5–8, 1983, pp. 22–28.

Dessler, A. J., Hanson, W. B., and Parker, E. N.: 1961, 'Formation of the Geomagnetic Storm Main-Phase Ring Current', *Geophys. Res.* **66**, 3631.

Dessler, A. J. and Michel, F. C.: 1966, 'Plasma in the Geomagnetic Tail', *J. Geophys. Res.* **71**, 1421.

Dusenbery, P. B. and Lyons, L. R.: 1981, 'Generation of Ion-Conic Distribution by Upgoing Ionospheric Electrons', *J. Geophys. Res.* **86**, 7627.

Erlandson, R. E., Pottelette, R., Potemra, T. A., Zanetti, L. J., Bahnsen, A., Lundin, R., and Hamelin, M.: 1987, *Geophys. Res. Letters* **14**, 431.

Evans, D. S., 1968, 'The Observations of a Near-Monoenergetic Flux of Auroral Electrons', *J. Geophys. Res.* **73**, 2315.

Evans, D. S.: 1974, 'Precipitating Electron Fluxes Formed by a Magnetic Field-Aligned Potential Difference', *J. Geophys. Res.* **79**, 2853.

Eviatar, A. and Siscoe, G. L.: 1980, 'Limit on Rotational Energy Available to Excite Jovian Aurora', *Geophys. Res. Letters* **7**, 1085.

Fälthammar, C.-G.: 1977, 'Problems Related to Macroscopic Electric Fields in the Magnetosphere', *Rev. Geophys. Space Phys.* **15**, 457.

Fälthammar, C.-G.: 1978, 'Generation Mechanisms for Magnetic-Field-Aligned Electric Fields in the Magnetosphere', *J. Geomagn. Geoelectr.* **30**, 419.

Fälthammar, C.-G.: 1983, 'Magnetic-Field-Aligned Electric Fields', *ESA J.* **7**, 385.

Fälthammar, C.-G.: 1985, *Magnetosphere-Ionosphere Coupling*, ESA SP-235, p. 107.

Fälthammar, C.-G., Akasofu, S.-I., and Alfvén, H.: 1978, 'The Significance of Magnetospheric Research for Progress in Astrophysics', *Nature* **275**, 185.

Fälthammar, C.-G., Block, L. P., Lindqvist, P.-A., Marklund, G. T., Pedersen, A., and Mozer, F. S.: 1987, 'Preliminary Results from the DC Electric Field Experiment on Viking', *Ann. Geophys.* **5A**, 171.

Friedman, M. and Lemaire, J.: 1980, 'Relationship Between Auroral Fluxes and Field-Aligned Electric Potential Differences', *J. Geophys. Res.* **85**, 664.

Gloeckler, G., Wilken, B., Stüdemann, W., Ipavich, F. M., Hovestadt, D., Hamilton, D. C., and Kremser, G.: 1985, 'First Composition Measurement of the Bulk of the Storm-Time Ring Current (1 to 300 keV/e) with AMPTE-CCE', *Geophys. Res. Letters* **12**, 325.

Gorney, D. J., Clarke, A., Croley, D., Fennell, J., Luhmann, J., and Mizera, P.: 1981, 'The Distribution of Ion Beams and Conics below 8000 km', *J. Geophys. Res.* **86**, 83.

Gorney, D. J., Chiu, Y. T., and Croley, Jr., D. R.: 1985, 'Trapping of Ion Conics by Downward Parallel Electric Fields', *J. Geophys. Res.* **90**, 4205.

Green, J. L., Waite, J. H., Chappell, C. R., Chandler, M. O., Doupnik, J. R., Richards, P. G., Heelis, R., Shawhan, S. D., and Brace, L. H.: 1986, 'Observations of Ionospheric Magnetospheric Coupling: DE and Chatanika Coincidences', *J. Geophys. Res.* **91**, 5803.

Greenspan, M. E.: 1984, 'Effects of Oblique Double Layers on Upgoing Ion Pitch Angle and Gyrophase', *J. Geophys. Res.* **89**, 2842.

Greenspan, M. E., Silevitch, M. B., and Whipple, E. C., Jr.: 1981, 'On the Use of Electron Data to Infer the Structure of Parallel Electric Fields', *J. Geophys. Res.* **86**, 2175.

Greenwald, R. A.: 1982, 'Recent Advances in Magnetosphere-Ionosphere Coupling', *Rev. Geophys. Space Phys.* **20**, 577.

Gurgioli, C. and Burch, J. L.: 1982, 'DE-1 Observations of the Polar Wind – A Heated and an Unheated Component', *Geophys. Res. Letters* **9**, 945.

Gurgioli, C. and Burch, J. L.: 1985, 'Composition of the Polar Wind – Not Just H^+ and He^+', *Geophys. Res. Letters* **12**, 69.

Gurnett, D. A. and Frank, L. A.: 1977, 'A Region of Intense Plasma Wave Turbulence on Auroral Field Lines', *J. Geophys. Res.* **82**, 1031.

Gurnett, D. A., Huff, R. L., Menietti, J. D., Burch, J. L., Winningham, J. D., and Shawhan, S. D.: 1984, 'Correlated Low-Frequency Electric and Magnetic Noise Along the Auroral Field Lines', *J. Geophys. Res.* **89**, 8971.

Haerendel, G.: 1980, 'Auroral Particle Acceleration – An Example of a Universal Plasma Process', *ESA J.* **4**, 197.

Haerendel, G.: 1981, 'Magnetospheric Processes Possibly Related to the Origin of Cosmic Rays, Origin of Cosmic Rays', in G. Setti, G. Spada, and A. W. Wolfendale (eds.), *IAU Symp.* **94**, 373.

Haerendel, G.: 1983, in B. Hultqvist and T. Hagfors (eds.), *An Alfvén Wave Model of Auroral Arcs, High-Latitude Space Plasma Physics*, Plenum Press, New York, p. 515.

Haerendel, G., Rieger, E., Valenzuela, A., Föppl, H., Stenbaek-Nielsen, H. C., and Wescott, E. M.: 1976, 'First Observation of Electrostatic Acceleration of Barium Ions into the Magnetosphere', *ESA* **SP-115**, 203.

Hall, D. S. and Bryant, D. A.: 1974, 'Collimation of Auroral Particles by Time-Varying Acceleration', *Nature* **251**, 402.

Hall, D. S., Bryant, D. A., and Bingham, R.: 1984, 'Electron Acceleration by Lower Hybrid Waves on Auroral Field Lines', *Adv. Space Res.* **4**, 515.

Hanson, W. B. and Patterson, T. N. L.: 1963, 'Diurnal Variation of the Hydrogen Concentration in the Exosphere', *Planetary Space Sci.* **11**, 1035.

Heelis, R. A., Winningham, J. D., Sugiura, M., and Maynard, N. C.: 1984, 'Particle Acceleration Parallel and Perpendicular to the Magnetic Field Observed by DE-2', *J. Geophys. Res.* **89**, 3893.

Heikkila, W. J.: 1982, 'Impulsive Plasma Transport Through the Magnetopause', *Geophys. Res. Letters* **9**, 159.

Hill, T. W.: 1984, in T. Potemra (ed.), *Rotationally-Induced Birkeland Current Systems, Magnetospheric Currents*, AGU Geophysical Monograph 28, Proc. Chapman Conf. on Magnetosphere Current Systems, Irvington, Viginia, April 5–8, 1983, p. 340.

Hill, T. W., Dessler, A. J., and Rassbach, M. E.: 1983, 'Aurora on Uranus: A Faraday Disc Dynamo Mechanism', *Planetary Space Sci.* **31**, 1187.

Horwitz, J. L.: 1982, 'The Ionosphere as a Source for Magnetospheric Ions', *Rev. Geophys. Space Phys.* **20**, 929.

Horwitz, J. L.: 1984, 'Residence Time Heating Effect in Auroral Conic Generation', *Planetary Space Sci.* **32**, 1115.

Horwitz, J. L., Brace, L. H., Comfort, R. H., and Chappell, C. R.: 1986, 'Dual-Spacecraft Measurements of Plasmasphere-Ionosphere Coupling', *J. Geophys. Res.* **91**, 11203.

Horwitz, J. L., Confort, R. H., and Chappell, C. R.: 1984, 'Thermal Ion Composition Measurements of the Formation of the New Outer Plasmasphere and Double Plasmapause During Storm Recovery Phase', *Geophys. Res. Letters* **11**, 701.

Hruška, A.: 1986, 'Field-Aligned Currents in the Earth Plasma Sheet', *J. Geophys. Res.* **91**, 371.

Huang, C. Y., Frank, L. A., and Eastman, T. E.: 1984, 'High-Altitude Observations of an Intense Inverted V Event', *J. Geophys. Res.* **89**, 7423.

Hudson, M. K., Lotko, W., Roth, I., and Witt, E.: 1983, 'Solitary Waves and Double Layers on Auroral Field Lines', *J. Geophys. Res.* **88**, 916.

Hultqvist, B.: 1971, 'On the Production of a Magnetic-Field Aligned Electric Field by the Interaction between the Hot Magnetospheric Plasma and Cold Ionosphere', *Planetary Space Sci.* **19**, 749.

Hultqvist, B.: 1983a, 'On the Origin of the Hot Ions in the Disturbed Dayside Magnetosphere', *Planetary Space Sci.* **31**, 173.

Hultqvist, B.: 1983b, 'On the Dynamics of the Ring Current', *J. Geophys.* **52**, 203.

Hultqvist, B.: 1985, 'Observations of Low-Energy Plasma Outside the Plasmasphere', *Space Sci. Rev.* **42**, 275.

Hultqvist, B., Lundin, R., and Stasiewicz, K.: 1986, 'Ion Interactions in the Magnetospheric Boundary Layer', in T. Chang (ed.), *Ion Acceleration in the Magnetosphere and Ionosphere*, AGU Geophysical Monograph 38, p. 127.

Iijima, T., Potemra, T. A., Zanetti, L. J., and Bythrow, P. F.: 1984, 'Large-Scale Birkeland Currents in the Dayside Polar Region During Strongly Northward IMF: A New Birkeland Current System', *J. Geophys. Res.* **89**, 7441.

Ipavich, F. M., Galvin, A. B., Gloeckler, G., Hovestadt, D., Klecker, B., and Scholer, M.: 1984, 'Energetic (> 100 keV) O^+ Ions in the Plasma Sheet', *Geophys. Res. Letters* **11**, 504.

Ipavich, F. M., Galvin, A. B., Scholer, M., Gloeckler, G., Hovestadt, D., and Klecker, B.: 1985, 'Suprathermal O^+ and H^+ Ion Behavior During the March 22, 1979 (CDAW 6) Substorms', *J. Geophys. Res.* **90**, 1263.

Johnson, R. G. (ed.): 1983, *Energetic Ion Composition in the Earths Magnetosphere*, Terra Scientific Publ. Co., Tokyo and D. Reidel Publ. Co., Dordrecht, Holland.

Kan, J. R., Williams, R. L., and Akasofu, S.-I.: 1984, 'A Mechanism for the Westward Travelling Surge During Substorms', *J. Geophys. Res.* **89**, 2211.

Kaufmann, R. L.: 1984, 'What Auroral Electron and Ion Beams Tell Us about Magnetosphere-Ionosphere Coupling', *Space Sci. Rev.* **37**, 313.

Kaufmann, R. L. and Kintner, P. M.: 1982, 'Upgoing Ion Beams. 1. Microscopic Analysis', *J. Geophys. Res.* **87**, 10487.

Kaufmann, R. L. and Kintner, P. M.: 1984, 'Upgoing Ion Beams: 2. Fluid Analysis and Magnetosphere-Ionosphere Coupling', *J. Geophys. Res.* **89**, 2195.

Kellogg, P. J., Monso, S. J., and Whalen, B. A.: 1984, 'Observation of Double-Layer-Like Structures at Rocket Altitudes', *Geophys. Res. Letters* **11**, 515.

Kelly, T. J., Russell, C. T., Walker, R. J., Parks, G. K., and Gosling, J. T.: 1986, 'ISEE 1 and 2 Observations of Birkeland Currents in the Earth's Inner Magnetosphere', *J. Geophys. Res.* **91**, 6945.

Kintner, P. M. and Gorney, D. J.: 1984, 'A Search for the Plasma Processes Associated with Perpendicular Ion Heating', *J. Geophys. Res.* **89**, 937.

Kintner, P. M., Kelley, M. C., Sharp, R. D., Ghielmetti, A., Temerin, M., Catell, C. A., Mizera, P. F., and Fennell, J. F.: 1979, 'Simultaneous Observations of Energetic (keV) Upstreaming Ions and Electrostatic Hydrogen Cyclotron Waves', *J. Geophys. Res.* **84**, 7201.

Kintner, P. M., LaBelle, J., Scales, W., Yau, A. W., and Whalen, B. A.: 1986, 'Observations of Plasma Waves Within Regions of Perpendicular Ion Acceleration', *Geophys. Res. Letters* **13**, 1113.

Kletzing, C., Catell, C., and Mozer, F. S.: 1983, 'Evidence for Electrostatic Shocks as the Source of Discrete Auroral Arcs', *J. Geophys. Res.* **88**, 4105.

Knight, S.: 1973, 'Parallel Electric Fields', *Planetary Space Sci.* **21**, 741.

Kremser, G., Stüdemann, W., Wilken, B., Gloeckler, G., Hamilton, D. C., and Ipavich, F. M.: 1987, 'Average Spatial Distributions of Energetic O^+, O^{2+}, O^{6+}, and C^{6+} Ions in the Magnetosphere Observed by AMPTE CCE', *J. Geophys. Res.* **92**, 4459.

Lemaire, J.: 1977, 'Impulsive Penetration of Filamentary Plasma Elements Into the Magnetospheres of the Earth and Jupiter', *Planetary Space Sci.* **25**, 887.

Lemaire, J. and Scherer, M.: 1974, 'Ionosphere Plasma Sheet Field-Aligned Currents and Parallel Electric Fields', *Planetary Space Sci.* **22**, 1485.

Lemaire, J. and Scherer, M.: 1983, 'Field-Aligned Current Density Versus Electric Potential Characteristics for Magnetospheric Flux Tubes', *Ann. Geophys.* **1**, 91.

Lemaire, J., Rycroft, M. J., and Roth, M.: 1979, 'Control of Impulsive Penetration of Solar Wind Irregularities Into the Magnetosphere by the Interplanetary Magnetic Field Direction', *Planetary Space Sci.* **27**, 47.

Lennartsson, W.: 1980, 'On the Consequences of the Interaction between the Auroral Plasma and the Geomagnetic Field', *Planetary Space Sci.* **28**, 135.

Lennartsson, W., Sharp, R. D., and Zwickl, R. D.: 1984, 'Substorm Effects on the Plasma Sheet Ion Composition on March 22, 1979 (CDAW 6)', *J. Geophys. Res.* **90**, 1243.

Lin, C. S. and Rowland, H. L.: 1985, 'Anomalous Resistivity and AE-D Observations of Auroral Electron Acceleration', *J. Geophys. Res.* **90**, 4221.

Lockwood, M.: 1982, 'Thermal Ion Flows in the Topside Auroral Ionosphere and the Effects of Low-Altitude, Transverse Acceleration', *Planetary Space Sci.* **30**, 595.

Lockwood, M.: 1983, 'Field-Aligned Plasma Flow in the Quiet, Mid-Latitude Ionosphere Deduced from Topside Soundings', *J. Atmospheric Terrest. Phys.* **45**, 1.

Lockwood, M. and Titheridge, J. E.: 1981, 'Ionospheric Origin of Magnetospheric O^+ Ions', *Geophys. Res. Letters* **8**, 381.

Lockwood, M. and Titheridge, J. E.: 1982, 'Departures from Diffuse Equilibrium in the Topside F-Layer from Satellite Soundings', *J. Atmospheric Terrest. Phys.* **44**, 425.

Lockwood, M., Waite, J. H., Jr., Moore, T. E., Johnson, J. F. E., and Chappell, C. R.: 1985, 'A New Source of Suprathermal O^+ Ions Near the Dayside Polar Cap Boundary', *J. Geophys. Res.* **90**, 4099.

Lotko, W.: 1986, 'Diffusive Acceleration of Auroral Primaries', *J. Geophys. Res.* **91**, 191.

Lundin, R.: 1984, 'Solar Wind Energy Transfer Regions Inside the Dayside Magnetopause. II. Evidence for an MHD Generator Process', *Planetary Space Sci.* **32**, 757.

Lundin, R. and Dubinin, E. M.: 1985, 'Solar Wind Energy Transfer Regions Inside the Dayside Magnetopause. III. Accelerated Heavy Ions as Tracers for MHD-Processes in the Dayside Boundary Layer', *Planetary Space Sci.* **33**, 891.

Lundin, R. and Sandahl, I.: 1978, 'Some Characteristics of the Parallel Electric Field Acceleration of Electrons over Discrete Auroral Arcs as Observed from two Rocket Flights', *ESA* **SP-135**, 125.

Lyons, L. R., Evans, D. S., and Lundin, R.: 1979, 'An Observed Relation between Magnetic Field-Aligned Electric Fields and Downward Electron Energy Fluxes in the Vicinity of Auroral Forms', *J. Geophys. Res.* **84**, 457.

Lysak, R. L., Hudson, M. K., and Temerin, M.: 1980, 'Ion Heating by Strong Electrostatic Ion Cyclotron Turbulence', *J. Geophys. Res.* **85**, 678.

Marklund, G. T., Raadu, M. A., and Lindqvist, P.-A.: 1985, 'Effects of Birkeland Current Limitation on High-Latitude Convection Patterns', *J. Geophys. Res.* **90**, 10864.

Mauk, B. H. and Zanetti, L. J.: 1986, *IUGG Quadrennial Report on: Magnetospheric Electric Fields and Currents*, APL/JHU 86–16, Space Physics Group, The Johns Hopkins University, Laurel, Md.

McIlwain, C. E.: 1960, 'Direct Measurements of Particles Producing Visible Auroras', *J. Geophys. Res.* **65**, 2727.

Menietti, J. D. and Burch, J. L.: 1981, 'A Satellite Investigation of Energy Flux and Inferred Potential Drop in Auroral Electron Energy Spectra', *Geophys. Res. Letters* **8**, 1095.

Mitchell, H. G. and Palmadesso, P. J.: 1984, 'O^+ Acceleration Due to Resistive Momentum Transfer in the Auroral Field Line Plasma', *J. Geophys. Res.* **89**, 7573.

Moore, T. E., Chappell, C. R., Lockwood, M., and Waite, J. H., Jr.: 1985, 'Superthermal Ion Signatures of Auroral Acceleration Processes', *J. Geophys. Res.* **90**, 1611.

Moore, T. E., Lockwood, M., Chandler, M. O., Waite, J. H., Jr., Chappell, C. R., Persoon, A., and Sugiura, M.: 1986, 'Upwelling O^+ Ion Source Characteristics', *J. Geophys. Res.* **91**, 7019.

Moore, T. E., Lockwood, M., Chandler, M. O., Waite, J. H., Jr., Chappell, C. R., Persoon, A., and Sugiura, M.: 1986, "Upwelling O^+ Ion Source Characteristics', *J. Geophys. Res.* **91**, 7019.

Moore, T. E., Pollock, C. J., Arnoldy, R. L., and Kintner, P. M.: 1986, 'Preferential O^+ Heating in the Topside Ionosphere', *Geophys. Res. Letters* **13**, 901.

Mozer, F. S., Carlson, C. W., Hudson, M. K., Torbert, R. B., Parady, B., Yatteau, J., and Kelley, M. C.: 1977, 'Observations of Paires Electrostatic Shocks in the Polar Magnetosphere', *Phys. Rev. Letters* **38**, 292.

Mozer, F. S. and Torbert, R. B.: 1980, 'An Average Parallel Electric Field Deduced from the Latitude and Altitude Variations of the Perpendicular Electric Field Below 8000 Kilometers', *Geophys. Res. Letters* **7**, 219.

Mozer, F. S., Cattell, C. A., Temerin, M., Torbert, R. B., von Glinski, S., Woldroft, M., and Wygrant, J.: 1979, 'The dc and ac Electric Field, Plasma Density, Plasma Temperature and Field-Aligned Current Experiments on the S3-3 Satellite', *J. Geophys. Res.* **84**, 5875.

Mozer, F. S., Cattell, C. A., Hudson, M. K., Lysak, R. L., Temerin, M., and Torbert, R. B.: 1980, 'Satellite Measurements and Theories of Low Altitude Auroral Particle Acceleration', *Space Sci. Rev.* **27**, 155.

Nagai, T.: 1987, 'Field-Aligned Currents Associated with Substorms in the Vicinity of Synchronous Orbit. 2. GOES 2 and GOES 3 Observations', *J. Geophys. Res.* **92**, 2432.

Nagai, T., Singer, H. J., Ledley, B. G., and Olsen, R. C.: 1987, 'Field-Aligned Currents Associated with Substorms in the Vicinity of Synchronous Orbit. 1. The July 5, 1979, Substorm Observed by SCATHA, GOES 3, and GOES 2', *J. Geophys. Res.* **92**, 2425.

Nagai, T., Waite, J. H., Jr., Green, J. L., Chappell, C. R., Olsen, R. C., and Comfort, R. H.: 1984, 'First Measurements of Supersonic Polar Wind in the Polar Magnetosphere', *Geophys. Res. Letters* **11**, 669.

Newman, A. L.: 1985, *Effect of Low Altitude Ambipolar Diffusion on High Altitude Particle Distributions*, paper presented at the 2nd Int. School of Space Simulations, Kapaa, 1985.

Nishikawa, K.-I., Okuda, H., and Hasegawa, A.: 1985, 'Heating of Heavy Ions on Auroral Field Lines in the Presence of a Large Amplitude Hydrogen Cyclotron Wave', *J. Geophys. Res.* **90**, 419.

O'Brien, B. J.: 1970, 'Consideration that the Source of Auroral Energetic Particle is not a Parallel Electrostatic Field', *Planetary Space Sci.* **18**, 1821.

Ohsawa, Y.: 1986, 'Resonant Ion Acceleration by Oblique Magnetosonic Shock Wave in a Collisionless Plasma', *Phys. Fluids* **29**, 773.

Okuda, H.: 1984, 'Turbulent Heating of Heavy Ions on Auroral Field Lines', *J. Geophys. Res.* **89**, 2235.

Okuda, H. and Ashour-Abdalla, M.: 1981, 'Formation of a Conical Distribution and Intense Ion Heating in the Presence of Hydrogen Cyclotron Waves', *Geophys. Res. Letters* **8**, 811.

Okuda, H. and Ashour-Abdalla, M.: 1983, 'Acceleration of Hydrogen Ions and Conic Formation Along Auroral Field Lines', *J. Geophys. Res.* **88**, 899.

Papadopoulos, K., Gaffey, J. D., Jr., and Palmadesso, P. J.: 1980, 'Stochastic Acceleration of Large M/Q Ions by Hydrogen Cyclotron Waves in the Magnetosphere', *Geophys. Res. Letters* **7**, 1014.

Persoon, A. M., Gurnett, D. A., and Shawhan, S. D.: 1983, 'Polar Cap Electron Densities from DE-1 Plasma Wave Observations', *J. Geophys. Res.* **88**, 10123.

Persson, H.: 1963, 'Electric Field Along a Magnetic Line of Force in a Low-Density Plasma', *Phys. Fluids* **6**, 1756.

Persson, H.: 1966, 'Electric Field Parallel to the Magnetic Field in a Low-Density Plasma', *Phys. Fluids* **9**, 1090.

Potemra, T. (ed.): 1984, *Magnetospheric Currents*, AGU Geophysical Monograph 28.

Potemra, T. A. and Zanetti, L. J.: 1985, 'Characteristics of Large-Scale Birkeland Currents in the Cusp and Polar Regions', in J. A. Holtet and A. Egeland (eds.), *Proc. of the NATO Advanced Workshop on Morphology and Dynamics of the Polar Cusp*, D. Reidel Publ. Co., Dordrecht, Holland (in press).

Potemra, T. A., Zanetti, L. J., Bythrow, P. F., Lui, A. T. Y., and Iijima, T.: 1984, 'By-Dependent Convection Patterns During Northward Interplanetary Magnetic Field', *J. Geophys. Res.* **89**, 9753.

Potemra, T. A., Zanetti, L. J., Erlandson, R. E., Bythrow, P. F., Gustafsson, G., Acuna, M. H., and Lundin, R.: 1987, 'Observations of Large-Scale Birkeland Currents with Viking', *Geophys. Res. Letters* **14**, 419.

Primdahl, F. and Marklund, G.: 1986, 'Birkeland Currents Correlated with Direct-Current Electric Fields Observed During the CENTAUR Black Brant X Rocket Experiment', *Can. J. Phys.* **64**, 1412.

Reiff, P. H. and Burch, J. L.: 1985, 'IMF By-Dependent Plasma Flow and Birkeland Currents in the Dayside Magnetosphere. 2. A Global Model for Northward and Southward IMF', *J. Geophys. Res.* **90**, 1595.

Retterer, J. M., Chang, T., and Jasperse, J. R.: 1983, 'Ion Acceleration in the Supraauroral Region: A Monte Carlo Model', *Geophys. Res. Letters* **10**, 583.

Retterer, J. M., Chang, T., and Jasperse, J. R.: 1986, 'Ion Acceleration by Lower Hybrid Waves in the Supraauroral Region', *J. Geophys. Res.* **91**, 1609.

Retterer, J. M., Chang, T., Crew, G. B., Jasperse, J. R., and Winningham, J. D.: 1987, 'Monte Carlo Modeling of Ionospheric Oxygen Acceleration by Cyclotron Resonance with Broad-Band Electromagnetic Turbulence', *Phys. Rev. Letters* **59**, 148.

Robert, P., Gendrin, R., Perraut, S., and Roux, A.: 1984, 'GEOS-2 Identification of Rapidly Moving Current Structures in the Equatorial Outer Magnetosphere During Substorms', *J. Geophys. Res.* **89**, 819.

Rothwell, P. L., Silevitch, M. B., and Block, L. P.: 1984, 'A Model for the Propagation of the Westward Traveling Surge', *J. Geophys. Res.* **89**, 8941.

Rowland, H. L. and Palmadesso, P. J.: 1983, 'Anomalous Resistivity Due to Low-Frequency Turbulence', *J. Geophys. Res.* **88**, 7997.

Serizawa, Y. and Sato, T.: 1984, 'Generation of Large-Scale Potential Difference by Currentless Plasma Jets Along the Mirror Field', *Geophys. Res. Letters* **11**, 595.

Sharp, R. D. and Shelley, E. G.: 1981, 'Magnetosphere-Ionosphere Coupling Through the Auroral Acceleration Region', reprint Conf. Proc. No. 295, *The Physical Basis of the Ionosphere in the Solar-Terrestrial System*, AGARD, Advisory Group for Aerospace Research and Development, North Atlantic Treaty Organization, Neuilly-sur-Seine, France.

Sharp, R. D., Lennartsson, W., and Strangeway, R. J.: 1985, 'The Ionospheric Contribution to the Plasma Environment in the Near-Earth Space', *Radio Sci.* **20**, 456.

Shawhan, S. D.: 1976, 'Io Sheath-Accelerated Electrons and Ions', *J. Geophys. Res.* **81**, 3373.

Shawhan, S. D., Fälthammar, C.-G., and Block, L. P.: 1978, 'On the Nature of Large Auroral Zone Electric Fields at One R_E Altitude', *J. Geophys. Res.* **83**, 1049.

Shelley, E. G.: 1986, 'Magnetospheric Energetic Ions from the Earth's Ionosphere', *Adv. Space Res.* **6**, 121.

Shelley, E. G., Johnson, R. G., and Sharp, R. D.: 1972, 'Satellite Observations of Energetic Heavy Ions During a Geomagnetic Storm', *J. Geophys. Res.* **77**, 6104.

Shelley, E. G., Peterson, W. K., Ghielmetti, A. G., and Geiss, J.: 1982, 'The Polar Ionosphere as a Source of Energetic Magnetospheric Plasma', *Geophys. Res. Letters* **9**, 941.
Shelley, E. G., Sharp, R. D., and Johnson, R. G.: 1976, 'Satellite Observations of an Ionospheric Acceleration Mechanism', *Geophys. Res. Letters* **3**, 654.
Singh, N. and Schunk, R. W.: 1984, 'Energization of Ions in the Auroral Plasma by Broadband Waves: Generation of Ion Conics', *J. Geophys. Res.* **89**, 5538.
Singh, N., Schunk, R. W., and Sojka, J. J.: 1981, 'Energization of Ionospheric Ions by Electrostatic Hydrogen Cyclotron Waves', *Geophys. Res. Letters* **8**, 1249.
Singh, N., Schunk, R. W., and Sojka, J. J.: 1983, 'Preferential Perpendicular Acceleration of Heavy Ionospheric Ions by Interactions with Electrostatic Hydrogen-Cyclotron Waves', *J. Geophys. Res.* **88**, 4055.
Smits, D. P., Hughes, W. J., Cattell, C. A., and Russell, C. T.: 1986, 'Observations of Field-Aligned Currents, Waves, and Electric Fields at Substorm Onset', *J. Geophys. Res.* **91**, 121.
Stasiewicz, K.: 1984, 'On the Origin of the Auroral Inverted V Electron Spectra', *Planetary Space Sci.* **32**, 379.
Stasiewicz, K.: 1985a, 'The Influence of a Turbulent Region on the Flux of Auroral Electrons', 'Generation of Magnetic-Field Aligned Currents, Parallel Electric Fields, and Inverted-V Structures by Plasma Pressure Inhomogeneities in the Magnetosphere', *Planetary Space Sci.* **33**, 591.
Stasiewicz, K.: 1985b, 'On the Formation of Auroral Arcs', *Planetary Space Sci.* **33**, 1037.
Stenbaek-Nielsen, H. C., Hallinan, T. J., Wescott, E. M., and Foeppl, H.: 1984, 'Acceleration of Barium Ions Near 8000 km Above an Aurora', *J. Geophys. Res.* **89**, 10788.
Stern, D. P.: 1983, 'The Origins of Birkeland Currents', *Geophys. Space Phys.* **21**, 125.
Stern, D. P.: 1984, 'Energetics of the Magnetosphere', *Space Sci. Rev.* **39**, 193.
Stokholm, M., Amata, E., Balsiger, H., Candidi, M., Orsini, S., and Pedersen, A.: 1985, 'Low Energy (< 130 eV) Oxygen Ions at the Geosynchronous Orbit During the CDAW 6 Event of March 22, 1979', *Geophys. Res.* **90**, 1253.
Stüdemann, W., Gloeckler, G., Wilken, B., Ipavich, F. M., Kremser, G., Hamilton, D. C., and Hovestadt, D.: 1986, in Y. Kamide and J. A. Slavin (eds.), *Ion Composition of the Bulk Ring Current during a Magnetic Storm: Observations with the CHEM-Instrument on AMPTE/CCE, Solar Wind-Magnetosphere Coupling*, Terra Scientific Publ. Co., Tokyo.
Sugiura, M.: 1984, 'A Fundamental Magnetosphere-Ionosphere Coupling Mode Involving Field-Aligned Currents as Deduced from DE-2 Observations', *Geophys. Res. Letters* **11**, 877.
Temerin, M.: 1986, 'Evidence for a Large Bulk Ion Conic Heating Region', *Geophys. Res. Letters* **13**, 1059.
Temerin, M., Cerny, K., Lotko, W., and Mozer, F. S.: 1982, 'Observations of Double Layers and Solitary Waves in the Auroral Plasma', *Phys. Rev. Letters* **48**, 1175.
Ungstrup, E., Klumpar, D. M., and Heikkila, W. J.: 1979, 'Heating of Ions to Superthermal Energies in the Topside Ionosphere by Electrostatic Ion-Cyclotron Waves', *J. Geophys. Res.* **84**, 4289.
Vasyliunas, V. M.: 1972, 'Interrelationship of Magnetospheric Processes', in B. M. McCormac (ed.), *Earth's Magnetospheric Processes*, D. Reidel Publ. Co., Dordrecht, Holland, p. 29.
Vasyliunas, V. M.: 1984, 'Fundamentals of Current Description, Magnetospheric Currents', in T. Potemra (ed.), *AGU Geophysical Monograph 28*, Proc. Chapman Conf. on Magnetosphere Current Systems, Irvington, Virginia, April 5–8, 1983, p. 63.
Waite, J. H., Nagai, T., Johnson, J. F. E., Chappell, C. R., Burch, J. L. Killeen, T. L., Hays, P. B., Carignan, G. R., Peterson, W. K., and Shelley, E. G.: 1985, 'Escape of Suprathermal O^+ Ions in the Polar Cap', *J. Geophys. Res.* **90**, 1619.
Whalen, B. A. and Daly, P. W.: 1979, 'Do Field-Aligned Auroral Particle Distributions Imply Acceleration by Quasi-Static Parallel Electric Fields?', *J. Geophys. Res.* **84**, 4175.
Whalen, B. A. and McDiarmid, I. B.: 1972, 'Observations of Magnetic-Field Aligned Auroral Electron Precipitation', *J. Geophys. Res.* **77**, 191.
Whipple, E. C., 1977, 'The Signature of Parallel Electric Fields in a Collisionless Plasma', *J. Geophys. Res.* **82**, 1525.
Wilhelm, K.: 1980, 'Natural and Artificially Injected Electron Fluxes Near Discrete Auroral Arcs', *ESA* **SP-152**, 407.
Wilhelm, K., Bernstein, W., Kellogg, P. J., and Whalen, B. A.: 1984, 'Acceleration of Electrons in Strong Beam-Plasma Interactions', *Geophys. Res. Letters* **11**, 1176.

Wilhelm, K., Bernstein, W., Kellogg, P. J., and Whalen, B. A.: 1985, 'Fast Magnetospheric Echoes of Energetic Electron Beams', *J. Geophys. Res.* **90**, 491.

Winningham, J. D. and Gurgiolo, C.: 1982, 'DE-2 Photoelectron Measurements Consistent with a Large-Scale Parallel Electric Field over the Polar Cap', *Geophys. Res. Letters* **9**, 977.

Yamamoto, T. and Kan, J. R.: 1985, 'Effects of Potential Distribution Along Field Lines on Field-Aligned Current Density', *J. Geophys. Res.* **90**, 11083.

Yamamoto, T. and Kan, J. R.: 1986a, 'Interruption of Field-Aligned Current Due to Electrostatic Turbulence', *J. Geophys. Res.* **91**, 7119.

Yamamoto, T. and Kan, J. R.: 1986b, 'Relationship between Field-Aligned Current Density and Potential Difference of Current-Driven Double Layer', *J. Geophys. Res.* **91**, 12113.

Yang, W. H. and Kan, J. R.: 1983, 'Generation of Conic Ions by Auroral Electric Fields', *J. Geophys. Res.* **88**, 465.

Yau, A. W., Shelley, E. G., Peterson, W. K., and Lenchyshyn, L.: 1985, 'Energetic Auroral and Polar Ion Outflow at DE-1 Altitudes: Magnitude, Composition, Magnetic Activity Dependence, and Long-Term Variations', *J. Geophys. Res.* **90**, 8417.

Yau, A. W., Whalen, B. A., Peterson, W. K., and Shelley, E. G.: 1984, 'Distribution of Upflowing Ionospheric Ions in the High-Altitude Polar Cap and Auroral Ionosphere', *J. Geophys. Res.* **89**, 5507.

Zanetti, L. J. and Potemra, T. A.: 1986, 'The Relationship of Birkeland and Ionospheric Current Systems to the Interplanetary Magnetic Field', in Y. Kamide and J. A. Slavin (eds.), *Solar Wind-Magnetosphere Coupling,* D. Reidel Publ. Co., Dordrecht, p. 547.

Zanetti, L. J., Potemra, T. A., Iijima, T., Baumjohann, W., and Bythrow, P. F.: 1984, 'Ionospheric and Birkeland Current Distributions for Northward Interplanetary Magnetic Field: Inferred Polar Convection', *J. Geophys. Res.* **89**, 7453.

ACCELERATION OF AURORAL PARTICLES BY MAGNETIC-FIELD ALIGNED ELECTRIC FIELDS*

L. P. BLOCK

Department of Plasma Physics, The Royal Institute of Technology, Stockholm, Sweden

(Received 30 September, 1987)

Abstract. Measurements on the S3-3 and Viking satellites appear to show that at least a large fraction of magnetic field-aligned potential drops are made up of multiple double layers. Solitons and double layers in *U*-shaped potential structures give rise to spiky electric fields also perpendicular to the magnetic field in agreement with satellite measurements. The large scale potential structures associated with inverted *V*-events are built up of many similar short-lived structures on a small scale. Viking measurements indicate that electric fields parallel to the magnetic field are almost always directed upward.

1. Introduction

The nature of auroral particle acceleration processes is still not clarified, although much progress has been achieved during the last decade. One can distinguish three different classes of processes.

(a) *Adiabatic acceleration* during $\mathbf{E} \times \mathbf{B}$ drift from the magnetospheric tail towards auroral geomagnetic field lines. Both Fermi and betatron acceleration occur, the latter being stronger in dipolar fields and the former in extended tail-like fields, see Heikkila (1974) and Tanskanen *et al.* (1987).

However, even in a dipole field some particles will eventually precipitate due to widening loss cone. The precipitation may be enhanced by generation of VLF-waves that pitch-angle scatter electrons into the loss cone, when the anisotropy in the equatorial plane exceeds a certain value (Cornilleau-Wehrlin *et al.*, 1985); Kremser *et al.* 1986).

(b) *Acceleration by waves along magnetic field lines*. This is described in detail by Bryant (1987).

(c) *Acceleration by d.c. electric fields along magnetic field lines.* A d.c. field is here defined as a field with time-scales for variations being long compared to the bounce time of particles along the geomagnetic field lines. If a sufficiently high voltage can be maintained for time periods of that order, it provides the most efficient mechanism for particle acceleration. A survey of possible mechanisms for producing such fields has been given by Fälthammar (1983). Mechanisms for generators driving the associated current systems have been described by Vasyliunas (1970), Taylor and Perkins (1971), Rostoker and Boström (1976), and Block (1983a, b). These mechanisms are due to the fact that inertia, gradient, and curvature drifts depend on the sign of the particle charge. The perhaps most well-known example of that is the Alfvén layer phenomenon.

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

A complete theory of auroral acceleration processes must correctly assess the relative roles of the three processes. It is, e.g., clear that the first process can only be of importance well within the magnetosphere. It can play no role for auroras on field lines reaching the magnetopause. Furthermore, d.c. electric fields must always induce waves through beam-plasma interaction. These waves must then cause diffusion in velocity space distorting the precipitation spectra as described by many authors, see, e.g., Kaufmann *et al.* (1976). The current controversy regarding the relative importance of wave versus d.c. acceleration is due to this fact. The observed precipitation spectra often obey some features indicating d.c. acceleration, but they also reveal important deviations from ideal d.c. accelerated spectra. The multitude of existing wave modes and instabilities in magnetized plasmas makes it extremely diffecult to judge conclusively whether or not wave-particle acceleration alone can account for the observed spectra.

2. Two General Principles

Before going into details of the nature of d.c. acceleration we point out two fundamental facts.

(a) Only electric fields can accelerate particles. Gravity is too weak by several orders of magnitude, and collisions are much too rare. D.c.-fields are much more efficient than a.c.-fields, since the latter must be in resonance with the particles along the entire orbit where the acceleration process is at work, in spite of the varying particle energy.

(b) Acceleration of some particles implies deceleration of others. The latter play the role of the circuit generator. Generator mechanisms are well known and have been described in detail in several papers; see, e.g., Vasilyunas (1970), Taylor and Perkins (1971), Rostoker and Boström (1976), and Block (1983a, b).

3. Mechanisms for Generation of Magnetic-Field Aligned Electric Fields

Reviews of mechanisms that can maintain electric fields parellel to magnetic fields ($E_{\parallel}$) have been published by several authors; see, e.g., Block and Fälthammar (1976) and Fälthammar (1983, 1985, 1986). The most frequently invoked mechanisms are: (i) anomalous resistivity, (ii) double layers, (iii) magnetic mirror supported fields.

3.1. Anomalous resistivity

Anomalous resistivity can be induced in two ways, either by a current above a certain critical limit or by a beam with particle energy well above the plasma thermal energy. The latter has been studied in detail by Papadopoulos and Coffey (1974a, b). They apply their results to the topside ionosphere (altitudes 500–1000 km). Their starting point is the mere existence of keV precipitation. Beam-plasma interaction produces plasma turbulence, which causes enhanced resistivity for currents carried by thermal electrons. $E_{\parallel}$ is, therefore, only a secondary effect which cannot explain the initial acceleration of the beam particles. This mechanism is, therefore, outside the scope of the present paper.

Anomalous resistivity produced by supercritical currents were first considered by

Kindel and Kennel (1971) and have since been invoked by many authors (Papadopoulos, 1977; Kindel *et al.*, 1981; Dum, 1981; Galeev, 1983; Lin and Rowland, 1985). However, a very simple analysis shows that, e.g., a current density of 1 μA m^{-2} driven by a 1 mV m^{-1} electric field dissipates a power of approximately 100 eV particle^{-1} s^{-1} if the particle density is 100 electrons cm^{-3} (Fälthammar, 1983; Block, 1975). That is clearly out of question, and it is therefore very hard to understand why anomalous resistivity is still discussed as a possible cause for acceleration of auroral particles. The above mentioned authors have not at all considered this profuse heating.

The other two mechanisms cannot be discarded by elementary arguments. They are in fact the most promising mechanisms for maintaining d.c. $E_{\|}$.

3.2. Magnetic mirror effects

The magnetic mirror supported $E_{\|}$ implies a balance between the magnetic mirror force and the electrostatic force. If, e.g., the ion pitch-angle distribution is relatively more field-aligned than the electron pitch-angle distribution, an upward electric field will be established. This field will raise the ion mirror altitudes and lower the electron mirror altitudes until quasi-neutrality is obtained. Obviously, a steady $E_{\|}$ requires steady supply of plasma with non-variable pitch-angle distributions. Thus, $E_{\|}$ must be expected to vary in concert with the properties of the injected plasma, with a delay of the order of a few bounce times for typical ions.

The magnetic mirror effect was first proposed in 1963 by Alfvén and Fälthammar (1963). Further developments of the theory have been made by Persson (1966), Knight (1973), Whipple (1977), and Lemaire and Scherer (1983). Important applications of this mechanism to auroral physics have been described by Lyons (1981), Menietti and Burch (1981), Chin and Cornwall (1980), Chiu *et al.* (1981), and Serizawa and Sato (1984).

3.3. Double layers

In double layers (DL) the electrostatic force is balanced by particle inertia. The properties of a DL is determined by the velocity distributions of all particle species streaming through or being reflected at the layer. In a laboratory discharge these velocity distributions can easily be maintained in steady state. Stable DLs are, therefore, commonly observed in the laboratory (Torvén and Andersson, 1979; Coakley *et al.*, 1979; and Baker *et al.*, 1981).

However, it is hard to believe that a steady supply of particles can be maintained in space. A DL is a local phenomenon with typical thickness of the order of some tens of Debye lengths. The supply of particles, on the other hand, is determined globally. Particles arriving simultaneously at a certain point have come from widely separated parts of the magnetosphere, depending on their total kinetic energy, adiabatic invariants and the detailed electric and magnetic field distributions along their drift orbits.

These considerations indicate that, if DLs are significantly contributing to $E_{\|}$ above the aurora, the most likely configuration is multiple, short-lived DLs which statistically sum up to a more slowly varying total potential drop from the ionosphere to the equatorial plane or top of the region with finite $E_{\|}$. This idea was suggested by Block

(1972) and later supported by Collin *et al.* (1982), who found that S3-3 observations of electron beams were best accounted for in this way. Furthermore, direct evidence of these multiple DLs was obtained by the electric field experiments on S3-3 (Temerin *et al.*, 1982) and on Viking (Holback *et al.*, 1986).

4. Electric Field Measurements on S3-3 and Viking

The most extensive measurements of electric fields on auroral field lines have been made with the S3-3 and Viking satellites. Some significant results obtained so far include the following.

4.1. Small Double Layers

The small DLs with potential drops of the order of a few volts, first discovered on S3-3 often exhibit small potential wells on the low potential side, as schematically illustrated in Figure 1 (Temerin *et al.*, 1982; and Holback *et al.*, 1986). Block (1983a, b) argued that

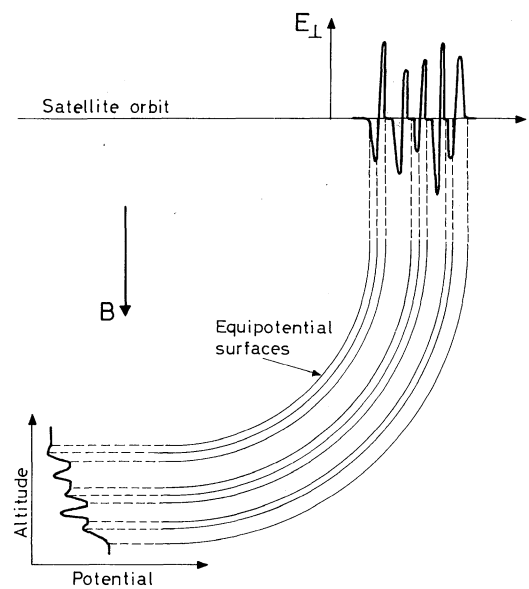


Fig. 1. *Lower left*: schematic picture of potential variation along a magnetic field line with small double layers and solitons. *Upper right*: electric field perpendicular to the magnetic field ($E_\perp$) along the satellite orbit if the magnetic-field-aligned potential can be projected along U-shaped equipotential surfaces.

these wells could explain observations showing that some of the precipitating electrons in auroral arcs have been accelerated through only part of the potential drop (Whalen and Daly, 1979). Temerin *et al.* (1982) suggest that the DL formation begins with a solitary wave consisting of a well with no net potential drop. The well on the low-potential side is then believed to be a residue of the solitary wave. Figure 1 shows the potential distribution along a magnetic field line with some DLs and solitons. Projection of this potential distribution along U-shaped equipotential surfaces would give very spiky electric fields perpendicular to the magnetic field as illustrated in Figure 1 (provided the electric field has a potential). The parallel electric fields seem to be distributed over 1000 km or more in altitude (Cattell, 1983), whereas the corresponding perpendicular fields are compressed in thinner regions of the order of 100 km or less. Hence, the perpendicular fields should have much larger magnitudes than the typical parallel fields.

4.2. Spiky Electric Fields and Large-Scale Potential Minimum over the Auroral Oval

The above inferences are corroborated by results from Viking. Figure 2 shows the electric field components in the satellite's spin plane, essentially downward and southward, and also the potential variation along the orbit during a pass over the evening side auroral oval during rather quiet conditions. The nearly southward component E_2 (middle panel) is defined to be strictly perpendicular to the magnetic field. It is also nearly along the satellite velocity direction. E_1 (top panel) is the component along the projection of the magnetic field onto the spin plane. The angle between the magnetic field and the spin plane was small during the interval displayed in Figure 2 ($\langle$ 8 deg).

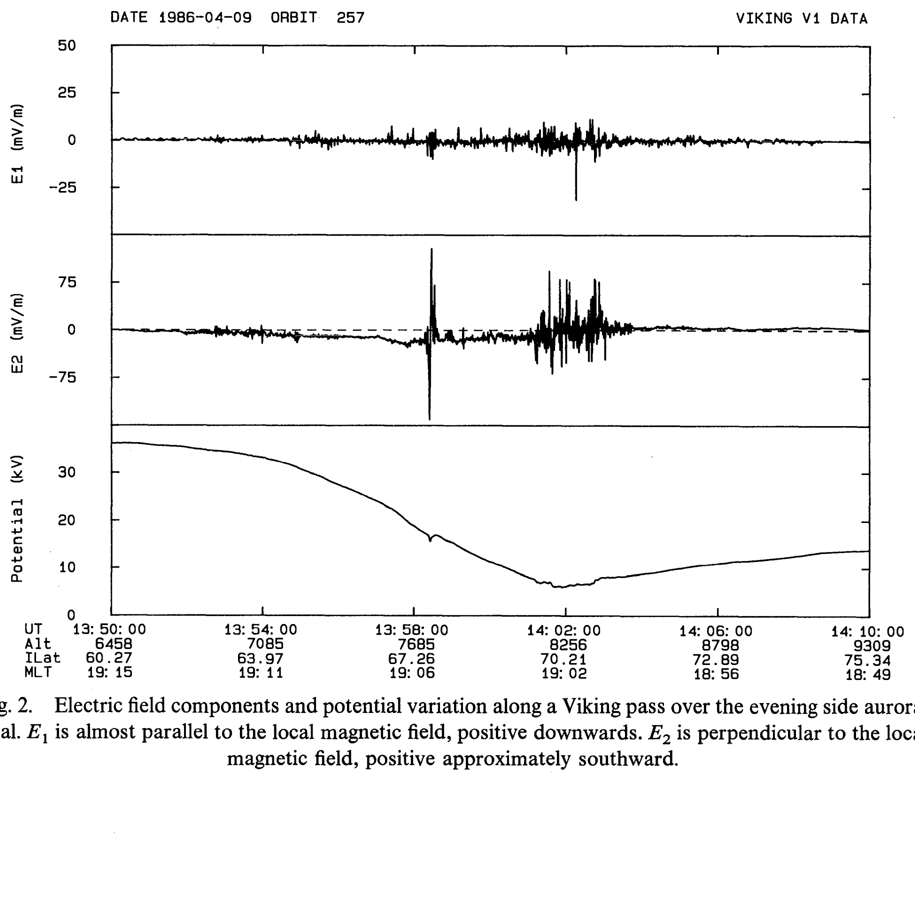


Fig. 2. Electric field components and potential variation along a Viking pass over the evening side auroral oval. E_1 is almost parallel to the local magnetic field, positive downwards. E_2 is perpendicular to the local magnetic field, positive approximately southward.

Hence, E_1 is positive downward and nearly parallel to the magnetic field. Unfortunately, the third component E_3 along the spin axis could not be measured due to a short circuit of one of the axial probes, which occurred already at launch.

Figure 2 illustrates three important agreements with the discussion in Section 4.1.

(a) E_1 and E_2 show very spiky structures.

(b) The magnitude of the spikes in E_2 is an order of magnitude larger than those in E_1.

(c) The potential along the orbit has a minimum within the oval. The depth of this minimum should correspond to the maximum potential drop below the satellite.

4.3. Potential minima with different scale lengths

Figure 3 shows the potential distribution along two evening-morning Viking passes. Note the difference in shape between the regions with depressed potential in the two

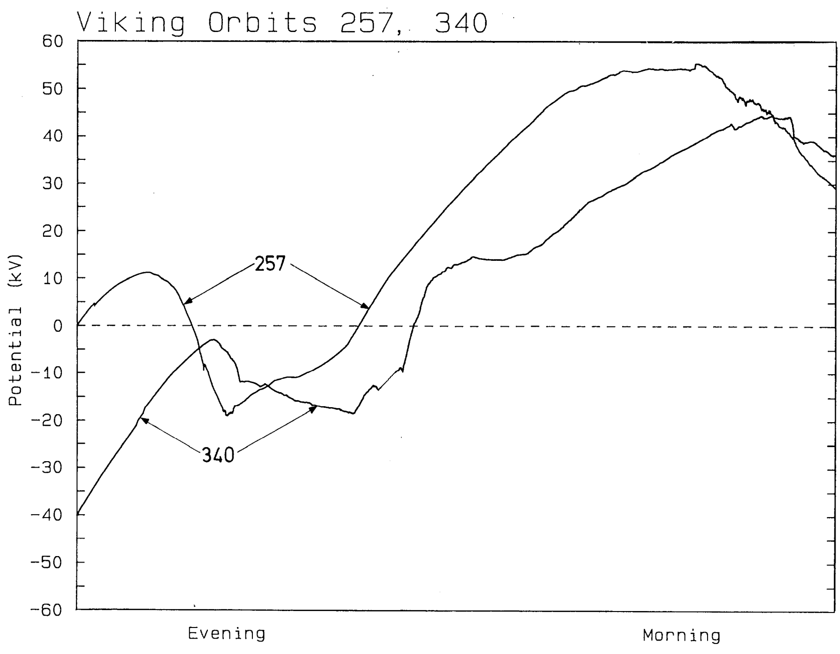


Fig. 3. Potential variation along two Viking passes over the polar regions, including the evening and morning sides of the auroral oval. Polar cap auroral activity was much stronger during pass 340.

orbits 257 (weak activity) and 340 (strong activity). The deepest potential minimum on orbit 340 occurs at a high latitude, as far north as about 79° invariant latitude. Note also that in addition to the main large-scale minimum there are several small-scale minima. Figure 4 shows details on a much shorter time-scale of such potential minima. There are small-scale dips within dips of larger scales with widths ranging from a

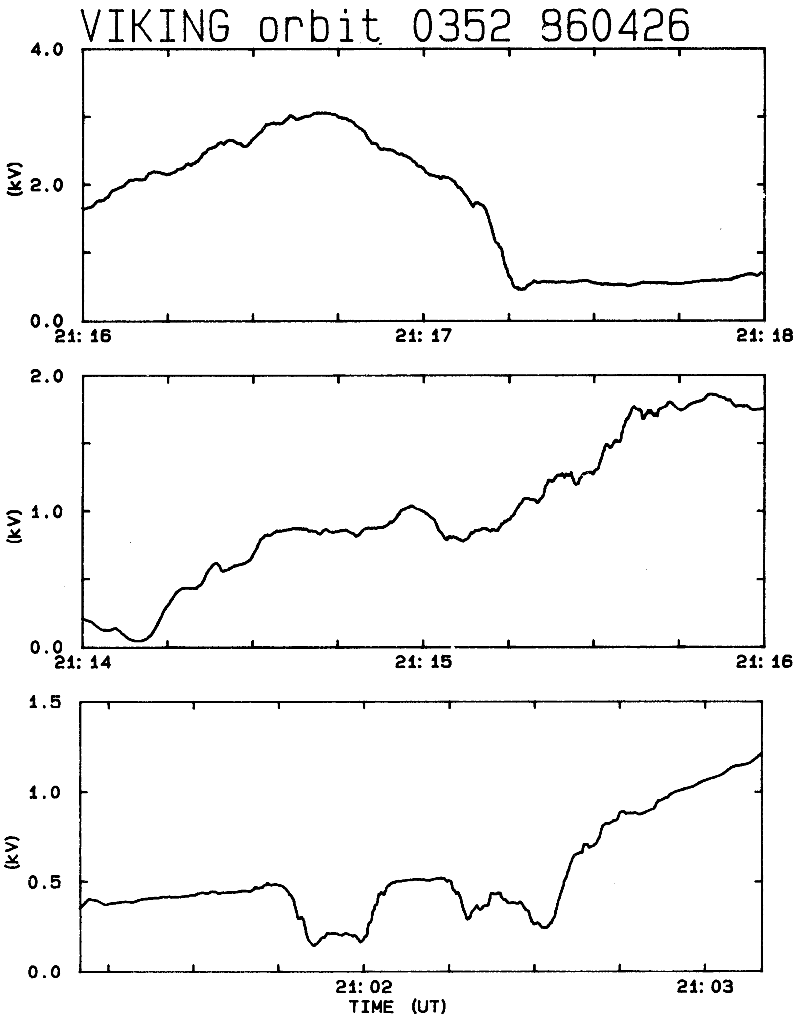


Fig. 4. Small scale potential minima, obtained from Viking electric field data during a pass over the morning side auroral oval.

fraction of a second to minutes. This is typical of all passes over the auroral oval. Figure 5 shows the southward electric field E_2 around some similar minima with four different time-scales, corresponding to scale lengths in the ionosphere of roughly 40, 3, 0.5, and 0.1 km.

We conclude, tentatively, that the large-scale U-shaped potential structures, corresponding to inverted V-events, are made up of several similar small-scale structures, possibly in a hierarchy of scales. The basic assumption behind this is of course that the electric fields are essentially due to charge separation, but they must not necessarily be

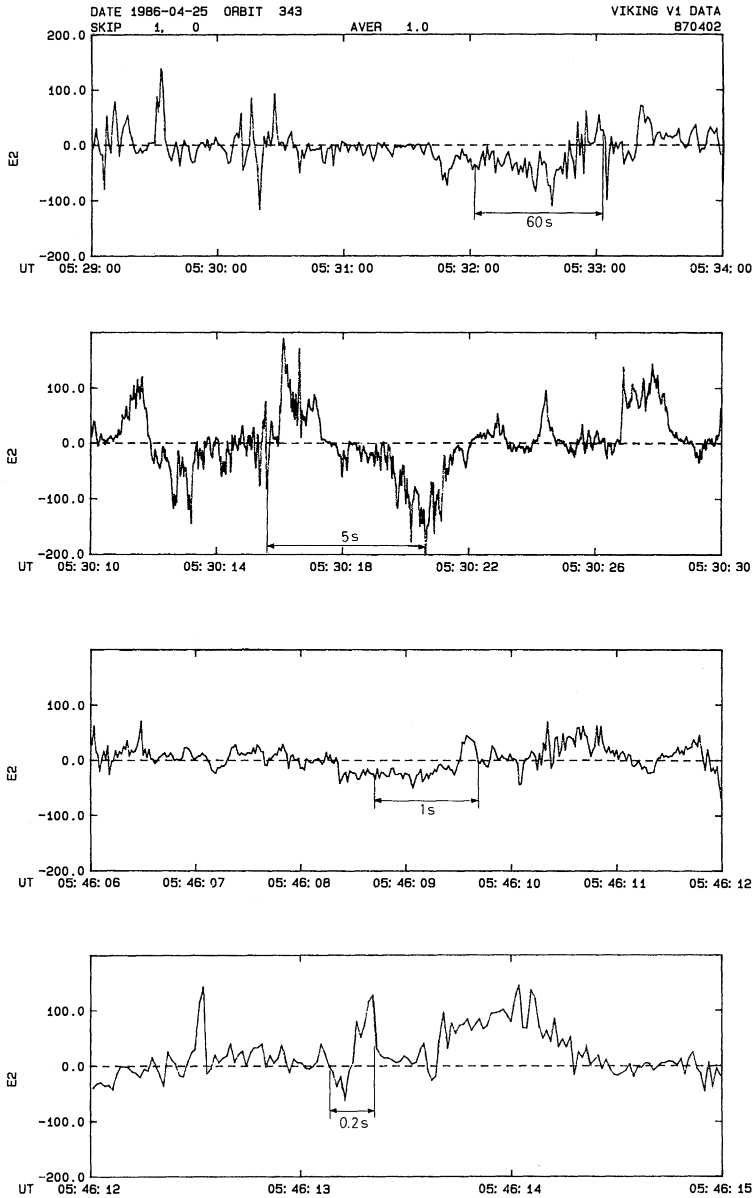


Fig. 5. The electric field perpendicular to the magnetic field, observed by Viking on auroral field lines. Electric field reversals with four different time scales are indicated.

stationary. On the contrary, it was argued in Section 3.3 that the DLs may be short-lived, but the total potential drop should be relatively constant along given magnetic field-lines, containing hundreds or thousands of small DLs. The small-scale perpendicular electric field reversals shown in Figure 5 may, therefore, correspond to DLs in the immediate neighbourhood, while larger-scale reversals corresponds to larger numbers of DLs in series within progressively larger altitude ranges below the satellite.

If this discussion is essentially correct, it means that the U-shaped equipotential surfaces are shaky, trembling, or oscillatory on a small scale but rather steady on a large scale.

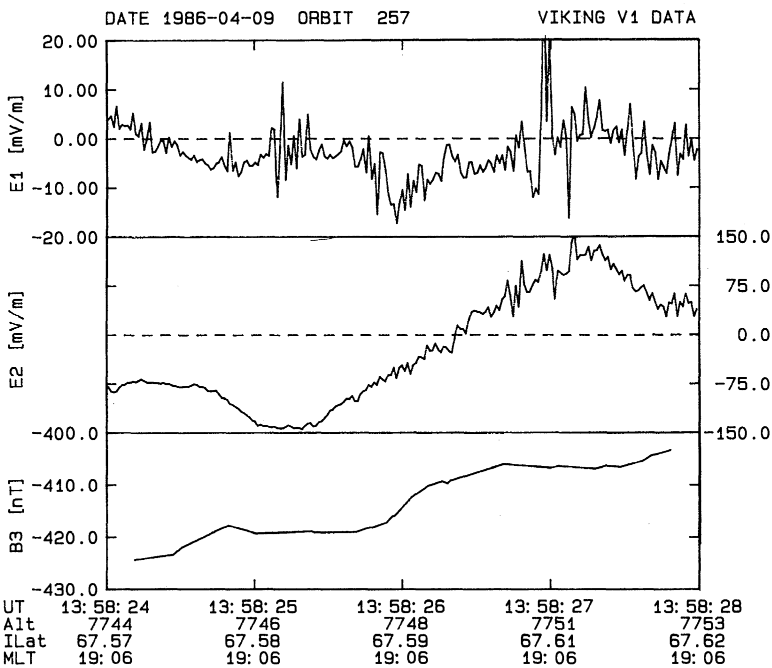


Fig. 6. Electric and magnetic field variations during a Viking pass over an evening side auroral arc. E_1 and E_2 have the same meaning as in Figure 2. B_3 is the westward magnetic field component perpendicular to the main magnetic field. An upward parellel electric field of about 17 mV m^{-1} is co-located with the maximum B_3 variation, corresponding to a Birkeland current density of 50 μA m^{-2} in the ionosphere.

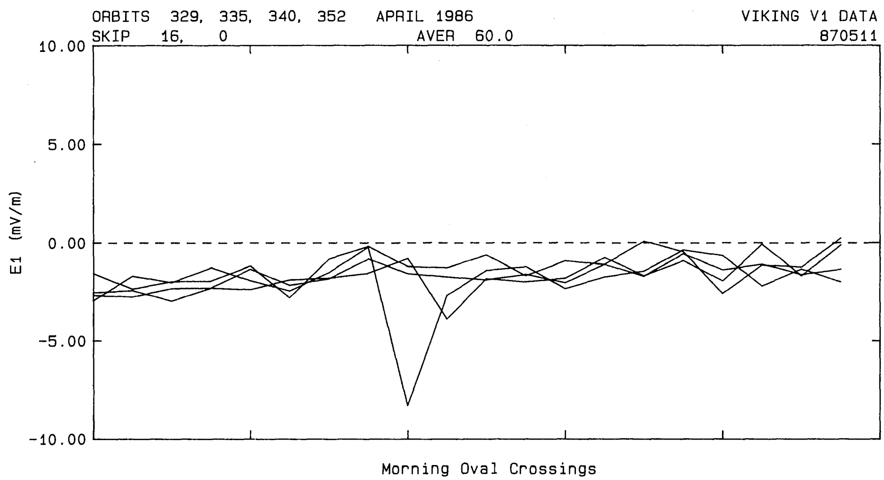
An unusually beautiful example of an intermediate scale electric field reversal is shown in Figure 6. The E_1 electric field component makes an angle of only 4 deg with the magnetic field, so it is most likely due to an upward directed $E_{\|}$ (E_1 is positive downward) of the order of 10 mV m^{-1}. The E_2 reversal occurs on a scale of a few seconds. The corresponding potential minimum is clearly seen in Figure 2. It is roughly 1 kV deep. The size of the dip is some 30–40 km at the satellite and about 10 km in the ionosphere.

The bottom panel of Figure 6 shows the westward magnetic field component B_3 (nearly perpendicular, within 4 deg, to the main magnetic field). With the usual interpretation in terms of a current sheet, the steep increase in B_3 around 13 : 58 : 26 UT corresponds to an upward current density of a s much as 50 μA m^{-2} in the ionosphere.

A more detailed description of this event is given in Block *et al.* (1987).

4.4. Predominance of upward directed electric fields

The geomagnetic mirror field and the presence of the dense ionospheric plasma imply that downward directed electric fields along the magnetic field are expected to be much weaker than upward directed fields. It was mentioned in Section 4.2. that one of the electric field probes was short-circuited at launch, which means that only the two components in the spin plane could be measured. Fortunately, the angle between the magnetic field and the spin plane is usually small (< 10 deg) during most passes over the auroral oval. Hence, E_1 is usually almost parallel to the magnetic field. The influence



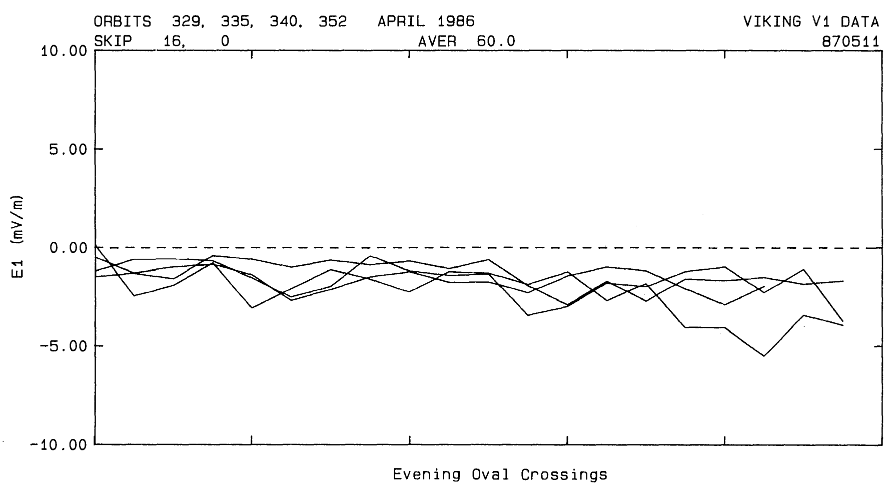


Fig. 7. Sixty second average values of E_1 (see caption for Figure 2) are almost always negative on auroral field lines, indicating that magnetic-field-aligned electric fields are predominantly directed upwards.

of east-west electric fields on E_1 should, therefore, be rather small. Moreover, there is no reason why eastward directed fields should dominate over westward fields, or *vice versa*. Thus, if E_1 is, on average, negative over the oval, it would strongly indicate, both that E_1 is often due to parallel electric fields and that these are predominantly directed upwards.

The data shown in Figure 7 do indeed support this concept. The curves show 60 s averages of E_1 for 8 different oval crossings, 4 on the morning side and 4 on the evening side. All points on these curves are negative except for a few points on the morning side, which are zero. No positive points are found. Downward fields seem to exist only on time-scales less than 1 min.

5. Summary and Conclusions

Viking and S3-3 data have been compared with early theories of d.c. electric field acceleration of auroral particles. It appears that:

(a) Most or at least a large fraction of the potential drop along auroral magnetic field lines is due to a large number of double layers. Individual double layers may be short-lived, but the total potential drop should stay rather constant during the lifetime of the associated aurora.

(b) The double layers have negative potential wells on the low potential side.

(c) The spiky nature of the electric field, both $E_\parallel$ and $E_\perp$, and the potential distribution over the auroral oval agree with the concept of *U*-shaped equipotential surfaces with multiple double layers and solitons on auroral field lines.

(d) The large scale *U*-shaped equipotential surfaces are built up of small scale similarly *U*-shaped structures, probably in a hierarchy of scale lengths.

(e) Measurable $E_\parallel$ is predominantly directed upward.

The above conclusions are not absolutely compelling. It appears, however, that the measurements on S3-3 and Viking surprisingly well support early hypothesis that d.c. fields are due to multiple double layers (Block 1972, 1975, 1983a, b).

A complete theory of auroral particle acceleration by multiple double layers must be extremely complicated. It must, among other features, also include the effects of energetic backscattered and trapped particles, as well as the interaction with the rest of the current circuit. This has been stressed by Alfvén (1981) and Kan and Lee (1980).

A correct assessment of the relative roles of waves and d.c. fields in the auroral acceleration process cannot yet be done. It will require full vector measurements on at least four satellites, flying relatively close to each other, as in the proposed project 'IMPACT'.

Acknowledgements

I wish to thank Per-Arne Lindqvist, Ramesh Mehra, Bengt Harald Nilsson, and Lars Blomberg for their extremely competent and hard work with the analysis of the vast amount of data from the electric field experiment on Viking. Thanks are also due to Professor Carl-Gunne Fälthammar and Dr Göran Marklund for valuable discussions.

References

Alfvén, H.: 1981, *Cosmic Plasma*, D. Reidel Publ. Co., Dordrecht, Holland, Plenum Press, New York, p. 229.

Alfvén, H. and Fälthammar, C.-G.: 1963, *Cosmical Electrodynamics, Fundamental Principles*, Clarendon Press, Oxford.

Baker, K. C., Singh, N., Block, L. P., Kist, R., Kampa, W., and Thiemann, W.: 1981, *J. Plasma Phys.* **26**, 1.

Block, L. P.: 1972, *Cosmic Electrodynamics* **3**, 349.

Block, L. P.: 1975, 'Double Layers' in B. Hultqvist and L. Stenflo (eds.), *Physics of the Hot Plasma in the Magnetosphere*, Plenum Press, New York, p. 229.

Block, L. P.: 1983a, 'Three-Dimensional. Potential Structure Associated with Birkeland Currents', in T. A. Potemra (ed.), *Magnetospheric Currents*, AGU Geophysical Monograph 28, American Geophysical Union, Washington D.C., p. 315.

Block, L. P.: 1983b, 'Three-Dimensional Potential Structure Above Auroral Arcs', in *Proc. Sixth ESA Symp. on European Rocket and Balloon Programmes*, ESA SP-183, p. 247.

Block, L. P. and Fälthammar, C.-G.: 1976, *Ann. Geophys.* **32**, 161.

Block, L. P., Fälthammar, C.-G., Lindqvist, P.-A., Marklund, G., Mozer, F. S., Pedersen, A., Potemra, T. A., and Zanetti, L. J.: 1987, *Geophys. Res. Letters* **14**, 435.

Bryant, D. A.: 1987, 'Wave Acceleration of Auroral Electrons', in *Proc. Eighth ESA Symp. on European Rocket and Balloon Programmes and on Related Research*, ESA SP-270, p. 273.

Cattell, C.: 1983, 'Association of Field-Aligned Currents with Small-Scale Auroral Phenomena', in *Magnetospheric Currents*, Geophysical Monograph, No. 28, p. 304.

Chin, Y. T. and Cornwall, J. M.: 1980, *J. Geophys. Res.* **85**, 543.

Chiu, Y. T., Newman, A. L., and Cornwall, J. M.: 1981, *J. Geophys. Res.* **86**, 10029.

Coakley, P., Johnson, L., and Herskowitz, N.: 1979, *Phys. Letters* **70A**, 425.

Collin, H. L., Sharp, R. D., and Shelley, E. G.: 1982, *J. Geophys. Res.* **87**, 7504.

Cornilleau-Wehrlin, N. J., Solomon, N. J., Korth, A., and Kremser, G.: 1985, *J. Geophys. Res.* **90**, 4141.

Dum, C. T.: 1981, 'Anomalous Resistivity and Plasma Dynamics', in S.-I. Akasofu and J. R. Kan (eds.), *Physics of Auroral Arc Formation*, AGU Geophysical Monograph 25, American Geophysical Union, Washington, D.C., p. 408.

Fälthammar, C.-G.: 1983, *ESA J.* **7**, 385.

Fälthammar, C.-G.: 1985, 'Magnetosphere-Ionosphere Coupling', *Proc. ESA Workshop on Future Missions in Solar, Heliospheric and Space Plasma Physics*, Garmisch-Partenkirchen, 30 April–3 May, 1985, ESA SP-235, p. 107.

Fälthammar, C.-G.: 1986, *IEEE Trans on Plasma Science* **PS-14**, 616.

Galeev, A. A.: 1983, 'Anomalous Resistivity on Auroral Field Lines', in B. Hultqvist and T. Hagfords (eds.), *High-Latitude Space Plasma Physics*, Plenum Press, New York and London, p. 437.

Heikkila, W. J.: 1974, *J. Geophys. Res.* **79**(16), 2496.

Holback, B., Boström, R., Gustafsson, G., Koskinen, H., Holmgren, G., and Kintner, P.: 1986, *EOS* **67**, 1156.

Kan, J. R. and Lee, L. C.: 1980, *J. Geophys. Res.* **85**, 788.

Kaufmann, R. L., Walker, D. N., and Arnoldy, R. L.: 1976, *J. Geophys. Res.* **81**, 1673.

Kindel, J. M. and Kennel, C. F.: 1971, *J. Geophys. Res.* **76**, 3055.

Kindel, J. M., Barnes, C., and Forslund, D. W.: 1981, 'Anomalous DC Resistivity and Double Layers in the Auroral Ionosphere', in S.-I. Akasofu and J. R. Kan (eds.), *Physics of Auroral Arc Formation*, AGU Geophysical Monograph 25, American Geophysical Union, Washington D.C., p. 296.

Knight, S.: 1973, *Planetary Space Sci* **21**, 741.

Kremser, G., Korth, A., Ullaland, S., Stadnes, J., Baumjohann, W., Block, L., Torkar, K. M., Riedler, W., Aparicio, B., Tanskanen, P., Iversen, I. B., Cornilleau-Wehrlin, N., Solomon, J., and Amata, E.: 1986, *J. Geophys. Res.* **91**, 5711.

Lemaire, J. and Scherer, M.: 1983, *Ann. Geophys.* **1**, 91.

Lennartsson, W.: 1980, *Planetary Space Sci.* **28**, 135.

Lin, C. S. and Rowland, H. L.: 1985, *J. Geophys. Res.* **90**, 4221.

Lyons, L. R.: 1981, 'The Field-Aligned Current versus Electric Potential Relation and Auroral Electrodynamics', in S.-I. Akasofu and J. R. Kan (eds.), *Physics of Auroral Arc Formation*, AGU Geophysical Monograph 25, Washington D.C., p. 252.

Menietti, J. D. and Burch, J. L.: 1981, *Geophys. Res. Letters* **8**, 1095.
Papadopoulos, K.: 1977, *Rev. Geophys. Space Phys.* **15**, 113.
Papadopoulos, K. and Coffey, T.: 1974a, *J. Geophys. Res.* **79**, 674.
Papadopoulos, K. and Coffey, T.: 1974b, *J. Geophys. Res.* **79**, 1558.
Persson, H.: 1966, *Phys. Fluids* **9**, 1090.
Rostoker, G. and Boström, R.: 1976, *J. Geophys. Res.* **81**, 235.
Serizawa, Y. and Sato, T.: 1984, *Geophys. Res. Letters* **11**, 595.
Tanskanen, P., Kangas, J., Block, L., Kremser, G., Korth, A., Woch, J., Iversen, I. B., Torkar, K. M., Riedler, W., Ullaland, S., Stadnes, J., and Glassmeier, K.-H.: 1987, *J. Geophys. Res.* **92**, 7443.
Taylor, H. E. and Perkins, F. W.: 1971, *J. Geophys. Res.* **76**, 272.
Temerin, M., Cerny, K., Lotko, W., and Shelley, E. G.: 1982, *Phys. Rev. Letters* **48**, 1175.
Torvén, S. and Andersson, D.: 1979, *Phys. D: Appl. Phys.* **12**, 717.
Vasyliunas, V. M.: 1970, 'Mathematical Models of Magnetospheric Convection and its Coupling to the Ionosphere, in B. M. McCormac (ed.), *Particles and Fields*, D. Reidel Publ. Co., Dordrecht, p. 60.
Whalen, B. A. and Daly, P. W.: 1979, *J. Geophys. Res.* **84**, 4175.
Whipple, E. C.: 1977, *J. Geophys. Res.* **82**, 1525.

PARALLEL ELECTRIC FIELDS ACCELERATING IONS AND ELECTRONS IN THE SAME DIRECTION*

BENGT HULTQVIST and RICKARD LUNDIN

Swedish Institute of Space Physics, Kiruna, Sweden

(Received 24 June, 1987)

Abstract. In this contribution we present Viking observations of electrons and positive ions which move upward along the magnetic field lines with energies of the same order of magnitude. We propose that both ions and electrons are accelerated by an electric field which has low-frequency temporal variations such that the ions experience an average electrostatic potential drop along the magnetic field lines whereas the upward streaming electrons are accelerated in periods of downward pointing electric field which is quasi-static for the electrons and forces them to beam out of the field region before the field changes direction.

1. Introduction

The first ideas about the aurora being the result of an electrical discharge around the Earth, and thus about electric fields playing a major role, date back to Lomonosov and Canton in the eighteenth century. However, Alfvén seems to have been the first to specify the role of the electric field and to introduce an electric field component along the magnetic field lines – generally called 'parallel electric fields' for brevity – as a major source of acceleration of the auroral energetic particles. He did this in his *Theory of Magnetic Storms*, I, II, III in 1939 and 1940. Later he extended his arguments in favour of the importance of parallel electric fields in space around the Earth on the basis of experience from laboratory experiments (Alfvén, 1955, 1958) and electrostatic double layers were also proposed to be of importance in space (see Alfvén, 1972).

Chapman and several of his collaborators were critical of Alfvén's 'electric field theory' of aurora and magnetic storms. Cowling (1942) wrote a special paper objecting to it. A common argument against the theory was that the conductivity along the magnetic field lines in an above the ionosphere is so high that any space charges and associated parallel electric fields will be eliminated very effectively. Chapman (1969) maintained this criticism of the role of parallel electric fields in aurora. The parallel electric field concept was thus dismissed as being contrary to generally accepted theories and it was only the direct measurements in space in the 1960s and 1970s that finally changed the situation completely. In the U.S.A. *Review and Quadrennial Report to the IUGG* in 1979 Stern (1979) states that "in the period 1975–1978... it became increasingly evident that the condition $E_{\parallel} = 0$... is often grossly violated in the magnetosphere. It would be only fair to say that the period marked a transition from general scepticism concerning the role of $E_{\parallel}$ in the magnetosphere to the acceptance of $E_{\parallel}$ as an essential ingredient in such phenomena as discrete auroras...".

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 149–154.

Among the many observational results that paved the way for the general acceptance of the importance of parallel electric fields, some were of special importance in breaking the resistance. When Evans (1974) eliminated the problem that the existence of an intense low-energy tail of the observed peaked electron spectra constituted, one big stumbling block disappeared. The S3-3 satellite, launched in 1976 gave many new results which supported the importance of a parallel electric field component in accelerating ions out of the upper ionosphere (Shelley *et al.*, 1976) and in driving Birkeland currents (e.g., Potemra, 1979). S3-3 also provided direct electric field observations strongly suggesting the existence of electric fields with $E_{\parallel} \neq 0$ above the auroral oval (Mozer *et al.*, 1977).

That the existence of parallel electric fields now is proved beyond reasonable doubt, does not, of course, mean that field aligned acceleration is the only important acceleration process. The observations – also by means of S3-3 – of ions accelerated primarily transverse to the magnetic field lines – so called ion conics – were reported first by Sharp *et al.* (1977). The ion conics are sometimes ‘elevated’ so that the lowest energy ions are magnetic field aligned (Klumpar *et al.*, 1984; Horwitz, 1986; Temerin, 1986). The existence of upward accelerated electron beams on the same field lines along which ion conics occur has been mentioned by several authors (Sharp *et al.*, 1980; Klumpar and Heikkila, 1982; Collin *et al.*, 1982; Lin *et al.*, 1982; Burch *et al.*, 1983; Kintner and Gorney, 1984). Here we will demonstrate the simultaneous observations by means of the Swedish satellite Viking of upward moving electrons and ions of similar parallel energy. This appears at first sight difficult to understand in terms of parallel electric field acceleration. In this brief contribution in honour of Hannes Alfvén we will outline how a parallel electric field can accelerate both ions and electrons upward along the magnetic field lines above the auroral zone ionosphere. The present report is only a preliminary one. A study using all available data from Viking is under way.

2. Observations

The Viking data to be discussed below are shown in Figure 1. It contains two colour spectrograms showing the counts per readout from three of the 11 spectrometers in the particle experiment on Viking. For a description of the experiment see Sandahl *et al.* (1985). The electron data were obtained with an electrostatic analyser covering an energy range of 0.01–40 keV and a field of view of $5° \times 5°$. With the satellite in cartwheel mode the spectrograph covered, in the period of interest, all pitch angles except $6°$ around the field line with an angular resolution of $\sim 5°$. The 100 km mirroring loss cone had a width of $11–12°$ at the satellite. The ion data are from two electrostatic analysers covering the energy ranges 0.04–1.2 keV and 1.2–40 keV, respectively. Their angular resolutions were $4°$ and $6°$. The pitch angles at which the measurements were made are shown in the bottom panel ($0°$ means downward moving particles and $180°$ upward moving). The energy scales are given along the lefthand vertical edge and below the panels are shown universal time, UT, satellite altitude, H, invariant latitude, I-LAT, and magnetic local time, MLT.

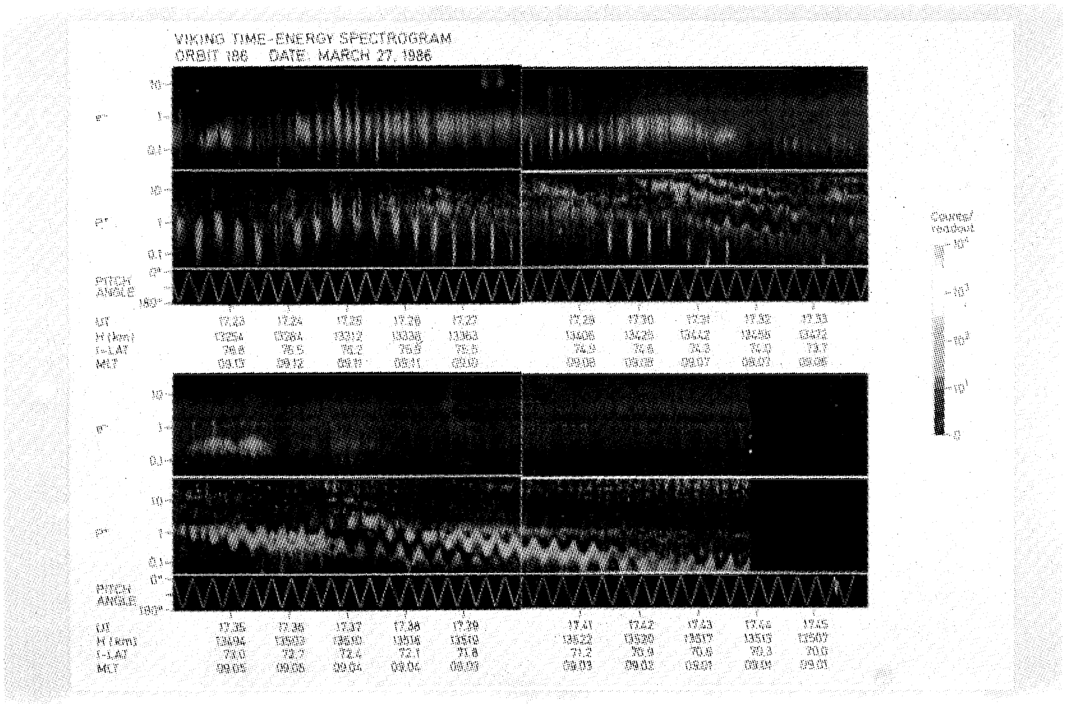


Fig. 1. Energy-time spectrograms for electrons and positive ions obtained from the particle experiment on Viking. Each spectrogram displays counts accumulated in 32 energy bins during 0.6 s, corresponding to a pitch angle resolution of 5°. The lower panel shows the pitch angle of the observations versus time. The position of the satellite is given by altitude (H), invariant latitude (I-LAT), and magnetic local time (MLT).

The data shown in Figure 1 are interesting for this study because it shows in the period 17 : 24–17 : 26 UT strongly elevated ion conics together with upward moving electron beams, some of them quite narrow in angular width. The "elevation" of the conics reaches almost 1 keV in the middle of the period mentioned. The ion fluxes along the magnetic field lines at ~ 1 keV in the conics where of the order of 6×10^5 (cm^2 s sr keV)$^{-1}$, whereas the upward electron flux reached more than 10^9 (cm^2 s sr keV)$^{-1}$ below 1 keV in the most intense beams and more than 10^8 (cm^2 s sr keV)$^{-1}$ at 1 keV. Data from ion mass spectrometers on Viking show that the ions in Figure 1 were mainly hydrogen ions.

These observations appear to be the first reported of ions *and* electrons moving upward along auroral zone magnetic field lines. They were obtained in a fairly quiet period a few hours after a series of substorms. AE was < 100 nT, K_p was 3- and Dst = -37 nT. There are many other examples of this kind in the Viking data base, but they will be reported in a later study.

3. Discussion

Klumpar and Heikkila (1982) reported observations from the ISIS-2 spacecraft at quite low altitudes of narrow upward moving electron beams with very soft energy spectra. The peak flux was below 100 eV and sometimes even below 10 eV. In a few cases they saw also perpendicularly accelerated ions, which they suggested had been accelerated above the altitude range with a parallel electric field. Since the observations were made in the upper ionosphere the potential drop had to be achieved over a narrow altitude range and thus $E_{\|}$ had to be so high that runaway electrons are produced (according to Dreicer, 1959, 1960). They did not report field aligned ion fluxes. Collin *et al.* (1982), Lin *et al.* (1982), Burch *et al.* (1983), and Kintner and Gorney (1984) also observed upward electron beams together with ion conics but did not report any field aligned ions.

Klumpar *et al.* (1984) were the first to report 'elevated' ion conics, i.e., ion distributions which are field aligned at low energies and conical at higher energies (ion bowl distributions according to Horwitz (1986)). They interpreted the distributions in terms of both perpendicular and parallel (later, at greater altitudes) acceleration of the ions. Horwitz (1986) showed that bowl-shaped distributions can also be formed by transverse heating in a region of finite horizontal extent, followed by essentially adiabatic convective flow to the observation location. Temerin (1986) found that the ion bowl distribution – or the 'elevated' ion conics as they are called above – can be understood in terms of only perpendicular acceleration, provided the perpendicular heating occurs in a broad altitude range below the satellite and provided the 'elevation' is not too large. Gorney *et al.* (1985) discussed a possible trapping of ion conics in downward parallel electric fields. None of these authors dealt with simultaneous observations of 'elevated' conics (ion bowl distributions) and upward moving electron beams. The observations shown in Figure 1, therefore, appear to be the first reported. None of the explanations proposed by the different above-mentioned authors appears to be able to explain the data in Figure 1 for the following reasons: With the mechanism of Klumpar *et al.* (1984) no upgoing electrons would be seen by Viking. The presence of a downward electrical potential fall of a magnitude of several hundred eV, which accelerates the beam electrons upward, would eliminate or strongly reduce the 'elevation' of ion conics formed in the ways suggested by Horwitz (1986) and Temerin (1986). The upward narrow electron beams occur only together with the 'elevated' ion conics in Figure 1 and as these narrow electron beams very clearly indicate the existence of a downward directed parallel electric field in this period (but not before or after) we would expect that the bottom of the conics would be at lower energies in the period 17 : 24–17 : 26 than before or after, rather than at higher energies as has been observed. According to Horwitz' (1986) mechanism the plasma convection would be in opposite directions before and after $\sim$17 : 25 UT which hardly seems reasonable. In addition, electric field measurements indicate (personal communication by Block and Lindqvist and by Gustafsson and Koskinen) that there existed in the period of elevated ion conics very strong broadbanded low-frequency turbulence but not before or after, indicating that Viking was on field lines which passed the transverse heating region, in which case Horwitz' mechanism is not in agreement with the particle observations. Even without a downward potential drop the observed large elevation of the conics at the center of the region (0.7–0.8 keV) may be more than Temerin's mechanism can produce (Temerin, personal communication) and with a downward potential drop of a few hundred eV it is even more unlikely that his mechanism can give rise to the observed distributions.

We propose the following model for the production of coincident upward flow of ions and electrons. The parallel electric field has a broad variation spectrum superimposed on an upward directed DC field. A very strong low frequence electric field turbulence was observed together with the elevated ion beams, as mentioned above. The effect of the electric field variations on an ion moving through the region largely averages out, so that the upward moving ions mainly increase their parallel kinetic energy with the potential change along the magnetic field lines. The variation spectrum of the parallel electric field is supposed to have the largest amplitudes at such frequencies that the electric field is effectively quasi-static for the electrons, but not for the ions. When these quasi-static parallel field components point downward the electrons are accelerated upwards to such high velocities that they leave the region with an electric field before the parallel field component switches direction. It is illuminating that an electron that has been accelerated through a potential increase of 100 V will pass a distance of 600 km along the field line in the next 100 ms, whereas a proton will only go 14 km in the next 100 ms after it has achieved the energy 100 eV.

Thus, whereas the ions are on the average accelerated only by the average potential change, which is directed upward, the upgoing electrons are accelerated also by larger but short lived downward directed fields, which are quasi-static for the electrons but not for the ions. The satellite only sees upward moving electrons when it is above the acceleration region and it cannot resolve the temporal variations in the electron flux.

Bryant *et al.* (1978) discussed possible consequences of a fluctuating electrostatic field in a different context of acceleration of primary auroral electrons.

The data shown in this brief report certainly do not prove the correctness of the proposed mechanism, which is speculative at the present state, put forward because no other proposed mechanism seems to be able to explain the observations. An investigation, in which data from all relevant experiments on Viking will be used, is under way.

Acknowledgements

The Swedish satellite Viking as well as the V3-K particle experiment was financed by grants from the Swedish Board for Space Activities. The Viking project was managed by the Swedish Space Corporation.

References

Alfvén, H.: 1939, *Kungl. Svenska Vet. Handl. 3* **18**, No. 3.
Alfvén, H.: 1940, *Kungl. Svenska Vet. Handl. 3* **18**, No. 9.
Alfvén, H.: 1955, *Tellus* **7**, 50.
Alfvén, H.: 1958, *Tellus* **10**, 104.
Alfvén, H.: 1972, in K. Schindler (ed.), *Cosmic Plasma Physics*, Plenum Press, New York, pp. 1–14.

Bryant, D. A., Hall, D. S., and Lepine, D. R.: 1978, *Planetary Space Sci.* **26**, 81.
Burch, J. L., Reiff, P. H. and Sugiura, M.: 1983, *Geophys. Res. Letters* **10**, 753.
Chapman, S.: 1969, in B. M. McCormac and A. Omholt (eds.), *Atmospheric Emissions*, Van Nostrand Reinhold Co., New York, pp. 11–26.
Collin, H. L., Sharp, R. D., and Shelley, E. G.: 1982, *J. Geophys. Res.* **87**, 7504.
Cowling, T. G.: 1942, *Terrest. Magn. Atmospheric Electr.* **47**, 209.
Dreicer, H.: 1959, *Phys. Rev.* **115**, 238.
Dreicer, H.: 1960, *Phys. Rev.* **117**, 329.
Evans, D. S.: 1974, *J. Geophys. Res.* **79**, 2853.
Gorney, D. J. Chiu, Y. T., and Croley, D. R., Jr.: 1985, *J. Geophys. Res.* **90**, 4205.
Horwitz, J. L.: 1986, *J. Geophys. Res.* **91**, 4513.
Kintner, P. M. and Gorney, D. J.: 1984, *J. Geophys. Res.* **80**, 937.
Klumpar, D. M. and Heikkila, W. J.: 1982, *Geophys. Res. Letters* **9**, 873.
Klumpar, D. M., Peterson, W. K., and Shelley, E. G.: 1984, *J. Geophys. Res.* **89**, 10779.
Lin, C. S., Burch, J., Winningham, J. D., and Menietti, J. D.: 1982, *Geophys. Res. Letters* **9**, 925.
Lundin, R., Eliasson, E., Hultqvist, B., and Stasiewicz, K.: 1987, *Geophys. Res. Letters* **14**, 443.
Mozer, F. S., Carlson, C. W., Hudson, M. K., Torbert, R. B. Paraday, B., and Yatteau, J.: 1977, *Phys. Rev. Letters* **38**, 292.
Potemra, T. A.: 1979, *Rev. Geophys. Space Phys.* **17**, 640.
Sandahl, I., Lundin, R. and Eliasson, L.: 1985, KGI preprint 077, Kiruna Geophysical Institute, Kiruna, submitted for publication to *IEEE Transactions on Geoscience and Remote Sensing*.
Sharp, R. D., Johnson, R. G., and Shelley, E. G.: 1977, *J. Geophys. Res.* **82**, 3224.
Sharp, R. D., Shelley, E. G., Johnson, R. G., and Ghielmetti, A. G.: 1980, *J. Geophys. Res.* **85**, 92.
Shelley, E. G., Sharp, R. D. and Johnson, R. G.: 1976, *Geophys. Res. Letters* **3**, 654.
Stern, D. P.: 1979, *Rev. Geophys. Space Phys.* **17**, 626.
Temerin, M.: 1986, *Geophys. Res. Letters* **13**, 1059.

BIRKELAND CURRENTS IN THE EARTH'S MAGNETOSPHERE*

T. A. POTEMRA

The Johns Hopkins University, Applied Physics Laboratory, Laurel, Maryland, U.S.A.

(Received 6 July, 1987)

Abstract. As a result of his polar expeditions at the beginning of this century, Kristian Birkeland determined that intense ionospheric currents were associated with the aurora. Birkeland suggested that these currents originated far from the Earth and that they flowed ointo and away from the polar atmosphere along the geomagnetic field lines. The existence of such field-aligned or 'Birkeland' currents was disputed because it was not possible to unambiguously identify current systems that are field-aligned (as suggested by Alfvén, 1939, 1940) and those which are completely contained in the ionosphere (as developed by Vestine and Chapman, 1938) with surface magnetic field observations. The presence of Birkeland currents has been absolutely confirmed with satellite-borne particle and magnetic field experiments conducted over the past two decades. These satellite observations have determined the large-scale patterns, flow directions, and intensities of Birkeland currents in the auroral and polar regions, and their relationship to the orientation and magnitude of the interplanetary magnetic field. The Birkeland currents are directly associated with visible and UV auroral forms observed with satellites. The results obtained from a variety of recently launched satellites are discussed here. These include Sweden's first satellite, VIKING, which has provided evidence for resonant Alfvén waves on the same geomagnetic field lines that guide stationary Birkeland currents. These observations demonstrate the important role that these currents play in the coupling of energy between the interplanetary medium and the lower ionosphere and atmosphere.

1. Introduction

1.1. Early history

The association of disturbances of the Earth's magnetic field with auroral displays dates back to Halley and Celsius in the 18th century. In the 19th century, Gauss suggested the possibility of electric currents from the Sun to explain magnetic disturbances, and at the end of the same century, Birkeland (1908) refined that idea to suggest that these solar currents were connected to currents in the auroral zones. Birkeland recognized that the geomagnetic disturbances recorded on the Earth's surface below the auroral region were due to intense currents flowing horizontally in the ionosphere above. They are referred to today as auroral electrojets, currents resulting from large-scale electric fields directed perpendicularly to the geomagnetic field. Birkeland suggested that the horizontal currents were connected to outer space by a system of currents flowing up and down along the geomagnetic field lines (see Figure 1 of Potemra, 1978).

The existence of field-aligned currents (later referred to as Birkeland currents) was disputed because it is not possible to identify them unambiguously from surface magnetic field measurements. Chapman (1927), for example, developed elegant mathematical descriptions of current systems contained completely within the Earth's

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

ionosphere. These currents could adequately account for surface magnetic variations. It has been suggested that Chapman was inspired by Lord Kelvin, President of the Royal Society, who proclaimed in 1893 that 'the supposed connection between magnetic storms and sunspots is unreal' (Dessler, 1984). Hannes Alfvén championed the idea of field-aligned currents and their importance to auroral physics, but had some difficulties in getting his ideas accepted. Vestine and Chapman (1938) stated that 'the electric current system of Birkeland gives rise to a disturbance-field shown to be inconsistent with observations in several important respects'. A year after this paper, Alfvén (1939) published an article that incorporated the Birkeland field-aligned current system into a theory of the aurora (see Figure 3 of Potemra, 1978).

The controversy concerning the existence of field-aligned currents continued through the 1960's when satellite data was provided that confirmed their existence. These were provided by Zmuda *et al.* (1966) with a magnetometer on board a polar orbiting Navy navigation satellite in a 1100 km altitude polar orbit. Disturbances transverse to the geomagnetic field and aligned approximately in the east-west direction were observed by Zmuda *et al.* (1966) on nearly every orbit through the auroral region. These magnetic disturbances were initially interpreted by Zmuda *et al.* as hydromagnetic waves. But it was soon realized that they were associated with field-aligned currents (Cummings and Dessler, 1967).

1.2. Patterns of Birkeland Currents

The most common method of detecting currents in the polar regions has been by virtue of the magnetic perturbations they produce. A satellite can pass through a (field-aligned) current and measure the *in situ* magnetic perturbations. The intensity and flow direction of the current is usually deduced from the formula $\mathbf{J} = 1/\mu_0 \operatorname{curl} \Delta\mathbf{B}$, where $\Delta\mathbf{B}$ is the measured perturbation transverse to the geomagnetic field. This technique is subject to some uncertainty because it is not possible to evaluate the three-dimensional 'curl' from a single satellite pass and not all transverse magnetic disturbances may be due to stationary field-aligned currents, because plasma waves may also produce magnetic perturbations. It has been usually argued that the large-scale magnetic perturbations (for example, 100 nT or larger observed over distances larger than 50 km in the low altitude ionosphere) can be explained reasonably in terms of large-scale Birkeland currents. Furthermore, the patterns of large-scale Birkeland currents deduced in this manner are supported by observations of precipitating electrons, which can account for almost all upward flowing currents, and by measurements of large-scale convective electric field patterns.

Since Zmuda *et al.*'s (1966) satellite observation of transverse disturbances, the large-scale characteristics of Birkeland currents have been studied extensively with magnetic field data collected by low altitude satellites (Potemra, 1987). And there have been several reviews of the characteristics of Birkeland currents compiled in the last few years (see, for example, Potemra, 1978, 1984; Saflekos *et al.*, 1982; Mauk and Zanetti, 1987). Some of these characteristics will be reviewed here and some of the recent results obtained with Sweden's first satellite, VIKING, on Birkeland currents will be presented here.

The characteristics of Birkeland currents have sometimes been separated into two categories based upon the spatial scale of the associated magnetic perturbation as measured by low-altitude polar-orbiting satellites. These are 'large-scale' defined as Birkeland currents associated with perturbations with spatial scales larger than 50 km or about 0.5° in latitude, equivalent to approximately a 6 s resolution for a magnetic field experiment on a low-altitude polar satellite. The 'small-scale' category is defined as Birkeland currents with spatial scales smaller than 50 km or 0.5° in latitude (at low satellite altitudes). Table I provides a list of characteristics of large-scale Birkeland currents and Table II lists some characteristics of small-scale Birkeland currents. Neither list is complete or comprehensive, because numerous ongoing satellite programs are providing new information on Birkeland currents. The major distinction between the large- and small-scale Birkeland currents may be in their source mechanisms. There is mounting evidence that the large-scale Birkeland current system is generated by the 'long-term' interaction of the solar wind with the Earth's magnetosphere, and that the small-scale system is associated with 'short-lived' plasma processes *within* the magnetosphere.

TABLE I

Large-scale Birkeland currents (spatial scale > 0.5° or 50 km)

– Stable patterns in auroral zone, cusp, polar cap (regions 1 and 2, cusp, and NBZ systems)
– Close association with large-scale convection and electric field patterns
– Close relationship to IMF
– Directly coupled to large-scale ionospheric currents detected on surface
– Carried by precipitating low-energy electrons (≤ few 100 eV) for upward flowing currents
– Carried by upward flowing ionospheric electron (thermal) for downward flowing currents
– Generated by solar wind/magnetospheric processes

TABLE II

Small-scale Birkeland currents (spatial size < 0.5° or 50 km)

– Irregular patterns embbeded in large-scale region 1, region 2, cusp and NBZ systems
– Close association with visible and UV auroral forms
– Close association with plasma waves, micropulsations, and ionospheric instabilities
– Associated with plasma processes within the magnetosphere and ionospheric processes
– Carried by high-energy (10's keV) particles

1.3. Birkeland Currents and the Interplanetary Magnetic Field

Figure 1 provides a summary of the large-scale Birkeland currents and their relationship to the interplanetary magnetic field (IMF) (from Zanetti and Potemra, 1986). This figure shows the magnetic perturbations often observed by low-altitude satellites over the polar region (the data from the MAGSAT satellite is shown in this case). Figure 1 illustrates the difference in transverse magnetic field disturbance observed during periods when the

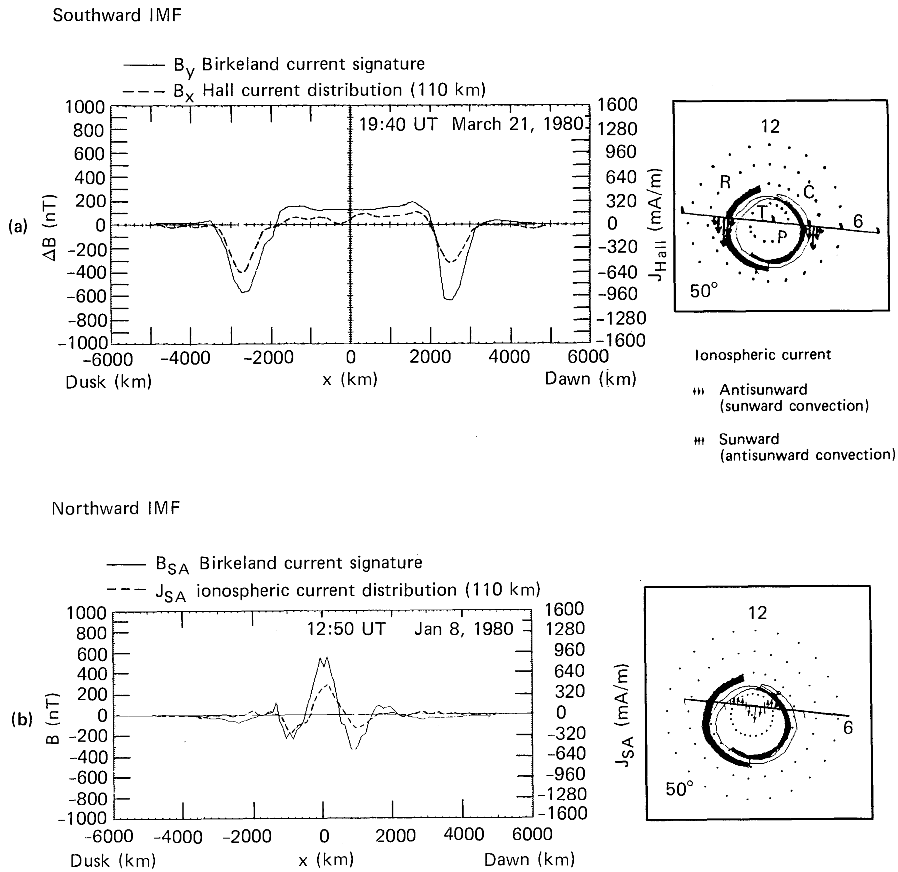


Fig. 1. Traces of magnetic field disturbance versus distance across the polar region measured with the MAGSAT satellite (Figure 2 from Zanetti and Potemra, 1986). The top panel corresponds to a period of southward-directed IMF and the bottom to a period of northward IMF. The small diagrams to the right of each data figure show MAGSAT's orbit superimposed upon the statistical pattern of Birkeland currents determined by Iijima and Potemra (1976). The solid trace in each figure is the magnetic perturbation associated with the Birkeland currents and the dashed trace is the distribution of ionospheric Hall current deduced from the MAGSAT observations (e.g., Zanetti and Potemra, 1986).

interplanetary magnetic field is predominantly southward (in the top panel) and northward (in the bottom panel). For southward conditions, the 'triangular-shaped' disturbances are most often observed which are due to the region 1–region 2 system of Birkeland currents flowing on the dawn and dusk sides of the auroral region (Iijima and Potemra, 1976). During periods of northward IMF, magnetic perturbations are often found at latitudes up to the magnetic pole, and a distinctive 'W-pattern' can be frequently found on the sunward half of the daylight hemisphere (as shown in the bottom panel of Figure 1). This perturbation is due to a major Birkeland current system called the

'NBZ' system (for 'northward B_z') which flows into and away from the polar cap ionosphere.

These relationships support the view that magnetospheric and auroral phenomena and, hence, the causative physical processes are intimately related to the orientation and magnitude of the interplanetary magnetic field. The polarity of the IMF B_z component determines whether these processes are active on geomagnetic field lines associated with either the auroral zones or the polar cap. The polarity of the B_y component is associated with dawn-dusk aand north-south hemispheric asymmetries of the current patterns.

1.4. Birkeland currents and aurora

The charged particles that carry the field-aligned Birkeland currents may excite the auroral atmosphere and produce emissions that can be detected on the ground or in space by satellites. Figure 2 shows a vacuum UV (1356 Å) auroral form observed in the

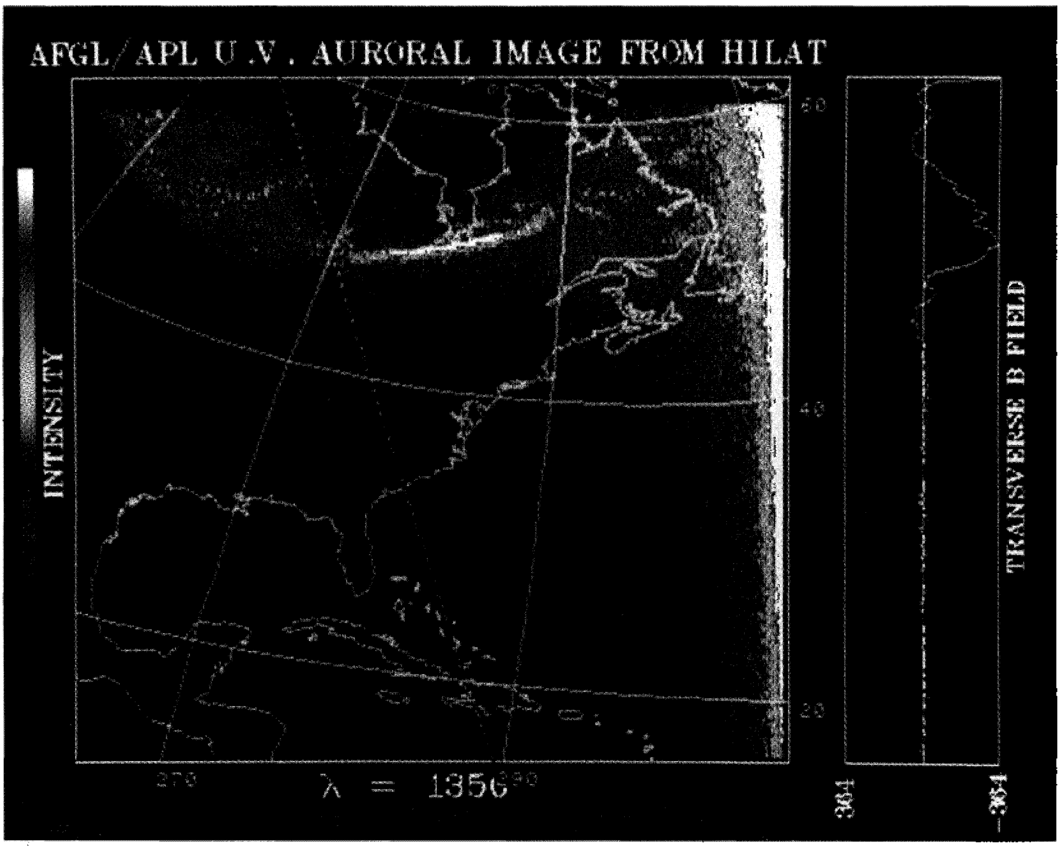


Fig. 2. A UV (1356 Å) image of the aurora on the sunlit side of the Earth acquired with the AIM instrument on HILAT on July 17, 1983 (Figure 11 from Potemra, 1983). The magnetic field observed by HILAT in approximately the east-west geomagnetic direction is shown in the line plot on the right. The day-night terminator at 115 km is indicated by the dotted line which extends from the top of the picture to the west of Hudson Bay to the bottom to the east of Florida. The bright arc below Hudson Bay occurs in complete daylight.

daylit atmosphere below Hudson Bay with the Auroral Ionospheric Mapper (AIM) instrument on the HILAT satellite (Figure 11 of Potemra, 1983). HILAT was launched on June 27, 1983 into a 850 km near-circular, 82.2° inclination polar orbit. It carried *in situ* plasma diagnostic instruments and VHF and UHF beacons for scintillation measurements in addition to the AIM instrument (Fremouw *et al.*, 1985, *Johns Hopkins APL Tech. Digest.*, April, 1984). The magnetic field component in approximately the east-west geomagnetic direction observed with the magnetic field experiment on HILAT is shown in the plot on the right-hand side of Figure 2. In this plot, the vertical axis is the distance along HILAT's trajectory (which occurs down the middle of the image map), and the horizontal axis is the magnetic disturbance. The large magnetic disturbance near the top of the plot (large-scale gradient directed toward the right and then to the left) reveals the presence of a pair of large-scale Birkeland currents. The large-scale current at the poleward side flows downward and the current at the equatorward side flows upward and is located at the same position of the bright auroral arc in the middle of the image. The downward flowing energetic electrons that carry a portion of the Birkeland current in this region produce this UV auroral form.

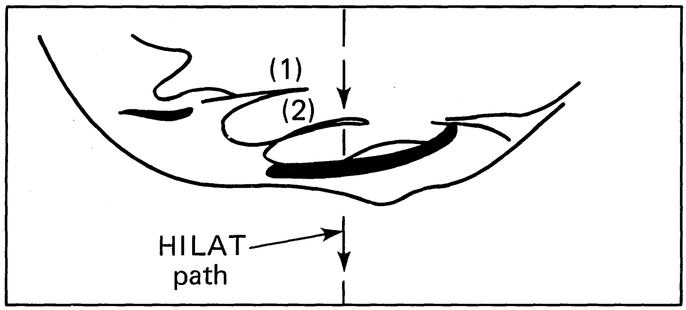


Fig. 3. A schematic drawing of the auroral features shown in Figure 2. The 'finger-like' arcs, labelled (1) and (2), are believed to the associated with a large-scale boundary wave in the distant magnetosphere.

The HILAT image shown in Figure 2 shows quasi-periodic forms which are depicted in cartoon form in Figure 3. The sub-satellite track of HILAT is shown in this figure as the vertical dashed line. Two 'finger-like' structures extend from the main auroral form poleward of the bright auroral feature discussed earlier. These forms are labelled '1' and '2' in Figure 3 and are associated with small-scale upward-flowing Birkeland currents that are embedded in the large-scale downward Birkeland current identified in Figure 2. These small-scale Birkeland currents appear as the small magnetic-field gradients embedded in the large-scale gradient labelled 'downward current' in Figure 2. Bythrow *et al.* (1986) have analyzed these data and the energetic electron and ion drift data acquired at the same time by HILAT. They suggested that these periodic structures of UV auroral forms and small-scale Birkeland currents are associated with waves at the interface of the Low Latitude Boundary Layer/Plasma Sheet (LLBL/PS) interface.

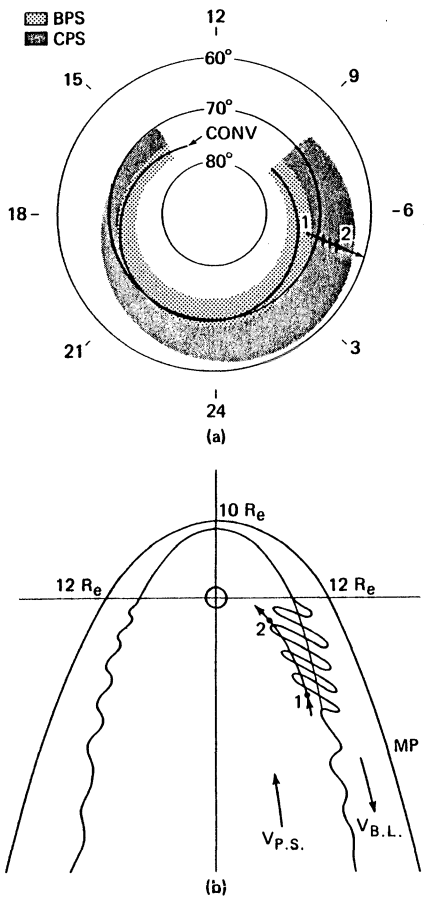


Fig. 4. A mapping of HILAT's orbit over the auroral region (*top*) to the distant magnetosphere (*bottom*) (Figure 4 of Bythrow *et al.*, 1986). The 'boundary plasma sheet' (BPS) and 'central plasma sheet' (CPS) are indicated. An exaggerated wave is shown in the magnetosphere. Points 1 and 2 on HILAT's orbit are indicated for reference purposes in the two regions.

Figure 4 of Bythrow *et al.* (their Figure 4) depicts the entire polar region (in the top) and corresponding regions in the distant magnetosphere (in the bottom). The orbital track of HILAT is depicted in the top figure as the line near 04 : 00 MLT with the points labelled '1' and '2' to provide in the two regions shown. This orbital track is mapped to the magnetosphere in Figure 4(b) near a wave on the LLBL/BL interface. The component of plasma drift associated with this boundary wave is directed normal to the boundary and it generates a polarization '$-\bar{v} \times \bar{b}$' electric field directly alternately sunward and tailward. The polarization current flow is directed opposite to the electric field and thus represents a generator for small-scale Birkeland currents. The mechanism also accelerates electrons to produce the auroral arcs (Bythrow *et al.*, 1986; Lundin and Evans, 1985).

The strong velocity shear and nearly parallel orientation of magnetic field lines in the LLBL to those of the inner magnetosphere at the interface of these two regions present an ideal site for generation of a Kelvin–Helmholtz instability (Sonnerup, 1980). The variations in thickness of the LLBL may be as large as 1 R_e and have periods of 2 to 5 min. These waves are a likely candidate as a source for the periodic Birkeland currents and UV auroral forms discussed here.

1.5. Observations of Birkeland Currents with Viking

Sweden's first satellite, VIKING, was launched on February 22, 1986 into a 817 km × 13 527 km polar orbit. It carried experiments to measure electric fields, magnetic fields, charged particles, waves, and auroral images (The Viking Science Team, 1986; Hultqvist, 1987). This program is directed toward an understanding of global phenomena, such as plasma convection, global current systems, and auroral morphology, as well as small-scale problems, including particle acceleration processes, wave-particle interactions, fine-structure currents, and auroral kilometric radiation.

The VIKING Magnetic Field Experiments includes a flux-gate magnetometer system with the sensors mounted on a 2 m boom. It has four automatically switchable ranges from $\pm$ 1024 to $\pm$ 65 536 nT (full scale) and resolutions commensurate with a 13 bit A/D converter in each range ($\pm$ 0.125 to $\pm$ 5 nT). Approximately 53 vector samples per second are acquired. This sampling rate corresponds to a spatial resolution of 60 m at VIKING's apogee altitude of approximately 3 R_e (geocentric). When projected along geomagnetic field lines, this resolution is about 12 m near the Earth's surface in the auroral atmosphere.

A summary plot of the eccentric dipole east-west component of magnetic field observed by VIKING near its apogee is shown in the top panel of Figure 5 (Figure 2 from Bythrow *et al.*, 1987). A plot of the magnetic field measured in approximately the east-west geomagnetic direction by the DMSP-F7 satellite is shown in the bottom panel. The DMSP-F7 satellite was launched into a 800 km near-circular, Sun-synchronous (08 : 30–20 : 30 local time plane) orbit in October, 1983. The magnetic field data shown in Figure 5 were acquired by VIKING and DMSP within 40 min of each other and at local times separated by about 2 hours. The shapes of the magnetic disturbances are remarkably similar even though the satellites are separated by nearly 2 R_e in altitude. The data from each satellite are plotted on different scales to emphasize their similar shapes.

The negative gradient of about 200 nT over 1300 km detected by VIKING near 21 : 30 UT 1see Figure 5) may be interpreted as a downward flowing Birkeland current with an intensity of 0.17 A m^{-1} and a density of 0.13 μA m^{-2}. The DMSP satellite measured a negative gradient of about 925 nT over 240 km at nearly the same location providing an intensity of 0.74 A m^{-1} and a density of 3.1 μA m^{-2}. A $r^{-3.2}$, $r^{3/2}$, and r^3 dependence for spatial extent, intensity, and density, respectively, is expected for observations of the same Birkeland current sheet at different altitudes (Bythrow *et al.*, 1987; Rich *et al.*, 1981). The parameter r is the ratio of the geocentric distance of VIKING to that of DMSP-F7, which is equal to 2.8 for the example discussed here.

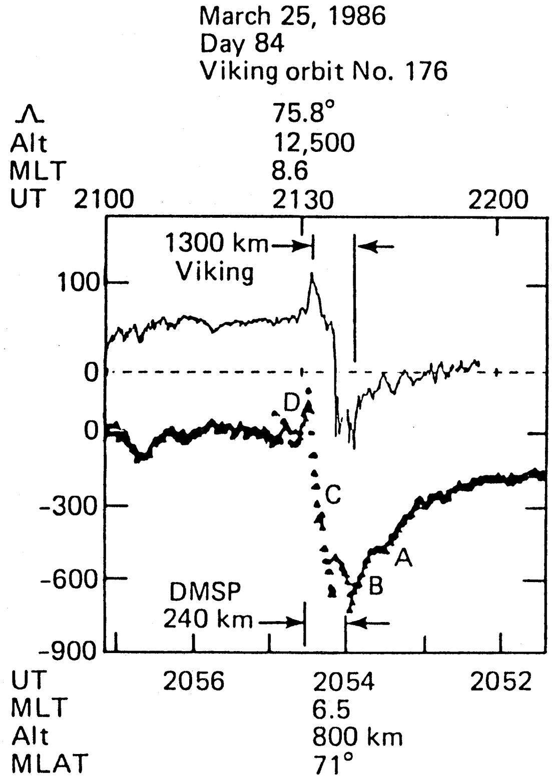


Fig. 5. Magnetic field disturbances in approximately the east-west direction measured by VIKING (*top*) and DMSP-F7 (*bottom*).

The observed spatial extent ratio is 240/1300 = 0.18 ($r^{-3/2} = 0.21$), the ratio of observed intensities is 4.4 ($r^{3/2} = 4.7$), and the ratio of densities is 23.8 ($r^3 = 22$). The ratios of the characteristics of the positive gradient due to the upward flowing region 2 Birkeland current shown in Figure 5, are also close to the predicted values (see Bythrow *et al.*, 1987). These observations support the view that the large-scale Birkeland currents comprise a stable and global element in the system that couples the lower auroral ionosphere to the outer magnetosphere.

Figure 6 is an expanded plot of the magnetic perturbation observed by VIKING shown in Figure 5. The eccentric dipole east-west and north-south components are shown with the energy-time spectrograms of electrons and positive ions (Plate 1 from Potemra *et al.*, 1987). The particle data was acquired with the Viking Hot Plasma instrument (Lundin *et al.*, 1987). The plot at the bottom of the figure shows the particle pitch angle, with 0° denoting precipitating particles and 180° denoting upward-flowing particles.

The period before 21 : 31 UT is characterized by intermittent fluxes of low-energy precipitating electrons and bursts of upward-flowing ions. The intense flux of precipitat-

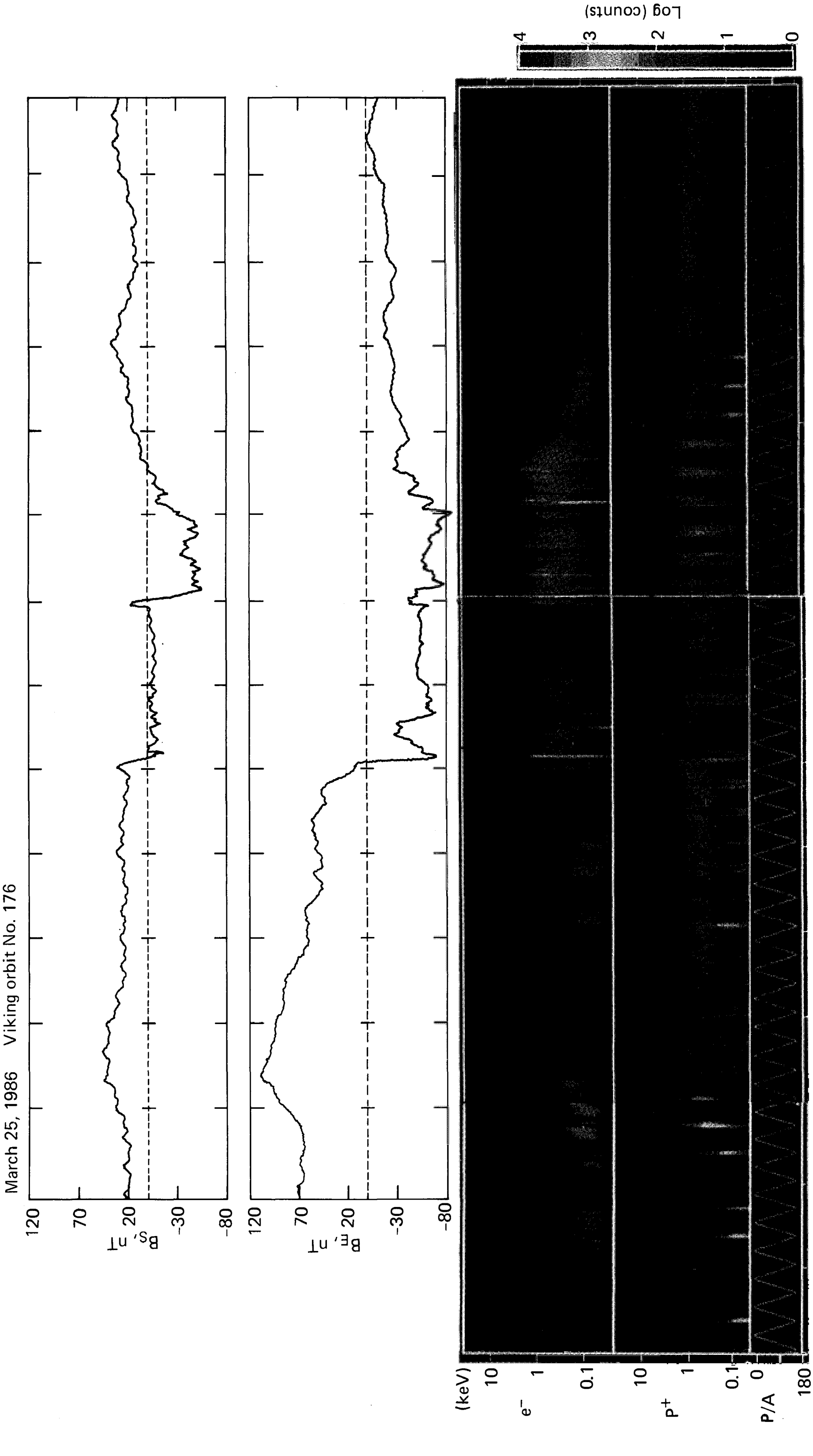


Fig. 6. The north-south and east-west components of magnetic field (in the top panels) and energy-time spectrograms for electrons and positive ions measured with VIKING. This is a high-resolution segment of the data shown in the top of Figure 5.

ing electrons between 21 : 31 and 21 : 31 : 30 UT coincides with the eastward magnetic field gradient denoting the upward-flowing Birkeland current. The westward magnetic field gradient from 21 : 31 : 30 to about 21 : 35 UT does not appear in the north-south component, and it is concluded that this is due to a large-scale region 1 Birkeland current that is aligned predominantly in the east-west direction. The particle spectrograms made during the same period show the general absence of precipitating electron fluxes with the exception of occasional bursts of electrons with energies below ~ 1 keV. The large-scale, downward-flowing region 1 Birkeland current may be carried by upward-flowing electrons with energies below the 10 eV threshold of the particle detector. The ion spectrogram shows the presence of ion conics characterized by the 'tuning fork' signatures (Gurgiolo and Burch, 1982) shown in Figure 6 beginning at about 21 : 31 : 30 UT.

A very sharp westward gradient in magnetic field of about 80 nT in 2.4 s occurs near 21 : 35 : 10 UT in Figure 6. There is also a significant north-south magnetic field perturbation at the same time, indicating that this current sheet is not strictly east-west aligned. The gradient corresponds to an 8 $\mu A\ m^{-2}$ Earthward-flowing current and is about 8 km wide. The particle spectrogram shows a very intense burst of upward-flowing electrons at the same time that the intense downward-flowing current is observed.

The region from about 21 : 35 to 21 : 39 UT is characterized by bursts of upward- and downward-flowing electrons with energies up to ~ 1 keV. This is accompanied by small-scale upward- and downward-flowing Birkeland currents, indicated by the perturbations in the north-south and east-west magnetic components, and is interpreted as a system of small-scale currents embedded in the large-scale region 1 system. The small-scale currents are carried by spatially-limited regions of hot electrons embedded in the upward-flowing thermal electrons population responsible for the large-scale region 1 system.

The eastward gradient of magnetic field beginning at about 21 : 38 UT is interpreted as the poleward edge of the large-scale upward-flowing region 2 current system. This boundary is close to the time of appearance of high-energy (~ 10 keV) electron fluxes in the spectrograms shown in Figure 6. These electron fluxes have been used to identify field lines that pass through the plasma sheet (see, for example, Bythrow *et al.*, 1981), confirming previous suggestions that, in the morning sector, the region 2 Birkeland current system maps to the plasma sheet and the region 1 system maps to the boundary layer (Bythrow *et al.*, 1981).

These characteristics may be summarized as follows. The large-scale ($> 0.5°$) region 1 and region 2 current systems are readily identified in the morning sector at altitudes of 12 500 to 13 000 km. The densities of these large-scale currents are about 0.1 $\mu A\ m^{-2}$, consistent with a mapping from lower altitudes, where typical densities are 2 to 4 $\mu A\ m^{-2}$ at 800 km altitude. Embedded in the large-scale region 1 system are small-scale intense currents that are not necessarily aligned in the east-west geomagnetic direction. One of them is an intense, 8 $\mu A\ m^{-2}$, Earthward-flowing, small-scale (~ 8 km wide) current that is carried by an upward flux of 'superthermal' (up to 1 keV) electrons.

The major carriers of the large-scale region 1 Birkeland currents appear to be upward-flowing electrons with energies below 10 eV. Embedded within this flux can be spatially limited regions of 'hotter' electron fluxes that carry small-scale currents embedded in the large-scale region 1 system, similar to those observed with the DE-1 satellite (Burch *et al.*, 1983). The region 1 system is located with particles that have characteristics associated with the boundary layer. The region 2 system coincides with particles with plasma-sheet characteristics. This supports the view that the source of the region 1 currents is in the boundary layer and that the region 2 currents are connected to the plasma sheet in the morning sector (Bythrow *et al.*, 1981).

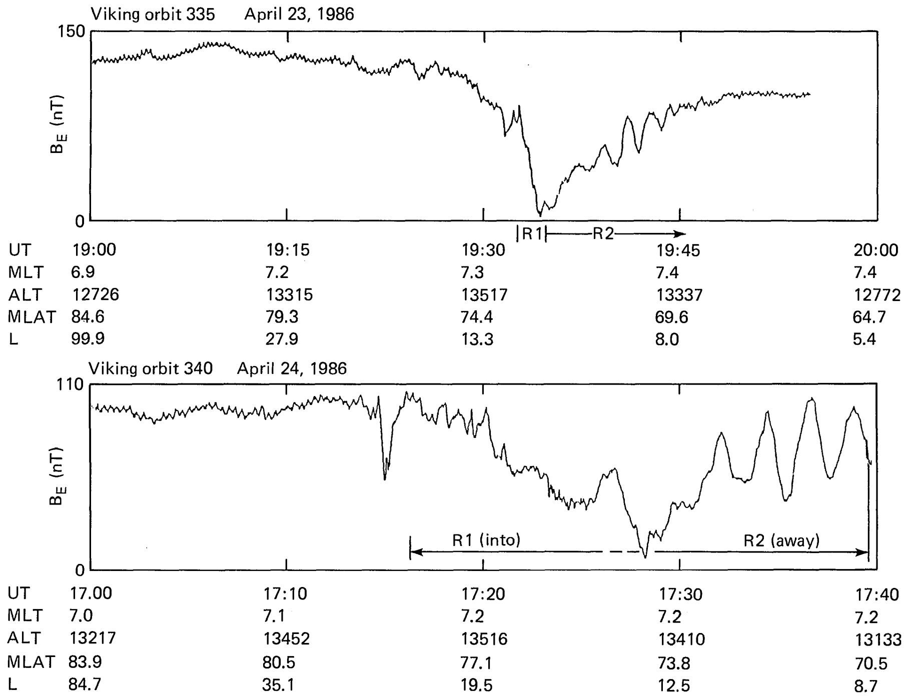


Fig. 7. Magnetic field disturbances in the east-west geomagnetic direction measured by VIKING near its apogee. Regions of large-scale region 1 and region 2 Birkeland currents are indicated as 'R1' and 'R2'.

Figure 7 shows the geomagnetic east-west components of magnetic field measured by VIKING on April 23 and 24, 1986 when the satellite was near its apogee over the morning auroral zone (near 07 : 00 MLT). The data acquired on April 23, 1986 show a steep negative (westward) gradient beginning approximately at 19 : 31 UT. This is interpreted as a downward flowing region 1 Birkeland current with a density of 0.4 $\mu A\ m^{-2}$. A positive (westward) gradient in the magnetic field may be seen after 19 : 35 UT which is interpreted as an upward flowing region 2 Birkeland current. Superimposed upon this positive gradient are oscillations with amplitudes up to 38 nT and periods of about 1.5 min. The data acquired on April 24, 1986 show a very wide region 1 Birkeland current extending from about 17 : 15 to 17 : 27 UT (4° of latitude) with a density of about 0.06 μA m^{-2}. A positive (eastward) gradient is evident at lower latitudes, but this is nearly hidden by the large oscillations in the field. The density of this region 2 Birkeland current is about 0.03 μA m^{-2} and the oscillations superimposed upon it have amplitudes up to 60 nT with a period of about 2 min.

Zanetti *et al.* (1987) have compared magnetic oscillations observed when VIKING and the AMPTE/CCE satellite (in an equatorial orbit) were very close to the same geomagnetic flux tube. Structured harmonic pulsations were observed by both satellites and they appeared to turn on and off simultaneously at both locations. Both the observations and the relative amplitudes along the magnetic field lines support the view that they are multiple field-line resonances of Alfvén waves. In a recent study using the VIKING magnetic field, electric field, and particle data, and the AMPTE/CCE magnetic field data, it has been suggested that resonant Alfvén waves can exist on the same geomagnetic field lines that guide the region 2 Birkeland currents in the morning sector (Potemra *et al.*, 1988). The sources of these waves are presently being investigated, and they include a high-latitude (cusp) entry mechanism (e.g., Engebretson *et al.*, 1987) and boundary waves driven by the Kelvin–Helmholtz instability as discussed earlier here with regard to periodic structures in auroral forms and Birkeland currents.

2. Summary

During the last two decades, numerous satellite observations have confirmed the presence of field-aligned curents which flow into and away from the Earth's auroral and polar regions. These observations support the suggestions of Birkeland and Alfvén of the important role that these currents have in auroral phenomena. When separated into categories depending upon their scale size and interplanetary conditions, the Birkeland currents can be described in terms of specific patterns. For example, the large-scale (spatial features larger than 50 km) Birkeland currents persist in predictable patterns when the interplanetary magnetic field (IMF) is either strongly northward or southward. When the IMF is southward, the Birkeland currents flow in two concentric circles around the geomagnetic pole in areas often referred to as 'region 1' (at the poleward side) and 'region 2' at the equatorward side. The region 1 Birkeland currents flow into the auroral zone on the morning side and away from the auroral zone on the afternoon side. The region 2 Birkeland currents flow in opposite directions of the region 1 system. During periods of strongly northward IMF, a different large-scale Birkeland current system dominates in the sunlit polar region which has been referred to as the 'NBZ' system. These general patterns appear to exist over a wide range of geophysical conditions, indicating that they are related to a fundamental coupling process between the Sun and Earth. The smaller scale Birkeland currents appear to be related to more complicated plasma phenomena contained completely within the magnetosphere. These processes are associated with particle acceleration, plasma waves, and aurora.

New results from the VIKING and AMPTE/CCE satellites indicate that geomagnetic field lines that guide stationary Birkeland currents can also support resonant Alfvén waves. The relationship of these waves to the current systems and their source in the magnetosphere is still under investigation. It is quite clear that Birkeland currents and Alfvén waves are fundamental to an understanding of the Earth's plasma environment.

Acknowledgements

The work reported here has been supported by the Office of Naval Research, the National Science Foundation, Defense Nuclear Agency, and the National Aeronautics and Space Administration. The magnetic field experiments on VIKING, HILAT, and DMSP were developed, fabricated and tested by the APL Space Department. The VIKING and AMPTE/CCE magnetic field instruments were developed under the guidance of M. H. Acuna. The NASA MAGSAT and DNA HILAT satellites were built by APL. The AIM image was provided through the courtesy of R. E. Huffman and C.-I. Meng. I am grateful for many stimulating discussions with L. J. Zanetti, P. F. Bythrow, and R. E. Erlandson.

References

Alfvén, H.: 1939, *Kungl. Sved. Vet-Akademiens Handl.* **III**, 18 (3).
Alfvén, H.: 1940, *Kungl. Sved. Vet-Akademiens Handl.* **III**, 18 (9).
Birkeland, K.: 1908, *The Norwegian Polaris Expedition 1092–1903*, Vol. 1, Sect. 1, Aschehoug, Oslo.
Burch, J. L., Reiff, P. H., and Sugiura, M.: 1983, *Geophys. Res. Letters* **10**, 753.
Bythrow, P. F., Heelis, R. A., Hanson, W. B., Power, R. A., and Hoffman, R. A.: 1981, *J. Geophys. Res.* **86**, 5577.
Bythrow, P. F., Doyle, M. A., Potemra, T. A., Zanetti, L. J., Huffman, R. E., Meng, C.-I., Hardy, D. A., Rich, F. J., and Heelis, R. A.: 1986, *Geophys. Res. Letters* **13**, 804.
Bythrow, P. F., Potemra, T. A., Zanetti, L. J., Erlandson, R. E., Hardy, D. A., Rich, F. J., and Acuna, M. H.: 1987, *Geophys. Res. Letters* **14**, 423.
Chapman, S.: 1927, *Proc. Roy. Soc.* **A115**, 242.
Cummings, W. D. and Dessler, A. J.: 1967, *J. Geophys. Res.* **72**, 1007.
Dessler, A. J.: 1984, in T. A. Potemra (ed.), *Magnetospheric Currents*, Am. Geophys. Union, Washington, D.C., p. 22.
Engebretson, M. J., Zanetti, L. J., Potemra, T. A., Baumjohann, W., .Luhr, H., and Acuna, M. H.: 1987, *J. Geophys. Res.* **92**, 10053.
Fremouw, E. J., Carlson, H. C., Potemra, T. A., Bythrow, P. F., Rino, C. L., Vickrey, J. F., Livingston, R. L., Huffman, R. E., Meng, C.-I., Hardy, D. A., Rich, F. J., Heelis, R. A., Hanson, W. B., and Wittwer, L. A.: 1985, *Radio Sci.* **20**, 416.
Gurgiolo, C. and Burch, J. L.: 1982, *Geophys. Res. Letters* **9**, 945.
Hultqvist, B.: 1987, *Geophys. Res. Letters* **14**, 379.
Iijima, T. and Potemra, T. A.: 1976, *J. Geophys. Res.* **81**, 5971.
Johns Hopkins University APL Tech. Digest: 1984, *Special HILAT Issue* **5**.
Lundin, R. and Evans, D. S.: 1985, *Planetary Space Sci.* **33**, 1389.
Lundin, R., Eliasson, L., Hultqvist, B., and Stasiewicz, J.: 1987, *Geophys. Res. Letters* **14**, 443.
Mauk, B. H. and Zanetti, L. J.: 1987, *Rev. Geophys.* **25**, 541.
Potemra, T. A.: 1978, *Astrophys. Space Sci.* **58**, 207.
Potemra, T. A.: 1983, *Johns Hopkins Univ. APL Tech. Digest* **4**, 276.

Potemra, T. A.: 1984, *Magnetospheric Currents*, Am. Geophys. Union, Washington, D.C.

Potemra, T. A.: 1987, *Phys. Scripta* **34** (in press).

Potemra, T. A., Zanetti, L. J., Erlandson, R. E., Bythrow, P. F., Gustafsson, G., Acuna, M. H., and Lundin, R.: 1987, *Geophys. Res. Letters* **14**, 419.

Potemra, T. A., Zanetti, L. J., Bythrow, P. F., Erlandson, R. E., Lundin, R., Marklund, G. T., and Block, L. P.: 1987, *J. Geophys. Res.* (in press).

Rich, F. J., Cattell, C. A., Kelley, M. C., and Burke, W. J.: 1981, *J. Geophys. Res.* **86**, 8929.

Saflekos, N. A., Sheehan, R. E., and Carovillano, R. L.: 1982, *Rev. Geophys. Space Phys.* **20**, 709.

Sonnerup, B. U. O.: 1980, *J. Geophys. Res.* **85**, 2017.

The Viking Science Team: 1986, *EOS, Trans. Am. Geophys. Union* **67**, 793.

Vestine, E. H. and Chapman, S.: 1938, *Terr. Magn. Atmos. Electr.* **43**, 351.

Zanetti, L. J. and Potemra, T. A.: 1986, in Y. Kamide and J. A. Slavin (eds.), *Solar Wind-Magnetosphere Coupling,* Terra Sci. Publ. Co., Tokyo, p. 547.

Zanetti, L. J., Potemra, T. A., Erlandson, R. E., Engebretson, M. J., and Acuna, M. H.: 1987, *Geophys. Res. Letters* **14**, 427.

Zmuda, A. J., Martin, J. H., and Heuring, F. T.: 1966, *J. Geophys. Res.* **71**, 5033.

AURORAL AND MAGNETIC VARIATIONS IN THE POLAR CUSP AND CLEFT – SIGNATURES OF MAGNETOPAUSE BOUNDARY-LAYER DYNAMICS*

P. E. SANDHOLT and A. EGELAND
Department of Physics, University of Oslo, Norway

(Received 2 July, revised 7 October, 1987)

Abstract. By combining continuous ground-based observations of polar cleft/cusp auroras and local magnetic variations with electromagnetic parameters obtained from satellites in polar orbit (low-altitude cleft/cusp) and in the magnetosheath/interplanetary space, different electrodynamic processes in the polar cleft/cusp have been investigated. One of the more controversial questions in this field is related to the observed shifts in latitude of cleft/cusp auroras and the relationship with the interplanetary magnetic field (IMF) orientation, local magnetic disturbances (DP2 and DPY modes) and magnetospheric substorms. A new approach which may contribute to clarifying these complicated relationships – simultaneous ground-based observations of the midday and evening-midnight sectors of the auroral oval – is illustrated. A related topic is the spatial relationship between the cleft/cusp auroras and the ionospheric convection currents. A characteristic feature of the polar cusp and cleft regions during negative IMF B_Z is repeated occurrence of certain short-lived auroral structures which seem to move in accordance with the local convection pattern. Satellite measurements of particle precipitation, magnetic field and ion drift components permit detailed investigations of the electrodynamics of these cusp/cleft structures. Information on electric field components, Birkeland currents, Poynting flux, height-integrated Pedersen conductivity, and Joule heat dissipation rate has been derived. These observations are discussed in relation to existing models of temporal plasma injections from the magnetosheath.

1. Introduction

The dynamical processes governing the particle, momentum, and energy transfer from the solar wind to the magnetosphere and upper atmosphere are main topics of solar-terrestrial research. Questions concerning the dynamics of the plasma-sheet of the nightside magnetosphere and its interaction with the auroral zone ionosphere have been studied in some detail (e.g., Akasofu, 1977; Shepherd *et al.*, 1980). In recent years attention has been focused on the physics of the dayside magnetospheric boundary layers (e.g., Paschmann, 1984; Eastman, 1984) and the coupling to the dayside polar ionosphere (e.g., Holtet and Egeland, 1985).

It has been speculated for a long time that the solar wind plasma can penetrate into the magnetosphere through magnetic neutral regions resulting from the coupling between the solar wind and the Earth's magnetic field. The first indication of the existence of a pair of magnetic neutral points (or lines) at high latitudes on the dayside magnetopause was provided by Chapman and Ferraro (1931), in a theoretical discussion of the interaction between a conducting plasma and the geomagnetic field. The first direct evidence for plasma entry in these regions was obtained around 1970 when satellite observations revealed the existence of plasmas of magnetosheath origin at low

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 171–199.

altitudes (Heikkila and Winningham, 1971) and at higher altitudes in the dayside magnetosphere (Frank, 1971).

Recent particle measurements from the Swedish spacecraft VIKING has confirmed earlier evidence (e.g., Gussenhoven *et al.*, 1985) of two different particle precipitation regions, one narrow (in longitude) cusp and a much more extensive cleft. The first VIKING results show the following distinguishing characteristics of the cusp 'proper', located in the 1100–1300 MLT sector (cf. Lundin *et al.*, 1987): (1) The hot plasma shows minute signatures of energization, the electrons and ions essentially showing magnetosheath characteristics. (2) The ions show some temporal flux variations, but in general they seem to be mainly affected by a poleward convection field. (3) It is characterized by an escape of ionospheric ions with energies below some 100 eV.

The cleft region, comprising the remainder of the dayside high-latitude portion of the auroral oval near noon, is quite different: (1) A significant amount of energization of both magnetospheric and ionospheric plasma (up to keV energies). (2) Temporal injections of magnetosheath plasma, showing characteristic time-dispersion signatures. (3) A strong ionospheric outflow of both electrons and ions.

During the International Geophysical Year 1957/1958 midday auroral emissions were recorded by all-sky cameras at polar stations (cf. Feldstein and Starkov, 1967). Photometric observations of these emissions were made by Eather and Mende (1971), who showed that the spectroscopic ratio $I(\mathrm{O\,I}\ 630.0\ \mathrm{nm})/I(\mathrm{N}_2^+\ 427.8\ \mathrm{nm})$ is enhanced by an order of magnitude relative to the typical midnight emissions.

There is now strong evidence that these red-dominated midday auroral emissions are due to magnetosheath plasma penetration into the magnetosphere and subsequent precipitation along field lines in the cusp (e.g., Holtet and Egeland, 1985).

Due to the inaccessibility of the polar regions which satisfy the observation conditions, i.e., correct distance to the geomagnetic pole and so far north in geographic latitude that the sunlight does not disturb the observations at magnetic midday, the cusp and cleft auroras have received markedly less attention than the night-time aurora. The only sites in the northern hemisphere which satisfy the above requirements for optical observations of the midday auroras are Svalbard (Norway) and Franz Josef's Land (U.S.S.R.) (cf. Figure 1).

The form, position and dynamics of the auroras near the cusp are known to reflect statistically the general character of the solar wind and the interplanetary magnetic field (Lassen and Danielsen, 1978). Recent results indicate that there may be a direct relationship between the solar wind interaction process at the dayside magnetopause and individual auroral forms observed near the projection of the magnetospheric cleft on the ionosphere (Sandholt *et al.*, 1985; 1986a). These emissions may then be used as a diagnostic tool in the investigation of plasma entry into the magnetosphere and the subsequent fate of that plasma under various solar wind and IMF conditions.

Concerning parameters like the scale size and recurrence rates of the dynamical plasma injection processes, satellite measurements permit only rough estimates. Continuous ground-based observations of the ionospheric 'footprints' may provide the necessary resolution in time and space.

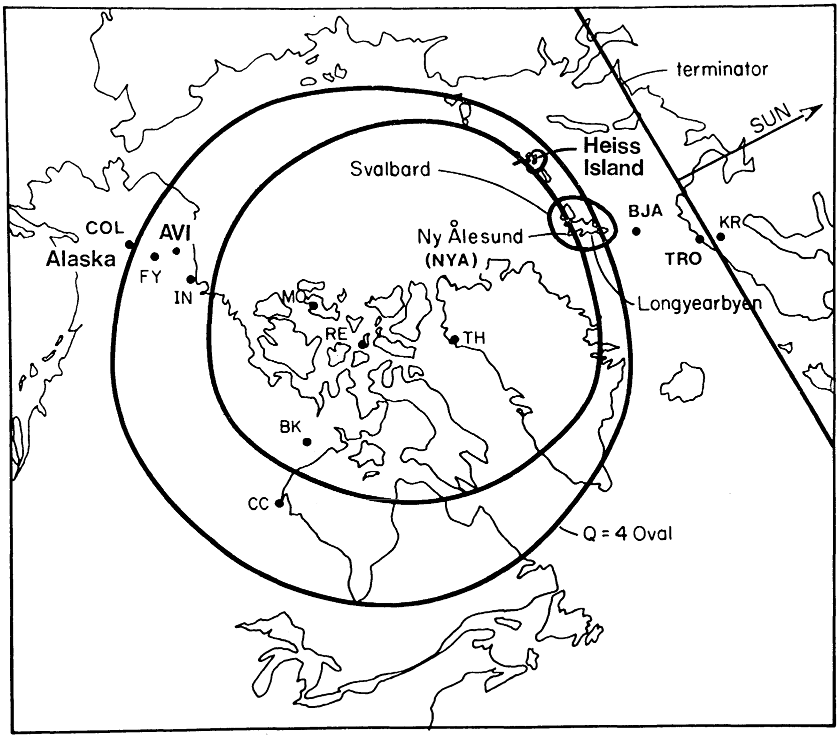


Fig. 1. Relationship of Svalbard and Heiss Island stations to auroral oval (Q = 4) (cf. Feldstein and Starkov, 1967) and the sunlit Earth in December at magnetic noon (~09 UT). The North American observing chains are indicated to show the possibility of simultaneous observations of developing substorms on the nightside and the midday aurora from Svalbard and Heiss Island.

In this report electrodynamical processes in the polar cusp/cleft with different temporal and spatial scales are discussed, based on ground-based observations of dayside auroras and local magnetic variations, combined with electromagnetic parameters obtained from satellites in polar orbit (above the ionosphere) and in the magnetosheath/interplanetary space.

2. Impulsive Plasma Transfer From the Magnetosheath – Signatures in the Cusp and Cleft Auroras?

2.1. Background

Owing to the dynamic nature of the shocked solar wind impinging on the front of the magnetosphere, i.e., irregularities in plasma density and magnetic field, plasma penetration across the dayside magnetopause should be a highly variable process (cf. Lemaire, 1977, 1979). This has been confirmed by *in situ* measurements from satellites (e.g., Sckopke *et al.*, 1981; Lundin and Dubinin, 1984). The plasma transfer mechanisms are intimately related to the electrodynamic coupling, constituting the power source for magnetospheric plasma convection (e.g., Cowley, 1982, 1984). Different mechanisms for plasma transfer are as follows (cf. Paschmann, 1984): (1) impulsive penetration events (IPEs), i.e., direct entry of plasma blobs (Lemaire, 1977; Lemaire *et al.*, 1979; Heikkila, 1982; Lundin, 1984), (2) quasi-steady state reconnection events (QSRs), i.e., plasma entry by merging of interplanetary and magnetospheric magnetic field lines in a quasi-steady state/large-scale process (Paschmann *et al.*, 1979; Sonnerup *et al.*, 1981; Aggson *et al.*, 1983, 1984), (3) flux transfer events (FTEs), i.e., transient and small-scale merging processes (Russell and Elphic, 1979; Rijnbeek *et al.*, 1984; Saunders *et al.*, 1984; Lee and Fu, 1985; Southwood, 1987), (4) viscous diffusion, i.e., plasma entry into a magnetically closed magnetosphere due to viscous interaction at the magnetopause (e.g., Eastman *et al.*, 1976; Mozer, 1984), and (5) gradient drift entry, i.e., plasma entry into a closed magnetosphere due to gradients in the magnetic field. This process injects plasma into the equatorial flanks and into the high-latitude dayside, separated according to the charge on the particle. This results in a charge buildup along the flanks of the magnetotail and the Sunward face of the plasma mantle. The latter condition may be the process which lengthens the cusp into a cleft (Olson and Pfitzer, 1985).

The impulsive injection model was first suggested for interaction of the solar wind with a magnetically closed magnetosphere; the flux transfer events, on the other hand, involve a direct local magnetic connection across the magnetopause. Heikkila (1982) discussed impulsive plasma injections in relation to an open magnetic field geometry, indicating that the impulsive penetration picture and flux transfer events refer to two aspects of the same process. Cowley (1984) has pointed out fundamental problems related to this model, which combines an open and equipotential magnetosphere. Lundin and Evans (1985) have introduced another model which is intended to incorporate both the IPE and FTE processes (cf. their Figure 8).

The detailed role of the polar cusp in the plasma entry mechanisms is not clear. Among the important tasks for future study are the processes and three-dimensional structure of the polar cusp region, and the electromagnetic effects in the polar cusp ionosphere (cf. Paschmann, 1984).

The optical aurora has been widely used as a sensor of magnetospheric processes. On the nightside the luminous auroral break-up is considered to be the most unambiguous signature of the onset of magnetospheric substorms (cf. Rostoker *et al.*, 1980). Similarly, the dayside aurora has been considered as an image of plasma processes in the boundary layers of the dayside magnetosphere (cf. Meng and Lundin, 1986; Lundin and Evans, 1985; Sandholt *et al.*, 1985, 1986a).

Possible signatures in the dayside aurora of plasma transfer across the magnetopause may be inferred from satellite observations and model considerations. Postulated auroral phenomena at the time of FTE signatures at the magnetopause are (S. W. H. Cowley, 1985, private communication; and L. Lee, 1986, private communication): (i) luminosity rapidly expanding southward from the existing cusp auroral display as the FTE is formed and then drifting poleward at speeds of $\sim 500\ \mathrm{m\ s^{-1}}$, (ii) luminosity disappearing after 5–10 min, when the flux tube has entered the tail

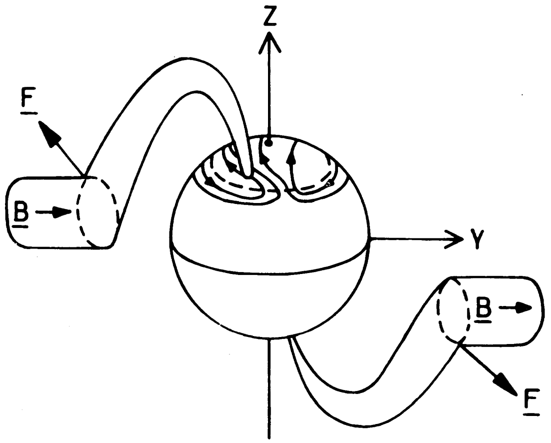


Fig. 2. Sketch illustrating the E–W forces F exerted on open flux tubes in the presence of an IMF B_Y field. The view is from the Sun. Also illustrated is the resulting azimuthal flow in the dayside cusp region (Svalgaard–Mansurov effect). The dashed line is the open field line boundary. (After Cowley, 1981.)

proper (cf. Figure 2), (iii) luminosity having a spectrum of scale sizes with the largest being several hundred km across, (iv) a series of events with recurrence time between 5 and 15 min, corresponding to the repeated formation and convection of magnetic islands (flux tubes or plasmoids), (v) injection of auroral particles with energy ~ 1 keV, due to acceleration by the reconnection electric field.

The recurrence time of FTEs is found to be $\tau = 10t_A/R_0 \sim 7-15$ min, where t_A is the Alfvén transit time across the width of the magnetopause current layer and $R_0 = V_1/V_A$ is the imposed driving rate at the incoming boundary (Fu and Lee, 1986).

Some of the predicted auroral signatures listed above are also consistent with a qualitatively different model, describing the impulsive penetration of solar wind plasmoids quite independently of any magnetic reconnection mechanism (cf. Lemaire, 1985). When the IMF has a southward component ($B_Z < 0$), the magnetic dipole moment M of a diamagnetic plasmoid with an excess density has a southward component ($M_Z < 0$). When such a plasmoid is at low latitudes near the frontside magnetopause, M being there antiparallel to M_E, both dipoles attract each other. Thus, solar wind irregularities with an excess momentum are attracted toward the inside of the magnetopause when the IMF is directed southward (Figure 3). A dawn-dusk asymmetry in the location of plasma injections, depending on the IMF B_Y component, is then predicted. The dusk (dawn) side is preferred during IMF $B_Y > 0$ (< 0) (anti-parallel field lines). Furthermore, according to Lemaire (1985), magnetospheric plasma will flow around the intruding plasmoid, i.e., in the opposite direction. A possible 'footprint' of this transient and localized flow of magnetospheric plasma will be rather short-lived, northward moving auroral structures, according to Lemaire.

An alternative possibility of plasma injection from the magnetosheath is at higher latitudes in the vicinity of the cusp, i.e., in the entry layer. After accessing terrestrial field

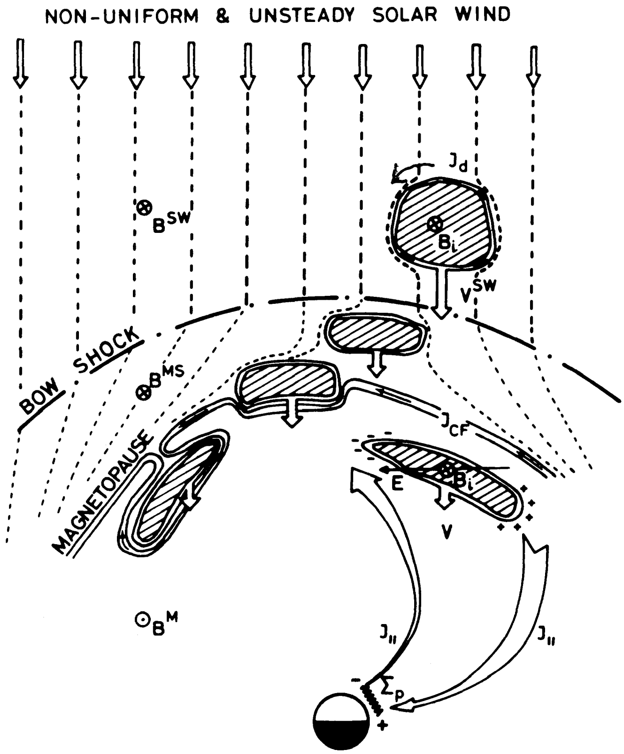


Fig. 3. Sequence of events representing the positions of a solar wind plasma irregularity penetrating through the Bow Shock and magnetopause. J_d is the sum of magnetisation, grad-B, and curvature currents; J_{CF} is the Chapman–Ferraro current density. $\mathbf{E}$ is the polarisation electric field ($-\mathbf{V} \times \mathbf{B}_\mathrm{I}$) induced in the magnetosphere by the plasma elements moving with the velocity $\mathbf{V}$; as soon as the plasma element is engulfed in a region with finite integrated Pedersen conductivity, the excess of kinetic energy of the intruding irregularity can be dissipated by Joule heating; the excess of momentum is transferred to the ionospheric plasma in the throat region. (After Lemaire, 1979.)

lines, the plasma could expand farther poleward owing to its high dynamical as well as static pressure. This, in turn, will be the free energy available for a dynamo process powering the dayside discrete aurora (cf. Heikkila, 1984; Lundin and Evans, 1985). Lifetimes of these penetrating plasma blobs have been estimated by Lundin (1984). According to his MHD model with energy dissipation regulated by the Pedersen conductivity (Σ_P) of the cusp ionosphere, typical lifetimes are a few minutes, corresponding to Σ_P-values of the order of 1 mho (cf. Section 3). The actual formula for the decay of the plasma element is of the form

$$t_D = \frac{n\,m\,\Delta x \Delta z}{B_0^2\,\Delta y}\,R_D\,,$$

where n is plasma density ($\sim 5 \times 10^6\,\mathrm{m}^{-3}$), $m \simeq m_P = 1.67 \times 10^{-27}\,\mathrm{kg}$, B_0 is the magnetic field in the cusp ionosphere ($\sim 5 \times 10^{-5}\,\mathrm{T}$), R_D is the electric resistance in the cusp ionosphere ($\sim 1/\Sigma_P$). Δx ($\sim 5\ R_E$), Δy ($\sim 1\ R_E$), and Δz ($\sim 1\ R_E$) express the spatial extension in the boundary layer of an injected plasma blob. The x- and y-axes point along and transverse to the boundary layer, respectively.

The underlying assumption is that the perpendicular current in the generator region, flowing across the boundary layer ($j_Y \sim \sigma_0(E_Y - uB_0)$), corresponding to an inertia current density, exerts a force per unit volume given by (cf., e.g., Alfvén, 1982):

$$\mathbf{J} \times \mathbf{B} = \rho \frac{\mathrm{d}u}{\mathrm{d}t} .$$

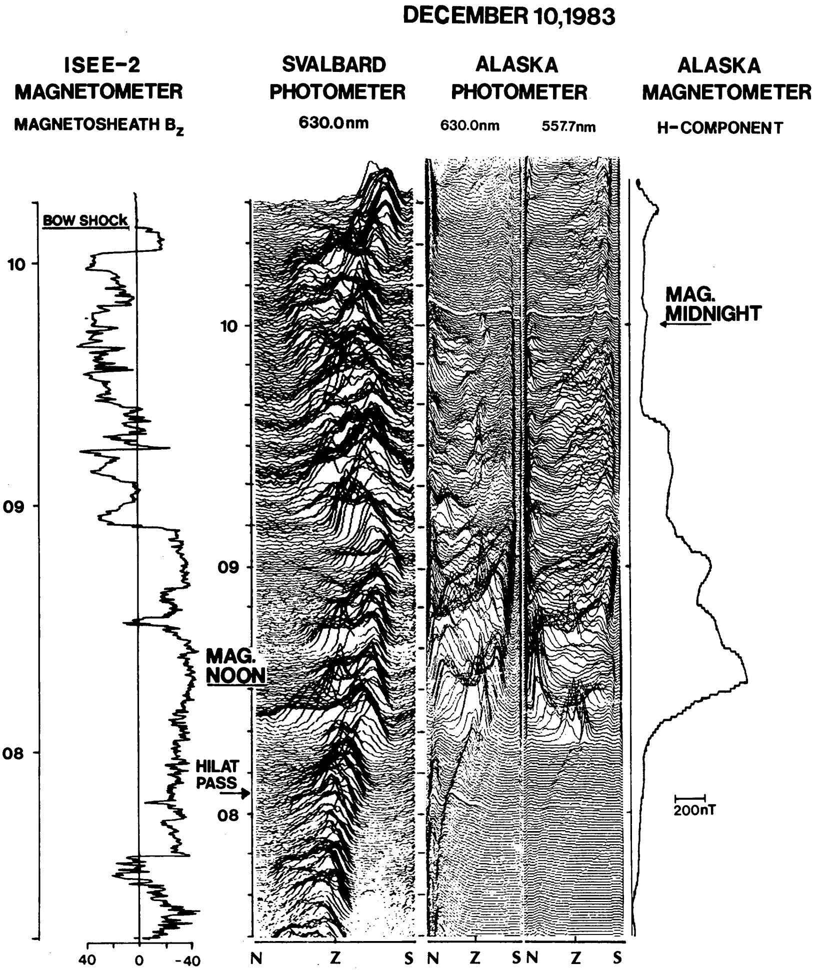


Fig. 4. The left panel shows magnetosheath magnetic field Z-component in nonteslas. Time is UT. The second panel shows photometer recording from Svalbard (north-south meridian scans) of the red oxygen line at 630.0 nm. The third and fourth panels show meridian photometer scans from Poker Flat (near Fairbanks), Alaska (65° geom. lat.). The fifth panel (far right) shows magnetometer recording from the station Arctic Village, Alaska (69° geom. lat.), close to the center of the westward electrojet at 08 : 30 UT. The IMF trace has been shifted by 15 min in relation to the observations from the ground, roughly corresponding to the time delay for the disturbance to reach the ionosphere. (After Sandholt *et al.*, 1986a.)

In the boundary layer the Lorentz term uB_0 corresponds to the injected magnetosheath plasma (H^+ and He^{2+}) moving with a higher 'convection' speed (u) as compared to the drift of the ambient plasma ($v(O^+) = E_Y/B$) (cf. Lundin, 1984).

2.2. Case studies

In this section we are going to present two observed cases showing characteristic dynamical features in the midday aurora which may be 'footprints' of impulsive electromagnetic coupling mechanisms at the magnetopause.

Figure 4 shows simultaneous observations by meridian scanning photometers of the midday aurora from Svalbard, Norway ($\sim 75°$ geom. lat.) and the evening aurora from Alaska ($\sim 65°$ geom. lat.). The north-south component of the magnetosheath magnetic field (GSE coordinate system), recorded by spacecraft ISEE-2, is shown in the left panel on the figure. The spacecraft was moving outward, crossing the bow-shock at 10:09 UT. A large and rather stable southward component was recorded between 07:35 and 08 : 55 UT. Between $\sim$ 07 : 50 and $\sim$ 08 : 20 UT both the day- and nightside auroras were moving southwards. At 08 : 20 UT a major substorm occurred on the nightside (cf. also AE-index in Figure 5). A series of poleward moving, transient

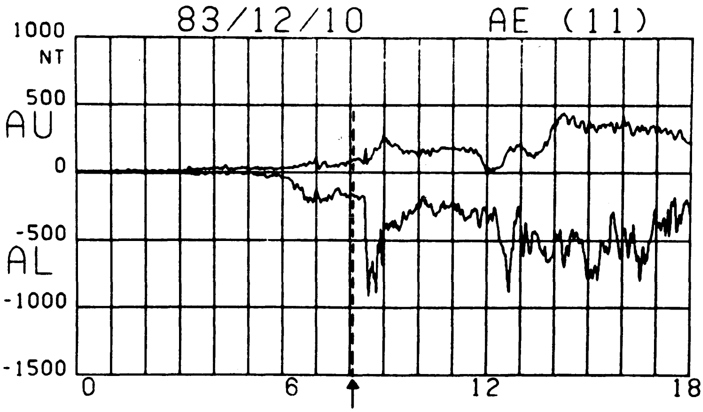


Fig. 5. Magnetic indices for December 10, 1983. Arrow and vertical dashed line mark the time of the satellite pass above the cusp ionosphere (cf. Figures 13, 14, 15).

luminosity structures were observed above Svalbard between $\sim$ 08 : 00 and 08 : 50 UT. A few of them are shown in more detail in Figure 6 (cf. also Figure 14). Notice the intensification at the equatorward boundary of the pre-existing auroral display from $\sim$ 08 : 15 UT and the subsequent poleward movement. After 7–8 min this luminosity had disappeared. The phenomenon is manifest in both channels displayed in the figure (red oxygen (O I) line and blue nitrogen (N_2^+) band. Another, similar event occurred between 08 : 26 and 08 : 33 UT.

Figure 7 shows an all-sky picture sequence of the midday aurora located to the south of Svalbard on December 30, 1981. The cusp arc is far south of the zenith at this time,

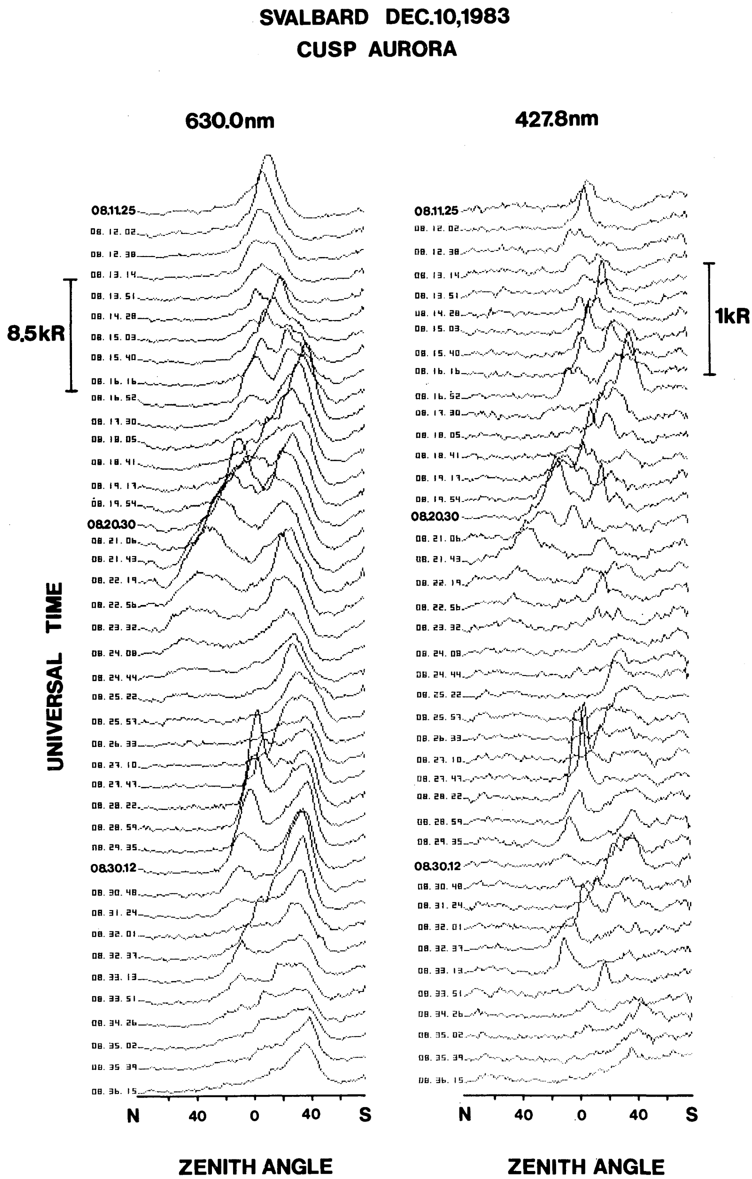


Fig. 6. North-south meridian profiles of red oxygen line and blue nitrogen band in cusp aurora during a 25 min period on December 10, 1983. (After Sandholt *et al.*, 1986a.)

when IMF $B_Z \simeq -10$ nT (cf. Figure 8). Notice the eastward motion, along the northern boundary of the pre-existing cusp arc, of a limited patch of enhanced luminosity. Another remarkable observation in this case, which may be of importance for the optical phenomenon, is the large negative IMF B_Y component (Figure 8) (cf. Section 2.3).

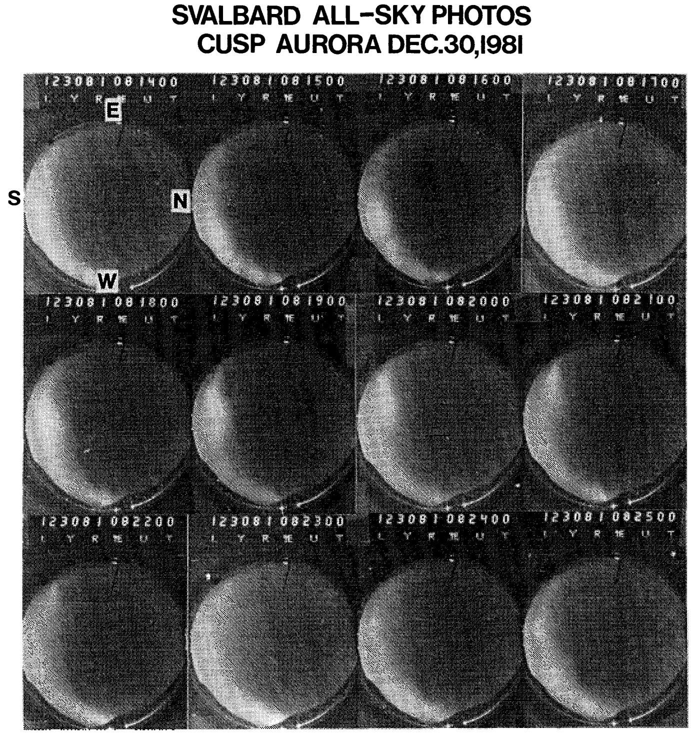


Fig. 7. Sequence of Svalbard all-sky photos of the midday aurora between 08 : 14 and 08 : 25 UT on December 30, 1981. The geomagnetic orientation of the pictures is indicated.

Summary of Observations

The cases presented here are examples taken from an extensive set of observations which may be divided into two main categories, according to the different spatial scales and dynamical behaviour (cf. Sandholt *et al.*, 1985, 1986a). Characteristics of the class we call 'small-scale' events are: (i) time of duration, 3–10 min; (ii) latitudinal extension, $\sim$ 50–100 km; (iii) longitudinal extension, $<$ 500 km (patches), (iv) large IMF B_Y component; and (v) longitudinal motion of luminosity, i.e., westward for $B_Y > 0$ and eastward for $B_Y < 0$.

The 'large-scale' cases are characterized in the following way: (i) time of duration, 8–15 min; (ii) longitudinal extension, $>$ 1000 km (elongated arcs); (iii) small IMF B_Y; and (iv) northward motion of these elongated arcs.

In addition to this, both categories show the following properties: (i) initial appearance at the equatorward boundary of the pre-existing cusp or cleft arc, located south of 75° geom. lat., (ii) occurrence limited to IMF $B_Z < 0$, (iii) repeated occurrence (series of events), and (iv) the optical spectral ratio $I630\ \text{nm}/I557.7\ \text{nm} > 2$ ($I427.8$ nm $\sim$ 0.2–1 kR) within these luminosity structures.

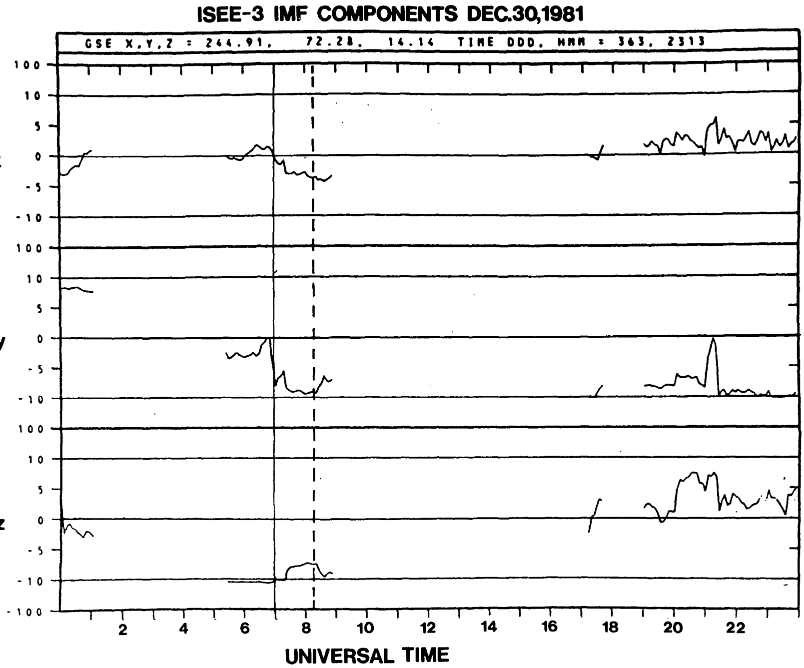


Fig. 8. Interplanetary magnetic field in geocentric Sun-ecliptic (GSE) coordinates measured from ISEE-3 at the libration point $\sim 250\ R_E$ upstream from the Earth. Vertical full line marks a transition towards more negative IMF B_Y. The dashed line indicates the time of the auroral recording shown in Figure 7.

2.3. Discussion

In Section 2.1 we discussed two plasma transfer models which are candidates to explain the auroral observations, i.e., impulsive penetration of diamagnetic plasma blobs and flux transfer events. Of the optical characteristics presented above the IMF B_Z dependence, the time of duration, and the repeated occurrence can be accounted for in both models.

Lundin's version of the IP model, with plasma injection at high latitudes (entry layer) is consistent with the northward motion of luminosity across the cusp/cleft (Lundin, 1984). Lemaire (1985) has suggested that the observed northward auroral motion corresponds to magnetospheric plasma directed toward the magnetopause, flowing around the intruding blobs. However, if the equatorial magnetopause projects down to somewhere near the cusp and cleft equatorward boundaries (last closed field line) the suggested plasma flow cannot explain the observed poleward motions, extending into the polar cap (cf. Figure 6). In addition, it seems morel likely that the optical phenomenon is related to the injected plasma blob itself, rather than the magnetospheric plasma flowing around.

According to the impulsive penetration model, the dusk (dawn) side of the magnetopause will be preferred injection site for IMF $B_Y > 0$ (< 0) and $B_Z < 0$

(antiparallel magnetosheath and magnetospheric field lines). After being injected the diamagnetic plasmoids are expected to proceed tailward along the boundary layer and subsequently be stopped after some minutes, due to adiabatic and non-adiabatic deceleration. Thus, a longitudinal motion across the noon meridian is not expected from this model. The observed IMF B_Y dependent longitudinal motions (cf. Figure 7) seem to be better accounted for in reconnection models (cf. Figure 2). The convection of open flux tubes are fundamentally different from the motion of the magnetically isolated plasma blobs, due to stresses exerted by the IMF (cf. Cowley, 1981).

It is noted here that the present reconnection models are highly controversial, even though accepted by a majority of magnetospheric physicists (cf. Axfprd, 1984; Cowley and Hughes, 1986). One view holds that the magnetic field description of magnetospheric plasmas applied in the present magnetic merging/reconnection theories is questionable (Alfvén, 1977).

Haerendel summarizes some of the properties of the magnetic merging/reconnection process, based on measurements in the boundary layers of the dayside magnetosphere from the ISEE-satellites (Haerendel, 1980): "It is transient and small-scale, i.e., the spatial scale is much smaller than the size of the overall magnetic configuration. – The short duration of the reconnection events leads to the erosion of magnetic flux in the form of rather discrete flux-tubes. – At the same time, direct acceleration of energetic particles is observed."

Southwood argues that the moving flux tube idea (Figure 2) – based on the frozen field concept – can be applied on time-scales large compared to the Alfvén wavez travel time along the flux tube. "For the tube to move as an entity, differential stresses must be redistributed along its length. The intermediate MHD (Alfvén) wave is the agent for redistribution. On a time-scale short compared with the travel time of such a wave back and forth along the tube between the equatorial magnetosphere and ionosphere (of the order of a minute or so for terrestrial flux tubes near the magnetopause) different parts of the tube move independently" (Southwood, 1987).

The open flux tubes (Figure 2) are embedded in the hot tenious plasma of the outer magnetosphere and in the relatively cool and dense plasma of the shocked solar wind, both environments highly collision-free and magnetized with plasma beta of the order unity (Haerendel, 1987). According to Alfvén, fluid models are not applicable in this media. "In collisionless plasma (which fill the active regions of space) the mean-free path is so long (compared to relevant dimensions) that the particles move in free ('ballistic') orbits. This makes the fluid model useless" (Alfvén, 1987).

A further discussion of the observations is given in Section 4.

3. Electrodynamics of the Cusp/Cleft Ionosphere

3.1. Background

Two separate systems of large-scale Birkeland currents in the cusp/polar cap have been proposed, based on satellite measurements during negative and positive IMF B_Z

components (NBZ currents), respectively (e.g., Potemra and Zanetti, 1985). A schematic illustration of the current system for $B_Z < 0$, with its dependence on IMF B_Y, is shown in Figure 9. The closure of these Birkeland currents in the magnetospheric boundary layers and the generators powering the respective current circuits are not known at present. Certain theories have been suggested, however, based on the available observations. Magnetic merging at the dayside magnetopause ($B_Z < 0$) and on field lines in the geomagnetic tail lobe ($B_Z > 0$) are considered to be important driving mechanisms (Iijima *et al.*, 1984). During periods of $B_Z < 0$ some part of the Chapman–Ferraro current can be diverted along geomagnetic field lines connected to the cusp, associated with a magnetopause rotational discontinuity (cf. Lee *et al.*, 1985).

A quantitatively different model has been suggested by Lemaire (1985). In that model field-aligned currents are generated at the boundary of diamagnetic plasma blobs when they enter the magnetosphere (cf. Section 2, Figure 3). When IMF B_Y is positive (negative) the preferred site of plasmoid penetration is on the dusk (dawn) side tail lobe. In this way one can explain the observed IMF B_Y-related dawn-dusk asymmetry in NBZ currents, in addition to the linear intensity relationship with IMF B_Z (> 0).

The local convection pattern associated with blob injection can be communicated to the ionosphere. Furthermore, a shower of penetrating plasmoids may eventually set up a quasi-steady-state convection flow pattern in the magnetosphere and ionosphere (Lemaire, 1985).

Plasma convection systems in the polar ionosphere predicted by the antiparallel merging hypothesis are shown in Figure 10 for different orientation of the IMF. The

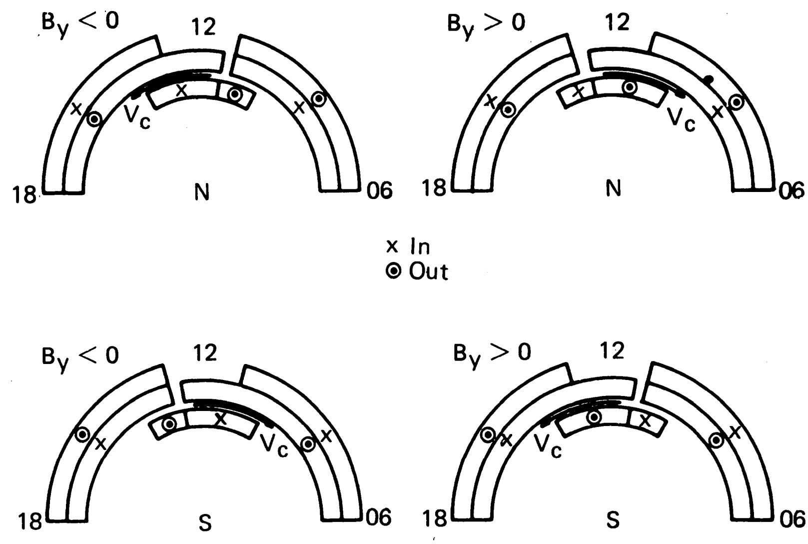


Fig. 9. Schematic of dayside, high-latitude field-aligned current systems for the north and south hemispheres. Left- and right-hand panels show the dominant polar cap convection direction and effects on the noon sector region 1 and cusp systems caused by IMF $B_Y < 0$ and > 0, respectively. (After Potemra and Zanetti, 1985.)

postulated reclosure convection cell is supposed to be driven by reconnection in the opposite hemisphere associated with magnetic closure between the northern and southern tail lobes. Experimental evidence on the reclosure cell is not unproblematic since its activation depends on a steady and favourable IMF orientation to exist for several hours. The existence of the lobe cell during southward IMF is also an open question.

A related problem, which will be addressed here, is the location of the polar cusp in relation to the patterns of plasma convection and Birkeland currents in the noon sector. These relationships can be studied, using optical recording of the cusp aurora combined with magnetograms from a latitudinal chain of stations and from polar orbiting satellites. We then use the established relationship between dayside high latitude convection patterns and geomagnetic signatures on the ground (cf. Friis-Christensen and Wilhjelm, 1975; Troschichev, 1982, and references given therein; Friis-Christensen *et al.*, 1985). One such case study is reported below. The other case presented in this section

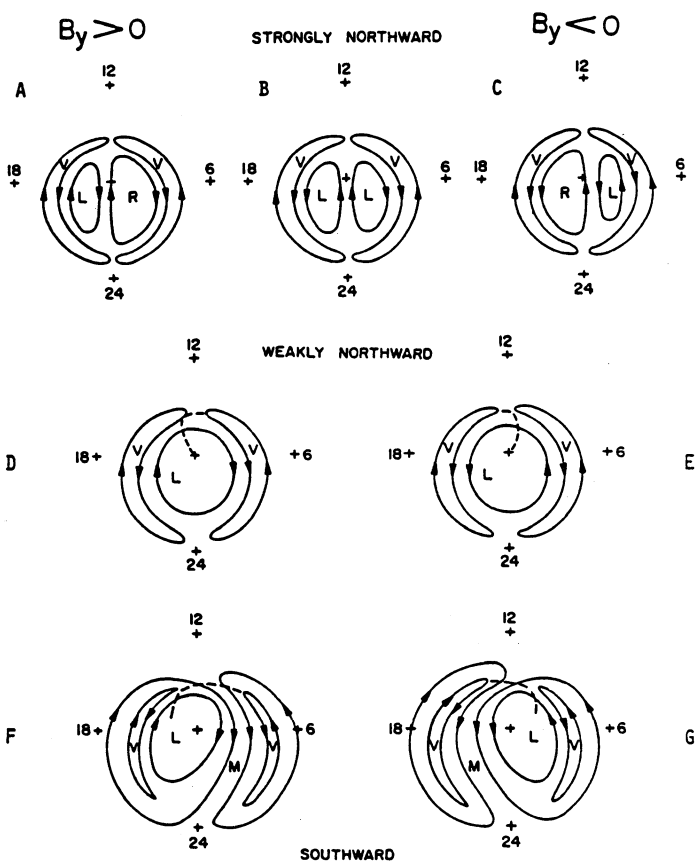


Fig. 10. Schematic northern hemisphere polar cap convection patterns for various IMF orientations. (After Reiff and Burch, 1985.)

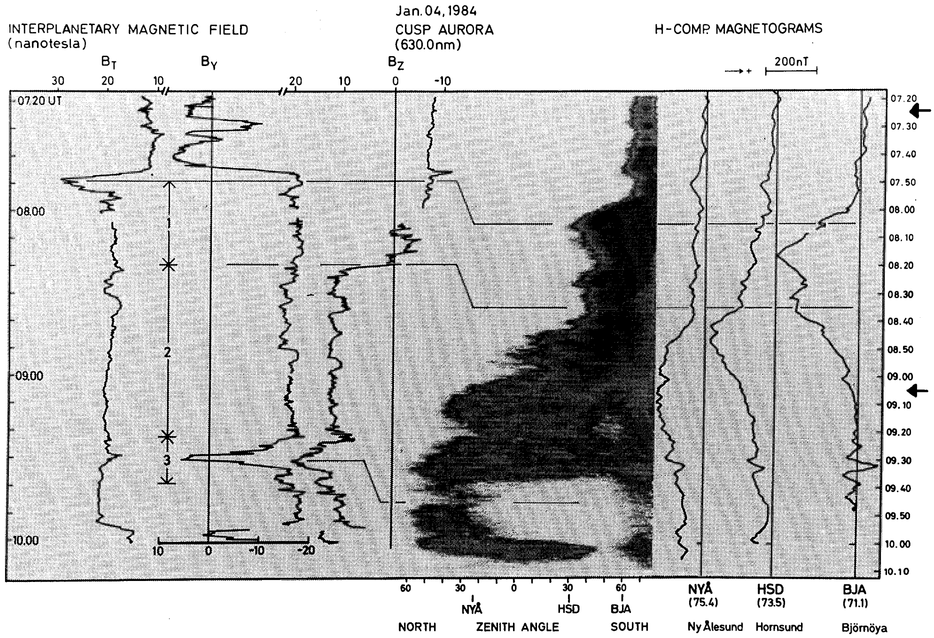


Fig. 11. Relationship between the amplitude and orientation of the interplanetary magnetic field (IMF) (left three panels) and intensity and location along the magnetic meridian of the midday polar cusp aurora above Svalbard (grey-scale representation of intensity versus zenith angle and time is shown in the middle panel), as well as the geomagnetic disturbance field recorded at three Svalbard stations (right three panels). The data recorded on the ground have been shifted by approximately 15 min relative to the IMF trace, in order to take into account the time delay between the IMF signal detected by the satellite (ISEE-2) in the solar wind and the geophysical response observed on the ground. Arrows in the time-scale to the right on the figure mark two successive DMSP F-7 passes above the cusp aurora (cf. Figure 12). The first pass occurred 1 hr to the west of Svalbard and the second one along the east-coast of Greenland. (After Sandholt *et al.*, 1986b.)

illustrates the electric field and current characteristics of the auroral forms in the polar cusp/cleft, which were described in Section 2. This study is based on combined observations from the ground and from a polar orbiting satellite.

3.2. Case Studies

3.2.1. The January 4, 1984 Case

Figure 11 illustrates the relationship between the amplitude and orientation of the interplanetary magnetic field (left three panels) and the intensity and location along the magnetic meridian of the midday polar cusp aurora above Svalbard. A grey-scale representation of intensity versus zenith angle and time is shown in the middle panel. The right three panels show the geomagnetic H-component disturbance field recorded at three stations on Svalbard. These observations can be separated into three periods with different IMF orientation.

Period 1. (cf. left panel in Figure 11). The magnetic signature indicates a southwestward DPY current with the amplitude modulated by IMF B_Z (B_Y is constant near -17 nT). The DPY signature (cf. Bjørnøya (BJA) magnetogram) can be interpreted as related to merging cell convection centered in the dawn sector of the polar ionosphere (Figure 10), with a northeastward flow (southeastward electric field) in the midday cusp region (cf. Crooker, 1979; Heelis, 1984; Reiff and Burch, 1985).

Period 2. The transition in IMF B_Z from negative to positive at 08 : 18 UT (B_Y not changing) was followed by changes in the optical aurora and the DPY signature. From 08 : 35 UT the magnetic disturbance decreases at Bjørnøya (BJA) while increasing farther north, at Hornsund (HSD) and Ny Ålesund (NYÅ). The poleward boundary of the cusp aurora moves northward during several tens of minutes after the IMF transition. This time period is comparable with the time for the IMF signal (08 : 18 UT) to reach the magnetotail (cf. Akasofu, 1977, p. 210). Cusp movements can be explained as the net result of dayside merging and nightside reconnection; i.e., the whole convection cycle is involved. According to this interpretation the cusp movement will continue until the convection cycle reaches a new equilibrium state.

At 09 : 00 UT the DPY equivalent current center is located near the poleward boundary of the cusp aurora (cf. Sandholt *et al.*, 1986b, Figure 5). This shows that the DPY current extends well into the polar cap, from approximately the center of the

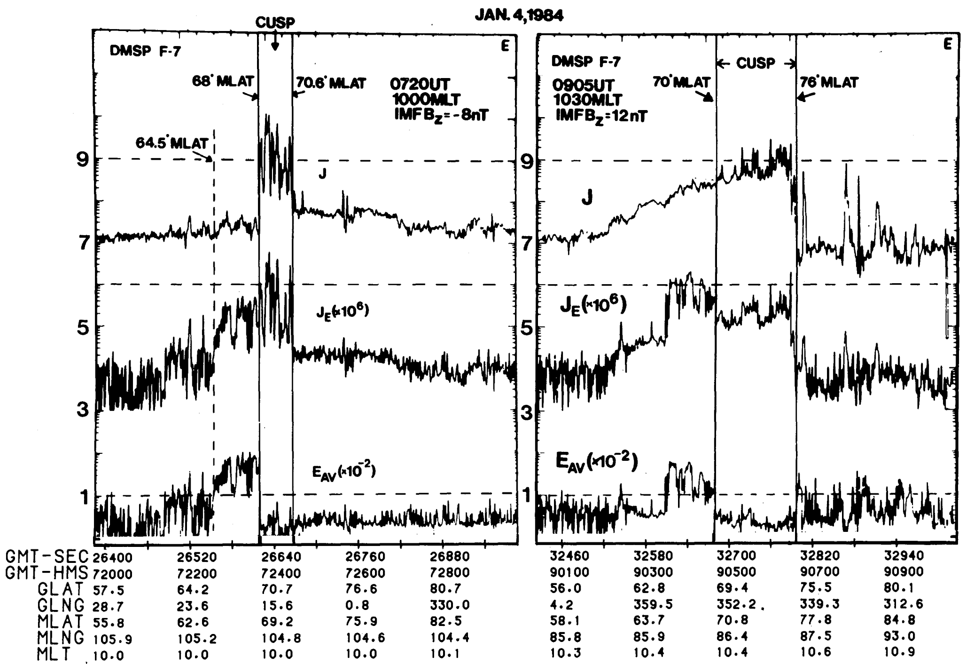


Fig. 12. Electron precipitation measurements during two successive passes of DMSP F-7 between Svalbard and Greenland on January 4, 1984. The polar cusp, marked by vertical full lines, is characterized by enhanced flux (J) (exceeding $\sim 10^8$ el. cm^{-2} s^{-1} sr^{-1}) and decreased average energy ((ε_{AV}) (below ~ 200 eV). Compare the optical data in Figure 11.

optical cusp. The DPY current observed in time interval 2 could be related to the lobe cell convection postulated for $B_Z > 0$ (cf. Figure 10) and identified by use of the dense magnetometer chain along the westcoast of Greenland (Friis-Christensen, 1985).

A DMSP satellite pass close to the eastcoast of Greenland at ~09 : 05 UT shows cusp-like particle precipitation within the 70–76° geom. lat. range (cf. Figure 12). The DPY mode of magnetic deflection is observed to reach a maximum level close to the poleward boundary of cusp precipitation. This is the same latitudinal relationship as was observed along the Svalbard meridian. The zone of 1–5 keV particles on the equatorward side of the cusp is manifested as a diffuse zone of enhanced green line aurora (cf. Sandholt *et al.*, 1986b, their Figure 2). Discrete arcs on the poleward side of the cusp are observed, corresponding to ~1 keV electron precipitation at those latitudes (cf. Figure 12). Thus, the midday auroral luminosity often shows latitudinal differences in spectral properties, indicating different sources of the associated particle precipitation.

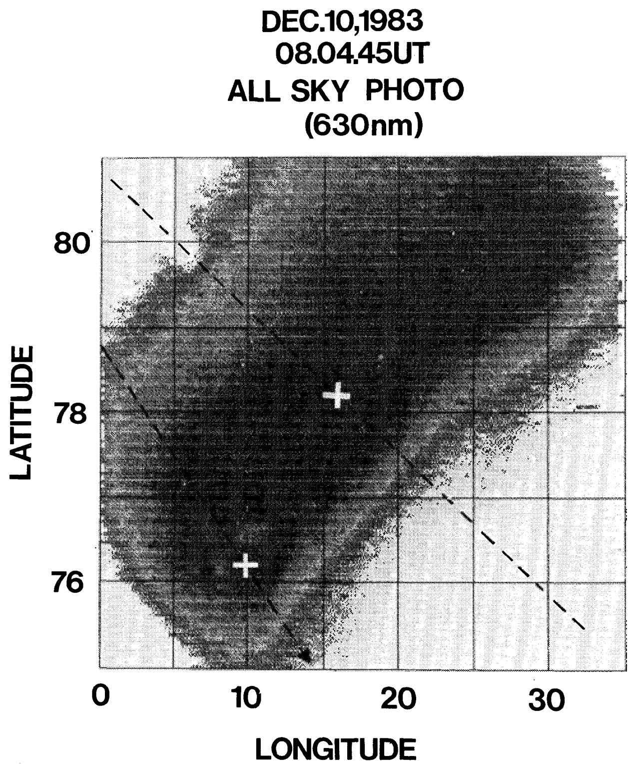


Fig. 13. Cusp aurora (red oxygen line at 630.0 nm) above Svalbard, at 08 : 04 : 45 UT, December 10, 1983, photographed by an image intensified all-sky camera in Longyearbyen (white cross in the center). The light intensities (digitized from photographic film) have been projected down to a flat Earth, with the geographic coordinates marked in the figure. The scanning direction of the Longyearbyen photometer system (dashed line through the zenith) and the HILAT trajectory (dashed line in lower left corner) are shown. White cross along the HILAT track marks the foot of field line coordinates of the satellite at 08 : 04 : 45 UT. The decreasing intensity towards magnetic west (lower left corner) is due to reduced camera sensitivity at the boundary of the field of view. (After Sandholt *et al.*, 1987.)

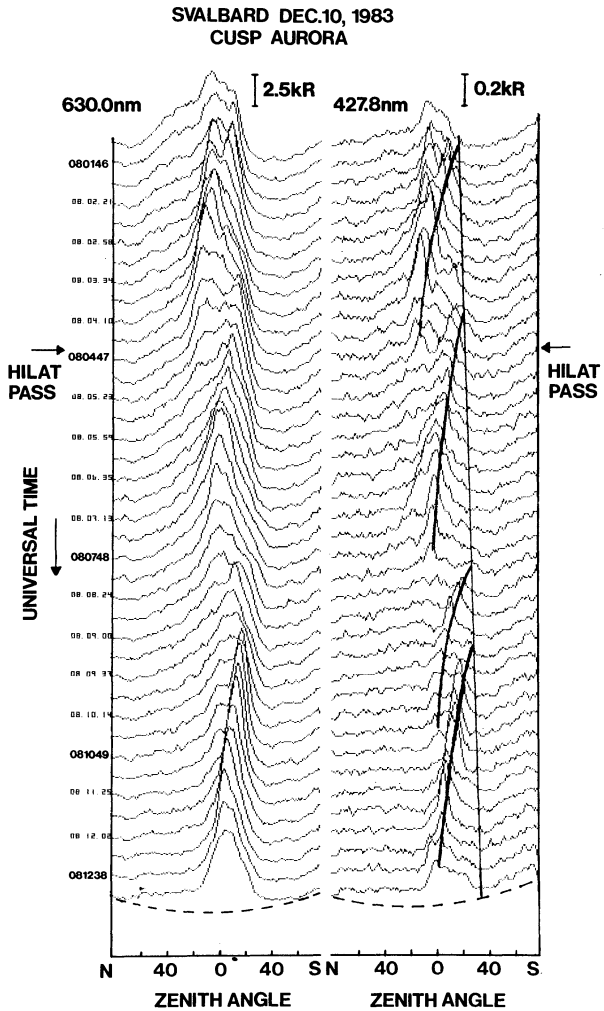


Fig. 14. North-south meridian photometer scans of auroral emissions at 630.0 nm (O I) and 427.8 nm (N_2^+) observed from Longyearbyen (LYR), Svalbard during an 11 min period (08 : 01–08 : 12 UT) including the HILAT pass at the Longyearbyen latitude which occurred at ~ 08 : 04 : 40 UT (cf. Figures 13 and 15). The equatorward boundary of the midday aurora and the poleward moving individual forms are marked in the right panel, while the calibration scales are given on the top. (After Sandholt *et al.*, 1987.)

The different particle source regions in the dayside magnetosphere, projecting to the ionosphere within the field of view of the scanning photometers at Svalbard, are: (1) the polar cap/plasma mantle, (2) the polar cusp/cleft, and (3) the dayside extension of the plasma sheet boundary layer.

Period 3. During this period, from 09 : 40 to 10 : 00 UT as observed from the ground (Figure 11), the IMF vector is pointing due north. According to the merging hypothesis this state should produce a net magnetic flux transfer from open lobe tubes to closed dayside tubes (cf. Cowley, 1981) with a resulting poleward motion of the equatorward boundary of the cusp. Such an effect is observed at ~ 09 : 40 UT. A weak cusp-like aurora is located to the north of Ny Ålesund between 09 : 35 and 09 : 55 UT.

The DMSP pass at ~07:20 UT (IMF $B_Z = -8$ nT) shows cusp electron precipitation between 68 and 70°.6 MLAT (Figure 12, left section), in good agreement with the optical recordings.

The main results of this event is incorporated in the summary section. It should be noted here that more events of this kind should be analyzed in order to establish the spatial relationship between the plasma convection patterns and the cusp auroral emissions. The problem is that the adequate events occur rather infrequently, mainly because of the very small magnetic deflection amplitudes in the winter polar cusp and cap during normal values of IMF B_Y and B_Z components.

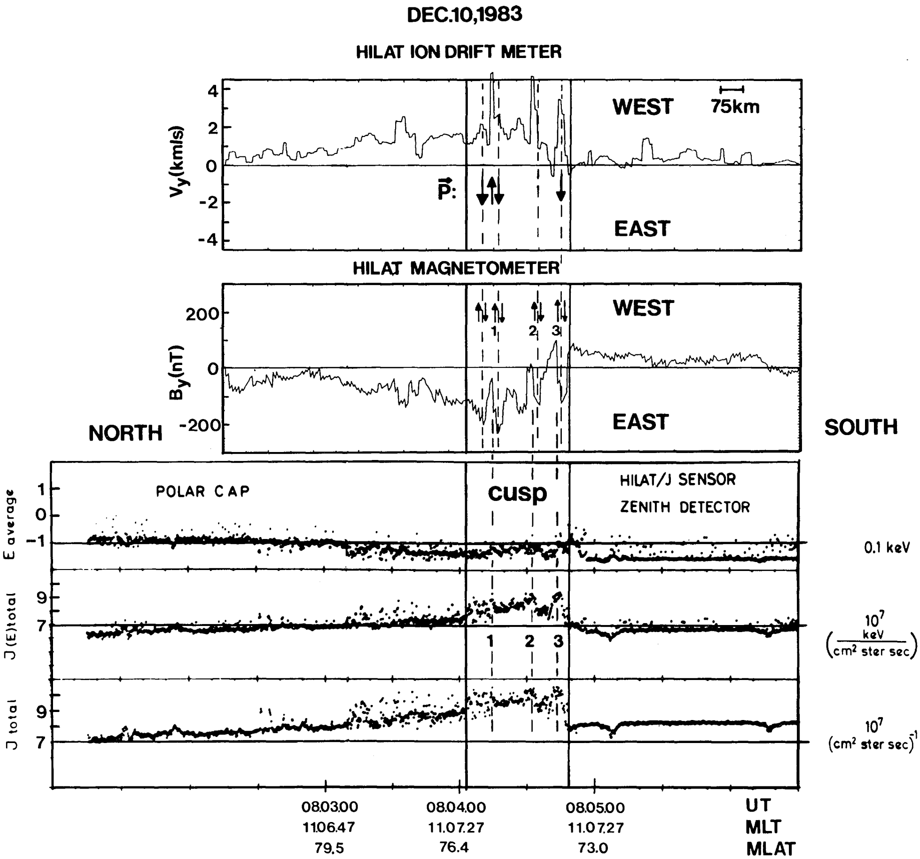


Fig. 15. *Upper panel*: Ion drift transverse to the satellite trajectory (positive values corresponding to westward drift). Vertical arrows indicate the direction of the Poynting flux. *Second panel*: Magnetic field east-west component deflections (positive westward). Arrows indicate the direction of Birkeland currents. *Lower panels*: Electron precipitation data in logarithmic scales: average energy (keV), energy flux (keV cm^{-2} $ster^{-1}$ s^{-1}) and number flux (cm^{-2} $ster^{-1}$ s^{-1}). Numbers 1, 2, and 3 mark the latitudinal location of three main structures within the cusp (cf. Figure 14). (After Sandholt *et al.*, 1987.)

3.2.2. *The December 10, 1983 Case*

An overview of the cusp aurora above Svalbard as well as the magnetosheath magnetic field and nightside auroral zone observations is shown in Figure 4. Between 08 : 03 and 08 : 05 UT satellite HILAT passed above the cusp aurora at an altitude of 800 km, slightly to the west of Svalbard, at 11 : 07 MLT (cf. Figure 13). Also shown in Figure 13 is the scanning direction of the photometer system at Longyearbyen, Svalbard.

Figure 14 shows more details of the photometer recordings around the time of the satellite pass. We notice certain characteristic time and space variations of the aurora within the 11 min period (08 : 01–08 : 12 UT) covered in the figure. A series of transient structures is observed, each appearing at the equatorward boundary of the pre-existing luminosity and then moving poleward during their limited lifetime. Also notice the equatorward motion of the equatorward boundary of the auroral luminosity. Between 08 : 04 and 08 : 05 UT two optical peaks are observed, one on each side of the zenith. A similar, but more pronounced optical event occurred 10 min after the satellite pass (cf. Figures 4 and 6).

A typical spectral ratio I630 nm/557.7 nm for those two peaks close to the zenith at $\sim$08 : 04 : 30 UT is $\sim$4. The luminosity shows a rather sharp equatorward boundary, in contrast to the more gradual decrease towards north.

Figure 15 shows HILAT measurements of ion drift (east-west component) as well as electron precipitation (average energy, energy flux, number flux). Positive drift velocity component v_Y (westward $E \times B$ drift) corresponds to northward electric field, with 1 km s^{-1} equivalent to $\sim$50 mV m^{-1}.

From the magnetometer trace B_Y in the second panel of Figure 15 local values of Birkeland current density can be derived, using the infinite current sheet approximation. Positive gradient in B_Y (along the satellite trajectory) means downward current flow.

From the properties of the downward (zenith detector) electron flux and the optical aurora the cusp region is defined by the two vertical lines in the figure. The energy flux in the cusp is between 10^8 and 10^9 keV cm^{-2} ster^{-1} s^{-1} or 5×10^{-4}–5×10^{-3} W m^{-2} = 0.5–5 ergs cm^{-2} s^{-2} with average energies generally $\simeq$0.1 keV. Three of the most prominent precipitation structures are marked by labels 1 to 3. One second averaged electron energy spectra show distinct shoulders in these regions of flux enhancements, indicating particle acceleration.

We will concentrate somewhat on feature No. 3, located close to the cusp equatorward boundary. The total width of that structure is $\sim$45 km. All three flux enhancements are associated with strong intensifications of the westward drift velocity and consequently enhanced northward electric field, reaching peak values between 175 and 250 mV m^{-1}. Pairs of field-aligned currents in the same regions are observed, with density values $\sim$10 μA m^{-2}, in a region with a more large-scale background downward current of $\sim$1 μA m^{-2}. Farther north the average B_Y trend indicates a less intense (average value $\sim$0.2 μA m^{-2}) upward current.

Figure 15 shows that there is a general good correlation between the eastward magnetic deflection and the westward ion drift component (northward electric field) in the cusp and cap regions. This relationship also holds in structure 3, but not in 1.

In a Cartesian coordinate system with the x-, y-, and z-axes pointing towards north, east, and downward, respectively, and assuming $\delta/\delta y \simeq 0$, we obtain (cf. Bythrow *et al.*, 1980; Sugiura *et al.*, 1984)

$$\mu_0^{-1} \frac{\delta \Delta B_y}{\delta x} = \Sigma_p \frac{\delta E_x}{\delta x} - E_y \frac{\delta \Sigma_H}{\delta x} . \tag{1}$$

Equation (1) can be integrated to give

$$\Delta B_y(x) = \mu_0(\Sigma_p E_x - \Sigma_H E_y) + \text{const.}, \tag{2}$$

where ΔB_y is the difference between the measured magnetic field and the Earth's dipole field.

From the assumptions $\delta/\delta y \simeq 0$ and $\nabla \times \overline{E} \simeq 0$ it follows that $E_y \simeq$ const. A good correlation between ΔB_y and E_x then indicates that the gradients in the Hall and Pedersen conductivities at the equatorward arc boundary are negligible. Thus, we have

$$\Delta B_y(x) \simeq \mu_0 \Sigma_p E_x(x) + \text{const.}, \tag{3}$$

where $\Sigma_p \simeq$ const. With ΔB in nT, E in mV m^{-1}, and Σ_p in mhos, Equation (3) assumes (cf. Smiddy *et al.*, 1980) the form

$$\Delta B_y(\text{nT}) = 1.25\Sigma_p(\text{mho}) E_x(\text{mV m}^{-1}) + \text{const.} \tag{4}$$

Equation (4) presupposes that effects of neutral winds can be ignored. We obtain $\Sigma_p \simeq 0.8$ mho for the case studied here. (The ΔB_y and E_x curves overlap each other within structure 3, indicating that $1.25\Sigma_p \simeq 1$.)

From the satellite measurements we make the following estimate of the Poynting flux vector associated with the northward E-field and the eastward magnetic deflection in the center of structure 3:

$$\mathbf{P} = \mu_0^{-1} \mathbf{E} \times \Delta \mathbf{B}, \tag{5a}$$

$$P_z \simeq \mu_0^{-1} E_x \Delta B_y . \tag{5b}$$

Using Equation (3) we obtain

$$P_z \simeq \Sigma_p E_x^2 . \tag{5c}$$

Thus, within structure 3 we have a downward Poynting flux with the amplitude approximately equal to the ionospheric Joule heat dissipation rate, which is $\sim 2 \times 10^{-2}$ W m^{-2}. This is a factor two larger than the value of the maximum electron energy flux in the same region of space.

Current continuity at the arc equatorward boundary gives the following expression for the northward electric field within the arc (cf. Marklund, 1984):

$$E_x^A = \frac{\Sigma_P^E}{\Sigma_p^A} E_x^E + \frac{\Sigma_H^A - \Sigma_H^E}{\Sigma_p^A} E_y^E + \frac{J_\|}{\Sigma_p^A} , \tag{6}$$

where the values inside and outside the arc are marked by superscripts A and E, respectively.

Based on the observed good correlation between ΔB_y and E_x we have already concluded that conductivity gradients and associated polarization electric fields play a minor role in the electrodynamics of this arc. This is so because of the soft particle precipitation in the midday cusp compared to other local time sectors of the auroral oval, where conductivity gradients and polarization effects are much more important (cf., e.g., Doyle *et al.*, 1986). Thus, Equation (6) is reduced to

$$E_x^A \simeq E_x^E + \frac{J_\parallel}{\Sigma_P^A} . \tag{7}$$

In our case, $J_\parallel$ (A m^{-1}) = $j_\parallel$ (A m^{-2})$\Delta x \simeq 0.15$ A m^{-1}, having used the average value of $j_\parallel = 5$ μA m^{-2} and the latitudinal width of the downward flowing current filament $\Delta x = 30$ km. (The corresponding values for the upward current farther north is $j_\parallel = 10$ μA m^{-2}, $\Delta x = 15$ km, and $J_\parallel = 0.15$ A m^{-1}. In both regions $j_\parallel \simeq \Sigma_p \,|\delta E_x/\delta x|$, with $\Sigma_p \simeq 0.8$ mho.) Thus, it follows that the last term in Equation (7) is ~ 185 mV m^{-1}. This is close to the peak value of the electric field measured within the arc. According to the terminology introduced by Marklund the cusp structure studied here is a Birkeland current arc.

From Figure 15 we notice an average westward ion drift of ~ 1.5 km s^{-1} within the cusp, corresponding to a northward electric field component of ~ 75 mV m^{-1}. The positive B_Y gradient in the cusp indicates downward flowing Birkeland current in this region. Further north the current flows upward. These large-scale structures of Birkeland current and electric field are consistent with the general statistic results for the actual IMF orientation ($B_Z < 0$; $B_Y > 0$, cf. Figure 9). Whether the Birkeland current close to the equatorward cusp boundary flows on closed or open field lines is an open question.

The large-scale features in the westward ion drift and the eastward magnetic deflection components are well correlated, indicating that the ionospheric Pedersen current constitutes the Birkeland current closure.

The smaller-scale structures in magnetic deflection and ion drift associated with precipitation features 1 and 2 in Figure 15 are more complex than in structure 3. In structures 1 and 2 the E-field peaks are displaced towards the north of the central line separating the up- and downgoing Birkeland currents.

A superposition of the smaller-scale structures on the large-scale pattern of currents seem to be a natural interpretation of the present measurements. As a consequence, the ion drift peak in structure 1 is associated with upward directed Poynting flux (northward E-field and westward magnetic deflection) as illustrated in Figure 15. Upgoing Poynting flux has been observed earlier in auroral break-up events (cf. Primdahl *et al.*, 1986). Their interpretation is that kinetic energy transferred to the neutral atmosphere can subsequently be converted to electric energy and delivered back to the magnetosphere.

3.2.3. *Summary of the December 10, 1983 Case*

Dynamical auroral phenomena with different temporal and spatial scales have been exemplified by one case with simultaneous day and nightside observations, combined with electromagnetic parameters obtained from satellites in polar orbit (above cusp ionosphere) and in the magnetosheath. The main findings are:

(i) Simultaneous equatorward movements of cusp and evening-midnight sector auroras occurred during a period of stable, negative magnetosheath magnetic field Z-component (positve B_Y), well before substorm expansion phase onset (Figures 4 and 5).

(ii) The dayside optical recording show a series of poleward moving transient structures, each appearing at the cusp equatorward boundary and disappearing after some minutes. The lifetimes of these auroral forms are not obtained by the meridian scanning photometer technique alone, when longitudinal motion is present; i.e., in cases with large IMF B_Y-component (Sandholt *et al.*, 1986a). In the present case with IMF $B_Y > 0$ a westward auroral drift component is expected. (This is not resolved in the all-sky photos. The similar, but stronger optical event at $\sim$ 08 : 15–08 : 25 UT (cf. Figures 4 and 6) show westward motion.)

(iii) The well-correlated structures in the aurora, the satellite magnetogram and the ion drift measurements indicate that the corresponding Birkeland current and E-field features are also transient in nature.

(iv) The structures show enhanced downward electron energy flux ($\sim$ 5–10 ergs cm^{-2} s^{-1}), multiple Birkeland current sheets ($J_{\parallel} \sim 10\ \mu\text{A m}^{-2}$) and strongly enhanced northward electric field component (~ 200 mV m^{-1}). The corresponding optical emission intensities at 557.7 and 630.0 nm are $\sim$ 1 and 5 kilorayleighs (kRs), respectively.

(v) Although electron precipitation is present over both upward and downward directed Birkeland currents the electron precipitation maxima occur within regions of upward Birkeland current flow (cf. Figure 15). The more large-scale Birkeland current in the cusp is directed downward, indicating outward flowing ionospheric electrons (counter-streaming electrons) and/or precipitating ions.

(vi) Within the auroral structure observed close to the cusp equatorward boundary at the time of the satellite pass, the main contribution to the electric field was found to be the Birkeland current term, with correspondingly minor polarization effects. A downward directed electromagnetic energy flux (Poynting flux), associated with a pair of Birkeland currents, is dissipated as ionospheric Joule heating. In the center of the structure this energy input rate is a factor two higher than the electron precipitation flux.

(vii) The height-integrated Pedersen conductivities inferred from correlated latitudinal variations in ion drift and magnetic field components were found to be in fair agreement with the value obtained from the particle precipitation measurements (Sandholt *et al.*, 1987).

(viii) Precipitation enhancements at higher latitudes in the cusp show more complex electromagnetic structures.

(ix) Electron energy spectra within all the three main structures are characterized by

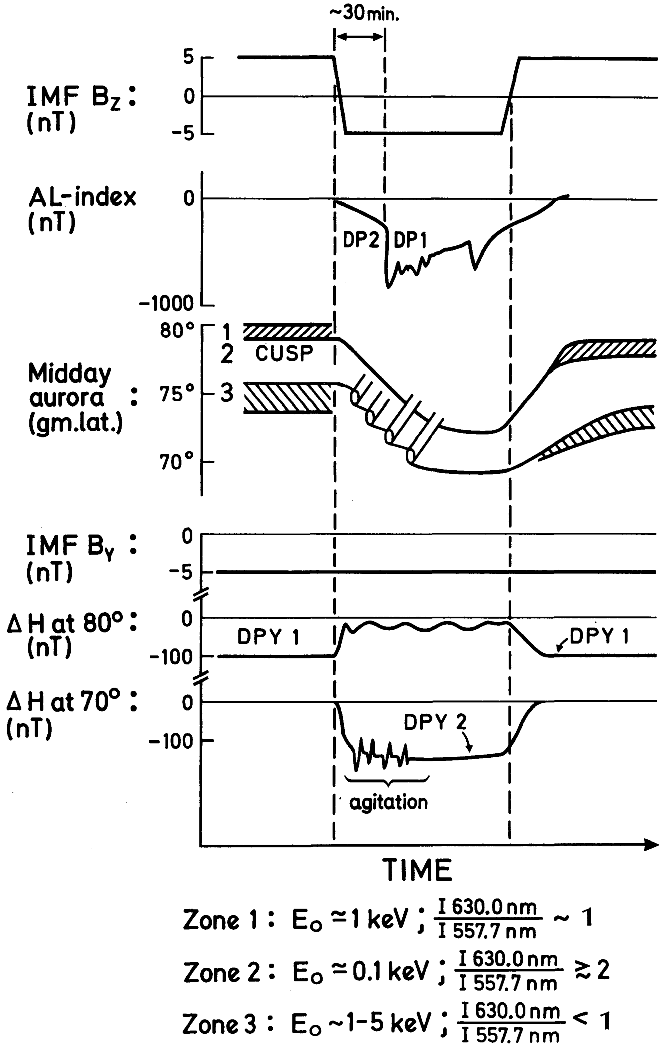


Fig. 16. Simplified sequence of changing IMF orientation with associated responses in geomagnetic activity (DP1, DP2, and DPY components) and midday auroras (including series of short-lived poleward moving forms during IMF $B_Z < 0$), according to the present case studies. Typical average energies (E_0) of electron precipitation zones corresponding to the different midday auroras are listed at the bottom of the figure.

peaks on the high-energy side at a few hundred electron volts, indicating field-aligned particle acceleration.

In this section we have discussed the electrodynamics of some characteristic auroral forms which are frequently observed in the polar cusp and cleft regions during negative IMF B_Z, based on one single satellite pass. A most critical question related to this case study is whether the equatorward part of the cusp, including structures 2 and 3, is on open or closed field lines. Information on the pitch-angle distribution of the precipitating particles may contribute to clarifying this question.

More events of this kind should be analyzed in order to establish the source of these transient auroral forms, i.e., the relationship with plasma entry and electrodynamic coupling at the dayside magnetopause.

4. Summary

Figure 16 is a schematic drawing which summarizes some of the observations presented in this article. The characteristics of different midday auroras and polar magnetic disturbance modes during a sequence of somewhat idealized IMF variations are illustrated. During IMF $B_Z \ll 0$ periods zone 3 in Figure 16 is usually outside the field of view of the optical instruments at Svalbard ($\sim 75°$ geom. lat.). Zone 1 is a characteristic feature of IMF $B_Z > 0$ intervals.

The initial state of the IMF vector, with $B_Z = 5$ nT and $B_Y = -5$ nT, means that the magnetosphere is in a low-energy state. An approximate empirical formula describing the electromagnetic energy input from the solar wind, i.e., the solar wind-magnetosphere coupling rate, is the ε-parameter given by Perrault and Akasofu (1978). This formula can be expressed in terms of the solar wind velocity and the IMF components transverse to the solar wind flow

$$\varepsilon \simeq 1/4v\{(B_T - B_Z)\}^2 l_0^2\,, \tag{8}$$

where $B_T = (B_Z^2 + B_Y^2)^{1/2}$ and $l_0 \simeq 7\,R_{\mathrm{E}}$.

$$\varepsilon(\mathrm{W}) \simeq 6 \times 10^6\, v\,(\mathrm{km\ s^{-1}})\,\{(B_T - B_Z)\,(nT)\}^2\,.$$

According to this formula the initial IMF state shown in the figure corresponds to $\varepsilon \sim 10^{10}$ W. This solar wind-magnetosphere coupling rate is accompanied by a weak custs aurora located at ~ 78–$79°$ geom. lat. The local magnetic disturbance component activated in this case (IMF $B_Y = -5$ nT) will be the DPY-1 mode, centered at the cusp poleward boundary and extending into the polar cap. It is supposed to be an effect of the lobe cell convection pattern (cf. Figure 10).

The transition of the IMF to the second state ($B_Z = B_Y = -5$ nT) is accompanied by a substantial increase of the coupling rate ($\varepsilon \sim 3 \times 10^{11}$ W). The ground-based observations in the cusp/cleft region show some clear changes in such cases. The aurora moves to lower latitudes. The response of the cusp poleward boundary is often higher than that of the equatorward boundary, resulting in a more narrow cusp when IMF B_Z is negative (cf. Figures 4 and 12). A decreasing DPY-1 magnetic perturbation is replaced by a growing DPY-2 component farther south, within the cusp/cleft. This may very well reflect the disappearance of the lobe cell and the activation of the merging convection cell (*M* cell in Figure 10).

During the first half hour period after the IMF southward transition is communicated to the magnetosphere, a significant polar cap expansion occur. This is inferred from the simultaneous equatorward motion of day- and nightside auroras (cf. Figure 4). This period may be interpreted as a growth phase, characterized by a net transport of magnetic flux, from the dayside to the nightside (cf. McPherron, 1979; Holzer *et al.*, 1986). At the end of this period the cusp aurora has moved down to $\sim 72°$ geom. lat. The onset of a magnetospheric substorm, following a period of negative IMF B_Z, typically do not influence the location of the cusp (cf. Figure 4). In some cases, however, the cusp is shifted farther southward at substorm onset. A tentative explanation of this different behaviour is that there are different kinds of substorms, sometimes referred to as spontaneous and *triggered* ordinary letters substorms (cf., e.g., Sergeev *et al.*, 1986). A spontaneous substorm is expected to have no immediate effect on the cusp location while the *triggered* one could be associated with instantaneous cusp movement, because of the close relationship with the IMF. A delayed cusp response related to substorms is expected due to the effect of the substorm activity on the magnetospheric convection cycle. It should be noted here that substorms are sometimes observed to be triggered by northward transitions of the IMF vector (cf. Rostoker, 1983; Sandholt *et al.*, 1986a, Figure 2A).

Soviet groups have studied the relative influence on dayside auroral luminosity of IMF variations and internal magnetospheric activity. They report the following results (G. V. Starkov, 1987, private communication): (1) Dynamics of the auroral luminosity band are defined by variations of the IMF B_Z-component under weakly changing magnetic activity. The IMF effect is that the B_Z variations at the subsolar point give rise to synchronous motion of the oval with a delay of 10–20 min (Vorobjev *et al.*, 1976; Zverev *et al.*, 1986). (2) Under increasing magnetic activity in the midnight sector the region of dayside auroral luminosity shifts equatorward and variations of its location associated with B_Z are observed at lower latitudes (Feldstein and Starkov, 1967; Vorobjev *et al.*, 1975; Zverev *et al.*, 1986).

A characteristic cusp/cleft observation during negative IMF B_Z, marked in Figure 16, is individual, poleward moving auroral forms (elongated arcs or patches). A characteristic magnetic agitation is usually observed in the local magnetic record from the ground when such an auroral structure is passing above (e.g., Sandholt, 1987, Figures 5, 6, 8; cf. also Lanzerotti *et al.*, 1987). Longitudinal auroral movements are observed in cases of large IMF B_Y component. It is suggested here that the phenomenon is a luminous 'footprint' of an electrodynamic coupling mechanism involving repeated plasma injection across the magnetopause. Boundary-layer dynamo processes like those proposed by Lemaire (1977) and Lundin (1984) are relevant in this connection. One attractive property of these models is that they explicitly describe a complete current circuit with a generator process that constitutes a power source for the observed discrete auroral forms and ionospheric Joule heat rates like those estimated in Section 3.

The observed cusp structures, characterized by latitudinal widths ~ 50 km and a mean northward electric field within the structure of $\sim 100\ \mathrm{mV\ m^{-1}}$ (cf. Figure 15) corresponds to a potential drop ~ 5 kV over a radial distance ~ 1000 km, when mapped to the frontside magnetopause boundary layer. This seems to be in reasonable agreement with inferred 'motional EMFs' across injected plasma filaments (cf. Lundin and Evans, 1985; their Figure 5).

The responses to the boundary-layer dynamo in ionospheric Pedersen current. Birkeland current and field-aligned potential drop required by current continuity then drive enhanced electron precipitation which causes the discrete auroras. Such a model does not explain all the observations listed in Table I, however. The IMF B_Y-related longitudinal movements seem to be better accounted for in a model involving magnetic merging (cf. Figure 2). The latitudinal motion of these transient auroral structures (cf. Figure 16), i.e., appearance at the cusp equatorward boundary (closed field lines ?/newly opened field lines ?) and subsequent disappearance near or poleward of the cusp poleward boundary (open field lines) are also consistent with this description.

An alternative explanation is that the different models, i.e., impulsive penetration into

TABLE I

Cusp auroral structures

Observation	Impulsive penetration	Reconnection[a]
Poleward drift	OK (high lat. only)	OK
Lifetime (3–15 min)	OK	OK
Occurrence: IMF $B_Z < 0$	OK	OK
Longitudinal motion related to IMF B_Y	No	OK
Series of events-recurrence time: 3–15 min	?	OK
Northward E-field: ~ 100–200 mV m^{-1}	OK	OK
Electron acceleration: $\Delta V_{\parallel} \sim 100$–$500$ eV	OK	OK
Latitudinal width: ~ 50 km	OK	OK

[a] Cf. comments in Section 2.3.

closed magnetospheric field lines and injection connected to open flux tubes after magnetic reconnection (cf. Table I), describe different aspects of the plasma transfer process. It has been suggested that the injected plasma could be trapped on closed field lines after reconnection (Sato *et al.*, 1986).

A final solution to this problem should be obtained after the Cluster mission, a multi-satellite project planned by ESA which will focus on the three-dimensional and small-scale structure of plasmas in the high-altitude cusp region (cf. European Space Science – Horizon 2000, 1984).

Acknowledgements

It is a great pleasure to thank C.-I. Meng and P. F. Bythrow, APL, Johns Hopkins University, D. A. Hardy and F. Rich, Air Force Geophysics Laboratory, Massachusetts, and R. Heelis, University of Texas at Dallas for particle measurements, magnetic field, and ion drift data from satellites HILAT and DMSP F-7, used in this study. IMF-data from missions ISEE-2/3 were kindly provided by R. C. Elphic, University of California, Los Angeles (principal investigator C. T. Russell). All-sky camera pictures were provided by H. C. Carlson/E. Weber, AFGL, Massachusetts and C. S. Deehr, Geophysical Institute, University of Alaska. Special thanks are due to Bjørn Lybekk, University of Oslo for operation of optical instruments at Svalbard and data analysis. This study has benefited from one author's (P.E.S.) visit to APL, Johns Hopkins University, Maryland, sponsored by NATO grant No. RG 85/0521.

References

Aggson, T. L., Gambardella, P. J., and Maynard, N. C.: 1983, *J. Geophys. Res.* **88**, 10000.

Aggson, T. L., Maynard, N. C., Ogilvie, K. W., Scudder, J. D., and Gambardella, P. J.: 1984, *Geophys. Res. Letters* **11**, 8.

Akasofu, S.-I.: 1977, *Physics of Magnetospheric Substorms*, D. Reidel Publ. Co., Dordrecht, Holland.

Alfvén, H.: 1977, *Rev. Geophys. Space Phys.* **15**, 271.

Alfvén, H.: 1982, TRITA-EPP-82 – 04, Dept. of Plasma Physics, Royal Institute of Technology, Stockholm, Sweden.

Alfvén, H.: 1987, *Geophys. Res. Letters* **14**, 475.

Axford, W. I.: 1984, in E. W. Hones, Jr. (ed.), *Magnetic Reconnection in Space and Laboratory Plasmas*, Geophys. Monogr. Ser., Vol. 30, AGU, Washington, D.C., p. 1.

Bythrow, P. F., Heelis, R. A., Hanson, W. B., and Power, R. A.: 1980, *J. Geophys. Res.* **85**, 151.

Chapman, S. and Ferraro, V. C. A.: 1931, *J. Geophys. Res.* **36**, 77.

Cowley, S. W. H.: 1981, *Planetary Space Sci.* **29**, 79.

Cowley, S. W. H.: 1982, *Rev. Geophys.* **20**, 531.

Cowley, S. W. H.: 1984, *Eur. Space Agency Spec. Publ. ESA* **SP-217**, 483.

Cowley, S. W. H. and Hughes, W. J.: 1986, *Planetary Space Sci.* **34**, 745.

Crooker, N. U.: 1979, *J. Geophys. Res.* **84**, 951.

Doyle, M. A., Burke, W. J., Hardy, D. A., Bythrow, P. F., Rich, F. J., and Potemra, T. A.: 1986, *J. Geophys. Res.* **91**, 6979.

Eastman, T. E.: 1984, *Eur. Space Agency Spec. Publ. ESA* **SP-217**, 77.

Eastman, T. E., Hones, E. W., Jr., Bame, S. J., and Asbridge, J. R.: 1976, *Geophys. Res. Letters* **3**, 685.

Eather, R. H. and Mende, S. B.: 1971, *J. Geophys. Res.* **76**, 1746.

European Space Science – Horizon 2000: 1984, *Eur. Space Agency Spec. Publ. ESA* **SP-1070**.

Feldstein, Y. I. and Starkov, G. V.: 1967, *Planetary Space Sci.* **15**, 209.

Frank, L. A.: 1971, *J. Geophys. Res.* **76**, 5202.

Friis-Christensen, E.: 1985, *Danish Met. Inst. Geophys. Paper R-72,* København, Denmark.

Friis-Christensen, E. and Wilhjelm, J.: 1975, *J. Geophys. Res.* **80**, 1248.

Friis-Christensen, E., Kamide, Y., Richmond, A. D., and Matsushita, S.: 1985, *J. Geophys. Res.* **90**, 1325.

Fu, Z. F. and Lee, L. C.: 1986, *J. Geophys. Res.* **91**, 13373.

Gussenhoven, M. S., Hardy, D. A., and Carovillano, R. L.: 1985, in J. A. Holtet and A. Egeland (eds.), *The Polar Cusp,* NATO Adv. Study Inst. Series C, Vol. 145, D. Reidel Publ. Co., Dordrecht, Holland, p. 85.

Haerendel, G.: 1980, in J. A. Holtet and C. S. Deehr (eds.), *Exploration of the Polar Upper Atmosphere*, NATO Adv. Study Inst. Series C, Vol. 64, D. Reidel Publ. Co., Dordrecht, Holland, p. 219.

Haerendel, G.: 1987, *Eur. Space Agency Spec. Publ. ESA SP-275*, 205.

Heelis, R. A.: 1984, *J. Geophys. Res.* **89**, 2873.

Heikkila, W. J.: 1982, *Geophys. Res. Letters* **9**, 159.

Heikkila, W. J.: 1984, in T. A. Potemra (ed.), *Magnetospheric Currents*, Geophys. Monogr. Ser., Vol. 28, AGU, Washington, D.C., p. 208.

Heikkila, W. J. and Winningham, J. D.: 1971, *J. Geophys. Res.* **76**, 883.

Holtet, J. A. and Egeland, A. (eds.): *The Polar Cusp*, NATO Adv. Study Inst. Series C, Vol. 145, D. Reidel Publ. Co., Dordrecht, Holland.

Holzer, R. E., McPherron, R. L., and Hardy, D. A.: 1986, *J. Geophys. Res.* **91**, 3287.

Iijima, T., Potemra, T. A., Zanetti, L. J., and Bythrow, P. F.: 1984, *J. Geophys. Res.* **89**, 7441.

Lanzerotti, L. J., Hünsücker, R. D., Rice, D., Lee, L. C., Wolfe, A., Maclennan, C. G., and Medford, L. V.: 1987, *J. Geophys. Res.* **92**, 7739.

Lassen, K. and Danielsen, C.: 1978, *J. Geophys. Res.* **83**, 5277.

Lee, L. C. and Fu, Z. F.: 1985, *Geophys. Res. Letters* **12**, 105.

Lee, L. C., Kan, J. R., and Akasofu, S.-I.: 1985, *J. Geophys. Res.* **57**, 217.

Lemaire, J.: 1977, *Planetary Space Sci.* **25**, 887.

Lemaire, J.: 1979, *Eur. Space Sci. Agency, ESA* **SP-148**, 365.

Lemaire, J.: 1985, *Proceedings from AGU Chapman Conf. on Magnetotail Physics*, Laurel, Maryland, October 28–31, 1985.

Lemaire, J., Rycroft, M. J., and Roth, M.: 1979, *Planetary Space Sci.* **27**, 47.

Lundin, R.: 1984, *Planetary Space Sci.* **32**, 757.

Lundin, R. and Dubinin, E. M.: 1984, *Planetary Space Sci.* **32**, 745.
Lundin, R. and Evans, D. S.: 1985, *Planetary Space Sci.* **33**, 1389.
Lundin, R., Eliasson, L., and Sandahl, I.: 1987, *Phys. Scripta* **35** (in press).
Marklund, G.: 1984, *Planetary Space Sci.* **32**, 193.
McPherron, R. L.: 1979, *Geophys. Res. Space Phys.* **17**, 657.
Meng, C.-I. and Lundin, R.: 1986, *J. Geophys. Res.* **91**, 1572.
Mozer, R. L.: 1984, *Geophys. Res. Letters* **11**, 135.
Olson, W. P. and Pfitzer, K. A.: 1985, *J. Geophys. Res.* **90**, 10823.
Paschmann, G.: 1984, *Eur. Space Agency Spec. Publ. ESA* **SP-217**, 53.
Paschmann, G., Sonnerup, B. U. Ö., Papamastorakis, I., Sckopke, N., Haerendel, G., Bame, S. J., Asbridge, J. R., Gosling, J. T., Russell, C. T., and Elphic, R. C.: 1979, *Nature* **282**, 243.
Perrault, P. and Akasofu, S.-I.: 1978, *Geophys. J.R. Astron. Soc.* **54**, 547.
Potemra, T. A. and Zanetti, L. J.: 1985, in J. A. Holtet and A. Egeland (eds.), *The Polar Cusp*, NATO Adv. Study Inst. Series C, Vol. 145, D. Reidel Publ. Co., Dordrecht, Holland, p. 203.
Primdahl, F., Marklund, G., and Sandahl, I.: 1986, *Danish Space Res. Inst. Rep. 7–86,* Lyngby, Denmark.
Reiff, P. H. and Burch, J. L.: 1985, *J. Geophys. Res.* **90**, 1595.
Rijnbeek, R. P., Cowley, S. W. H., Southwood, D. J., and Russell, C. T.: 1984, *J. Geophys. Res.* **89**, 786.
Rostoker, G.: 1983, *J. Geophys. Res.* **88**, 6981.
Rostoker, G., Akasofu, S.-I., Foster, J., Greenwald, R. A., Kamide, Y., Kawasaki, K., Lui, A. T. Y., McPherron, R. L., and Russell, C. T.: 1980, *J. Geophys. Res.* **85**, 1663.
Russell, C. T. and Elphic, R. C.: 1979, *Geophys. Res. Letters* **6**, 33.
Sandholt, P. E.: 1987, *Ann. Geophys.* **5A**, 219.
Sandholt, P. E., Deehr, C. S., Egeland, A., Lybekk, B., Viereck, R., and Romick, G. J.: 1986a, *J. Geophys. Res.* **91**, 10063.
Sandholt, P. E., Egeland, A., and Lybekk, B.: 1986b, *J. Geophys. Res.* **91**, 12108.
Sandholt, P. E., Egeland, A., Holtet, J. A., Lybekk, B., Svenes, K., Asheim, S., and Deehr, C. S.: 1985, *J. Geophys. Res.* **90**, 4407.
Sandholt, P. E., Lybekk, B., Egeland, A., Bythrow, P. F., and Hardy, D. A.: 1987, *J. Geophys. Res.* (submitted).
Sato, T., Shimada, T., Tanaka, M., Hayashi, T., and Watanabe, K.: 1986, *Geophys. Res. Letters* **13**, 801.
Saunders, M. A., Russell, C. T., and Sckopke, N.: 1984, *Geophys. Res. Letters* **11**, 131.
Sckopke, N., Paschmann, G., Haerendel, G., Sonnerup, B. U. Ö, Bame, S. J., Forbes, T. G., Hones, E. W., Jr., and Russell, C. T.: 1981, *J. Geophys. Res.* **86**, 2099.
Sergeev, V. A., Dmitrieva, N. P., and Barkova, E. S.: 1986, *Planetary Space Sci.* **34**, 1109.
Shepherd, G. G., Bostrøm, R., Derblom, H., Fälthammar, C. G., Gendrin, R., Kaila, K., Korth, A., Pedersen, A., Pellinen, R., and Wrenn, G.: 1980, *J. Geophys. Res.* **85**, 4587.
Smiddy, M., Burke, W. J., Kelley, M. C., Saflekos, N. A., Gussenhoven, M. S., Hardy, D. A., and Rich, F. J.: 1980, *J. Geophys. Res.* **85**, 6811.
Sonnerup, B. U. Ö., Paschmann, G., Papamastorakis, I., Sckopke, N., Haerendel, G., Bame, S. J., Asbridge, J. R., Gosling, J. T., and Russell, C. T.: 1981, *J. Geophys. Res.* **86**, 10049.
Southwood, D. J.: 1987, *J. Geophys. Res.* **92**, 3207.
Sugiura, M., Iyemori, T., Hoffman, R. A., Maynard, N. C., Burch, J. L., and Winningham, J. D.: 1984, in T. A. Potemra (ed.), *Magnetospheric Currents, Geophys. Monograph*, Vol., 28, AGU, Washington, D.C., p. 96.
Troshichev, O. A.: 1982, *Space Sci. Rev.* **32**, 275.
Vorobjev, V. G., Starkov, G. V., and Feldstein, Ya. I.: 1975, *Planetary Space Sci.* **23**, 269.
Vorobjev, V. G., Starkov, G. V., and Feldstein, Ya. I.: 1976, *Planetary Space Sci.* **24**, 955.
Zverev, V. L., Vorobjev, V. G., Starkov, G. V., and Feldstein, Ya. I.: 1986, in *Proceedings from International Symposium on Polar Geomagnetic Phenomena*, Souzdal, May 1986, p. 89.

PLASMA ACCELERATION, INJECTION, AND LOSS: OBSERVATIONAL ASPECTS*

C. E. McILWAIN

Department of Physics, and Center for Astrophysics and Space Science, University of California, San Diego, California, U.S.A.

(Received 10 September, 1987)

Abstract. The sudden and dramatic acceleration of charged particles seems to be a universal phenomenon which occurs in plasmas occupying a wide range of spatial scales. These accelerations are typically accompanied by intrusions of the energized plasma into adjacent regions of space. A physical understanding of these processes can only be obtained by carefully coordinated experimental and theoretical studies which are designed to let nature display what is happening without imposing limitations associated with existing paradigms. Studies of the Earth's magnetosphere are hampered by the lack of adequate sampling in space and time. The 'feature matching' technique of building magnetic and electric field models can help compensate for the extreme sparseness of experimental data but many future studies will still require large numbers of spacecraft placed in carefully coordinated orbits. History shows that magnetospheric research has sometimes faltered while various attractive conjectures were explored, but that direct observations play the role of a strict teacher who has little concern for the egos of scientists. Presumably this teacher will also discard the author's pet notion: that the 'ignition' of portions of the 'auroral shell' in association with 'Earth flares' results in the heating of ionospheric particles (and some particles of solar origin) that are then convected inward to form the ring current. The author, of course, hopes that at least some aspects of this notion will surive and will help lead the way to a better understanding of the Earth's neighbourhood.

1. Introduction

Parker (1984) has pointed out that magnetized plasmas universally exhibit activity. He further noted that this fact 'is forced upon us by observations', and that:

> "Theory anticipated none of it and has been succesful in understanding activity only where detailed observations are available to define the problem. For the fact is that a magnetized plasma has too many degrees of freedom to allow direct deduction of its behaviour from the basic dynamical equations for an aggregate of charged particles... Consequently, a subtler approach has been necessary, sparked by observations of discovery, followed by mapping and probing along with the development of theoretical understanding of each special situation."

This dependence of our understanding of plasmas upon observation has been long recognized by H. Alfvén, who recently noted (Alfvén, 1986a) that

> "*In situ* measurements in magnetospheric plasmas (including the solar wind) have caused drastic changes in our views of the properties of cosmic plasmas. What was considered sacrosanct ten or even five years is now hopelessly obsolete."

One reason measurements of plasmas in space have provided insights not forthcoming from laboratory experiments has been the ability to separately measure the

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 201–213.

distribution function of each charged particle species over a wide range of velocities and with high resolution. It is typically found that these distributions are replete with fascinating features (McIlwain and Whipple, 1986). These features are created by the plasma processes previously experienced by the measured particles, and thus they are a historical record of these processes, and can be studied in much the spirit that geologists study rock formations.

2. Magnetospheric Research Objectives and Data Gaps

The objective of studying plasmas in the vicinity of the Earth is simply to discover what is happening and why. In other words, we wish to obtain a physical understanding and accounting for all particles and fields within each volume of space. This includes knowledge of the history and probable future behaviour of each particle.

Unfortunately, experimental observations by themselves are incapable of providing a good picture of what is happening. Dramatic strides are being made in the development of instruments for measuring almost all aspects of the fields and particles in the vicinity of a spacecraft. Regrettably, however, there are few remote sensing techniques available for studying plasmas of the densities found in the Earth's magnetosphere above ionospheric altitudes. This means that most measurements will be sampled at no more points in space than there are spacecraft. In the past, this has meant that most studies had to be performed using data from only a few points in space and often only one. The coordinated launching of spacecraft by many nations during the International Solar Terrestrial Physics Program (ISTP) will make a substantial improvement in this situation but even an armada of 300 spacecraft as proposed for the 'Global Current Mission' (National Research Council, 1988) would still be woefully short of providing adequate sampling in the Nyquist sense.

The sparsity of the typical data set is highlighted by the difficulties encountered in any serious attempt to solve an inverse problem, for example, the determination of the electrical current systems from magnetic field measurements. The poor space-time coverage available using the data from any reasonable number of spacecraft precludes any formal solution that would reflect the true nature of the complex three-dimensional time-dependent vector field which represents the Earth's magnetospheric current systems. One exciting component of ISTP is the 'CLUSTER' mission in which a set of spacecraft will be put in coordinated orbits that are periodically adjusted to optimize the data on small- and medium-sized magnetospheric structures. Analysis of data from the CLUSTER spacecraft will not be easy but surely solutions for the local currents will be feasible. Perhaps analysis techniques can be generated analogous to the semi-norm minimization procedure which Parker (1987) developed in applying inverse theory to seamount magnetization.

In a paper recommending the establishment of additional magnetic observatories, Fleming (1932) commented:

"In geophysics we must regard our Earth as one great laboratory in which nature is constantly carrying on experiments. Unfortunately means are not yet available by which the magnetic and electric phenomena of the Earth may be observed from a distance as is the case in the astrophysical laboratory of space. In many ways this places the geophysicist at a much greater disadvantage than the astrophysicist because he must establish many observatories to get data for statistical studies."

While the space age has only compounded the problem of having too few observatories (three dimensions need to be sampled, not just two), at least the techniques for 'observing from a distance' that are available for the study of 'the astrophysical laboratory' are now available for studies of the Earth and its environment. The observation of global auroral patterns from space has, of course, been very helpful in providing clues concerning the large-scale behaviour of the magnetosphere.

3. Gap Filling with the 'Feature Matching' Technique

One way of filling the huge data gaps in observations of magnetospheric plasmas is to construct model magnetic and electric fields and to compute where the observed particles were in the past and where they will be in the future. As mentioned above, the observed distribution functions almost always have numerous well-defined features, i.e., peaks, discontinuities, or simply changes of slope with respect to velocity, space, or time. If the model fields are locally correct, it should be possible to 'match up' similar features at neighbouring spacecraft. If they are globally correct, it should be possible to identify a spatial boundary, an event in time, or some special process corresponding to each feature.

Obviously the field models should be consistent with the available field data, but in general models constructed using only the measured fields cannot be expected to accurately 'match up' the particle features measured at different locations. If the particle and field data are not too sparse, it should be possible to correctly identify which features should be matching up and to adjust the field models to obtain a more accurate matching.

This procedure of using the naturally present particles as tracers to find a consistent model can be viewed as an inverse solution (albeit one without a formal method for guessing the model fields) and in common with all inverse solutions involving real data, there can be no guarantee of uniqueness. Finding physically meaningful solutions obviously calls for considerable human judgment, a great deal of science and art, and the assistance of fast computers.

4. Real Data and Undiscovered Processes

In the magnetosphere, unlike in computer simulations, all microscopic and macroscopic processes are unavoidably present and it cannot be assumed *a priori* that any particular process can be ignored. This, of course, includes processes which are as yet undiscovered. Studies of real data, therefore, must be ready to reveal the unexpected. In other words, when searching for models which are consistent with the data, we should somehow let nature tell us what is going on and take care not to ignore her messages when they seem to be contrary to our current beliefs.

Rapid variations in space or time, such as found at boundaries or in transient events, can cause an almost absurd degree of undersampling and thus present special problems in obtaining an understanding of the inner details. Likewise, strong wave activity and associated stochastic processes require special treatment.

5. Moving Field Lines and Coordinate Systems

One of the first steps in the solution of any physical problem is to choose an appropriate coordinate system. In plasma physics, coordinate systems are often constructed such that the electric field is minimized. In some circumstances, for certain types of plasmas, this greatly simplifies studies of the large-scale behaviour. On the other hand, a coordinate system which has a complicated space and time dependence (i.e., 'writhes') and which depends upon the solution being sought, would seem to be a poor choice for analytic work. Now that it is almost certain that inductive electric fields (Mauk, 1986) and electric fields parallel to the magnetic field (Fälthammar, 1986) play key roles in determining space plasma behaviour, it also seems rather dangerous (Alfvén, 1986b) to use the concept of moving field lines even for 'cartoon' physics.

6. Partcle Motions in the Earth's Magnetosphere

The study of charged particle motions in the analytical models of the Earth's magnetosphere began in earnest in the winter of 1902 when Störmer (1930):

> "saw the remarkable experiments of my friend Professor Birkeland on cathode-rays in a magnetic field. ... At once it occurred to me that here was an interesting mathematical problem, namely, to find the trajectories of an electric corpuscle moving in the magnetic field around such a magnetized sphere. ... But in the case of the aurora, a series of simplifying hypotheses had to be made to make mathematical problem not too difficult."

These simplifications included the use of a dipole magnetic field and an electric field that was everywhere zero. After he and his students had performed about 5000 hours of calculations during the period of 1904 to 1907, he was able to state (Störmer, 1930):

> "A long series of characteristic properties of the aurora can be explained by this theory better than by any previous theory, but in the case of the auroral belts of the Earth, the theory gives too small a diameter for this zone."

While electronic computers did not arrive in time for Störmer to explore more realistic models of the Earth's fields, his work provided an excellent foundation for future studies. One valuable concept produced by his work was the existence of allowed and forbidden regions. The dark regions in Figure 1 show how he found the forbidden regions in a dipole field to change with respect to one of the constants of motion.

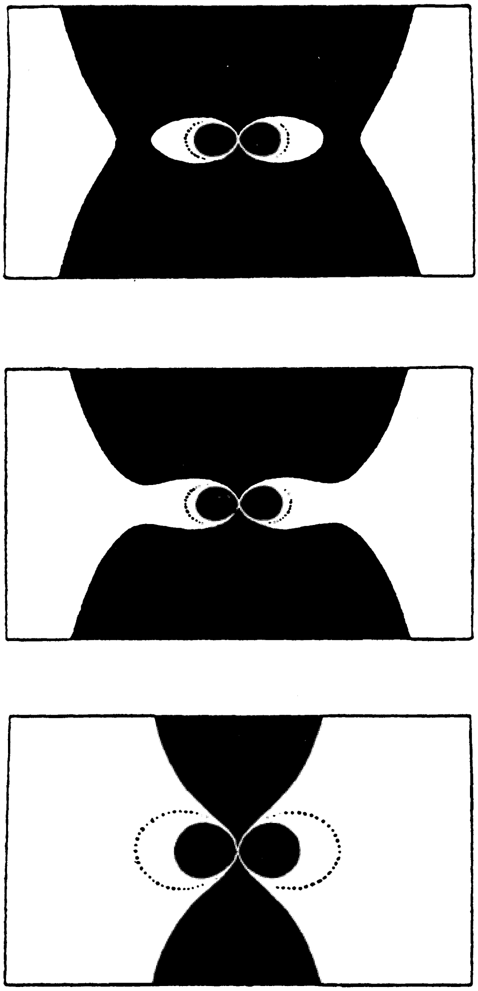


Fig. 1. Forbidden regions in a dipole magnetic field for three different constants of the motion of a charged particle (Störmer, 1930).

7. The Magnetopause Current System

In their impressive work on magnetic storms, Chapman and Ferraro (1931) considered 'the phenomena accompanying the advance of a solar stream into the Earth's field'. From this work, it became clear that the magnetic field in the vicinity of 10 earth radii and beyond was unlikely to resemble a dipole field. While they made considerable progress in understanding the interfaces between the solar wind and the Earth's magnetosphere (as have many others in the following half century), this topic will remain a controversial subject far into the future.

8. The Ring Current

Each magnetic storm is accompanied by a world wide depression of the Earth's magnetic field. Schmidt (1924) felt that the uniformity and slow decay of these perturbations were evidence for an external 'ring-current' circulating round the Earth in the plane of the magnetic equator.

Chapman and Ferraro (1932) concluded that the solar stream would induce a polarization electric field and suggested that:

> "... the outermost ions (and electrons after them) will tend to spiral round the Earth in a ring, possibly in a manner suggested by some of Störmer's trajectories of a single ion moving in the Earth's field; and we do not think it improbable that the current-ring will be formed by this tendency which the positives have to spiral around the lines of force."

In the absence of high altitude measurements, unfortunately, it is not possible to deduce the vertical distribution of external currents, so Chapman preferred to analyze all magnetic field perturbations in terms of equivalent currents flowing at ionospheric altitudes and stated (Chapman and Bartels, 1940):

> "... until further supporting evidence for the existence of a ring is obtained... the ring hypothesis must remain in suspense."

9. The Solar Wind Electric Field and Magnetospheric Convection

'A Theory of Magnetic Storms and the Aurorae' (Alfvén, 1939), was a landmark paper in many ways (Dessler and Wilcox, 1970; Dessler, 1983). One way was the recognition of the importance of the electric field associated with the solar wind (which was then referred to as the 'beam'). In his book *Cosmical Electrodynamics*, Alfvén (1950) writes:

> "Considering also the polarization in the solar magnetic field, a large-scale electric field is obtained with the same extension as the beam. ... Hence, the entry of a beam into the neighbourhood of the Earth means that the Earth is placed in a general electric field."

By assuming that this solar wind electric field penetrated into the Earth's magnetosphere, he was able to show (see Figure 2) how the 'forbidden region' could be made to agree more closely with auroral observations than in Störmer's theory. He also noted (Alfvén, 1950) that:

> "... the drifting electrons are equivalent to a ring current of about the type needed to explain the equatorial disturbance."

10. Electric and Magnetic Field Observations

The space age has resulted in many measurements of the magnetic field in almost all regions of the Earth's magnetosphere. While there will probably never be a sufficient number of simultaneous observations of the magnetic field to yield an adequate sampling for many types of analyses, the situation with regard to the Earth's electric field is far

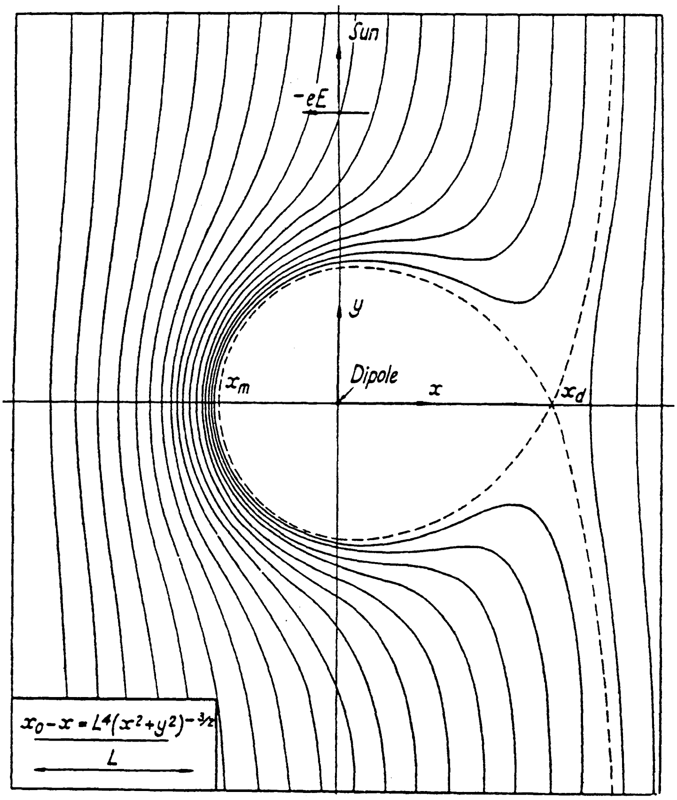


Fig. 2. Motion of electrons in a dipole plus a uniform magnetic field and a uniform electric field (Alfvén, 1939).

worse. Measurements of the electric field in the outer magnetosphere are, therefore, essential. However, they are also very difficult and have been attempted only to a limited extent. A combination of techniques may be necessary to achieve measurements in all regions of interest.

At ionospheric altitudes the direct measurements using double probes and the indirect measurements using incoherent radar are presumably reliable. The usefulness of these electric field observations for studies of the remainder of the magnetosphere is often limited, however, by the uncertainty in where the magnetic field lines extend into space and by the known presence of electric fields parallel to the magnetic field.

11. Initial Observations at 6.6 Earth Radii

The first high-resolution plasma observations in the Earth's magnetosphere, such as those made by the UCSD instrument on ATS-5, showed a wealth of unexpected detail (DeForest and McIlwain, 1971). As mentioned above, each separate detail in the distribution functions can be regarded as a piece of a historical record. Obviously a key aspect of a particle's history is what region of space it has been traversing. Sharp features can thus be expected at the energy and pitch angles corresponding to the trajectories which just trace back to a special region of space such as the magnetopause, magnetotail, or ionosphere.

The allowed and forbidden regions given by the Störmer and Alfvén studies show that even the simplest of field models will predict a strong local time dependence in the energy and sign of charge required for a particle to reach ATS-5's radial distance of 6.6 Re from a boundary such as the magnetopause. Following the iterative procedure suggested above, an electric field model was constructed (McIlwain, 1972) which seemed to achieve a reasonable 'match up' between different sets of ATS-5 observations that were chosen to be 'equivalent' (to be sure, this was a poor substitute for simultaneous observations at multiple points in space). A 'map' of the boundaries between the different categories of trajectories predicted by this model is shown in Figure 3.

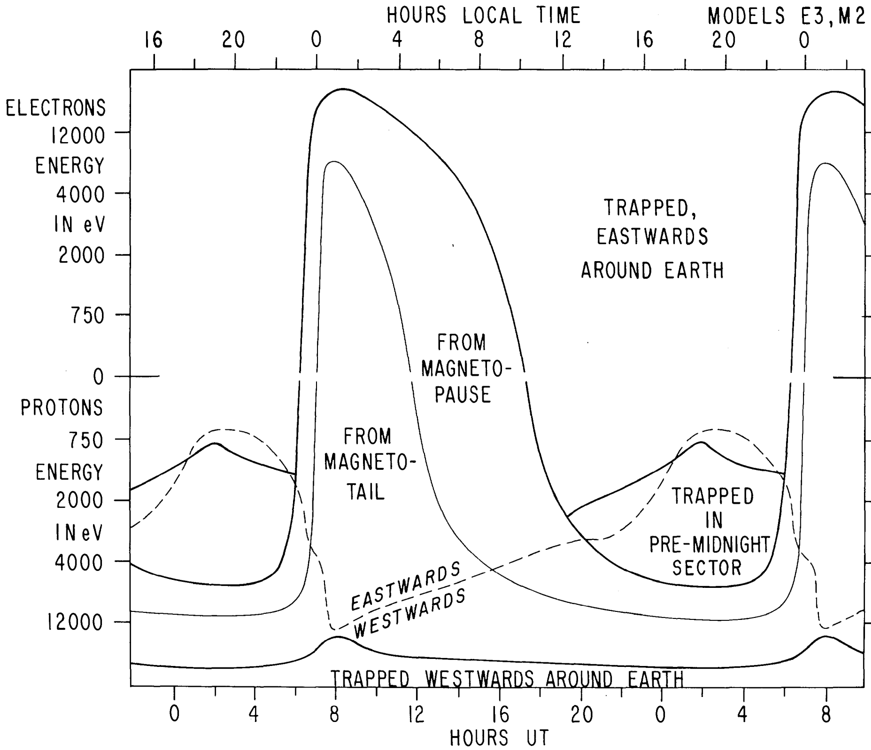


Fig. 3. The charge times energy versus local time for various kinds of trajectories encountered in geosynchronous orbit (McIlwain, 1972).

The distribution function features produced by the loss processes in the ionosphere and the atmosphere are particularly useful. The size of the loss cone generated by these losses changes as the particles drift in radius and longitude. The shape of the measured angular distributions thus contains clues as to the previous history of the particles. The wide deep loss cone exhibited by the ions returning from deep radial excursions into the inner magnetosphere is easily identifiable and forms a deep minimum in the energy spectra at small and medium pitch angles. In time-independent models, it can be shown that the energy of this minimum is a simple function of the electrical potential.

The presence of electric fields parallel to the magnetic field can of course leave clear signatures in the angular distribution (Whipple, 1977; Kaufmann, 1984). For particles of the appropriate sign of charge, a 'source cone' (McIlwain, 1975) can be generated by parallel fields.

12. Substorms and Injection Boundaries

The most dramatic features in the ATS-5 observations were clearly associated with magnetospheric substorms. It was found (McIlwain, 1974) that a consistent picture was obtained if it was assumed that the leading edge of enhanced fluxes were all located along an 'injection boundary' at the onset of a substorm. Similar studies have now been performed with observations at other radial distances (Greenspan *et al.*, 1985; Williams, 1987) and latitudes (Newell and Meng, 1987) providing considerable confirmation of this picture.

13. The Auroral Shell

'The auroral shell' has been defined (McIlwain, 1985) to be the surface formed by the magnetic field lines connecting the equatorward edges of the auroral northern and southern ovals. As indicated above, the 'injection boundary' is the inner edge of the region promptly filled with hot plasma in association with magnetospheric substorms. The equatorward edge of the auroral zone has long been known to be the first region to brighten at substorm onset and thus also marks the beginning and the inner limit of renewed active plasma acceleration. There is thus every reason to believe that the auroral shell and the injection boundary are identical. Analyses of simultaneous auroral and equatorial data (Akasofu *et al.*, 1974; Eather *et al.*, 1976) confirm that they are on the same field lines within the accuracy of the available magnetic field models.

The average location of the auroral shell for a given local time LT and a given magnetic disturbance level K_p can be found using a model magnetic field and computing field lines starting at the intersection of the auroral shell with the magnetic equator. The radial distance of this intersection R_{ar} is approximated by the equation (McIlwain, 1986)

$$R_{ar} = 9.8 - 1.4\cos(\mathrm{LT}) - [0.9 + 0.3\cos(\mathrm{LT})]K_p/(1 + 0.1K_p)\,.$$

14. Energetic Ionospheric Ions in the Magnetosphere

Some aspects of the picture presented above, in which the ring current is injected by dynamic processes during each substorm, is not radically different from some earlier pictures. Almost all earlier pictures, however, fail to accommodate the dramatic new observations which show that substantial fractions of the energetic particles found in most regions of the magnetosphere are of ionospheric origin (Johnson, 1979; Chapell *et al.*, 1987). Perhaps substorms are 'Earth flares' (McIlwain, 1969) in which plasma of local origin is heated, sometimes quite suddenly. Presumably the solar wind is the source of the energy (perhaps with some assistance from the Earth's rotational motion) but the path this energy takes has not been clearly revealed by direct measurements.

15. Substorm Currents and Inductive Surges

Mauk (1986) has shown how the inductive electric field produced by the rapid magnetic field 'dipolarization' observed in association with substorms can produce the bouncing clusters of ions also observed at such times (Quinn and McIlwain, 1979).

Kaufmann (1987) has constructed model magnetic fields and particle distributions and considered the flow of the associated currents. He finds that:

> "... a modest increase in particle energy density can produce a substantial increase in cross-tail current. ... This process involves a positive feedback effect, with pre-existing particles carrying more cross-tail current as soon as any perturbation begins to stretch tail field lines."

He then models a substorm onset by diverting current to the ionosphere in a wedge near midnight and finds that:

> "field lines within the wedge collapse dramatically even if only a portion of the cross-tail current is diverted. ... It is suggested that diversion of only the electron cross-tail current to the ionosphere is enough to initiate a substorm."

Figure 4 shows the result of some of his calculations in which cross-tail currents are added to a quiet magnetosphere model and then progressively diverted to the ionosphere along magnetic field lines. While the behaviour of the outer portion of the magnetotail must be included in any comprehensive picture of the magnetosphere, Kaufmann's studies indicate that it may not play a crucial role with respect to substorms. He comments:

> "Finally, if the particles near 8 Re are very strongly confined to the equator ..., it is possible that they could have been missed by the satellites with highly elliptical orbits which have studied the inner plasma sheet.
>
> ... the above points emphasize the importance of good pitch angle and flux measurements by a satellite that remains at a nearly fixed location between $R = 6.6$ and 10 Re throughout substorm growth phase."

The images of auroral patterns obtained by the Viking spacecraft have revealed that typically only a portion of the auroral shell is ignited at substorm onset (Shepherd *et al.*, 1987; Rostoker *et al.*, 1987) and thus lend support to Kaufmann's picture of wedges of current being diverted each substorm.

16. The Ring Current in Hindsight

In retrospect, there was probably enough theoretical development and sufficient observations to have surmised significantly more about the nature of the ring current

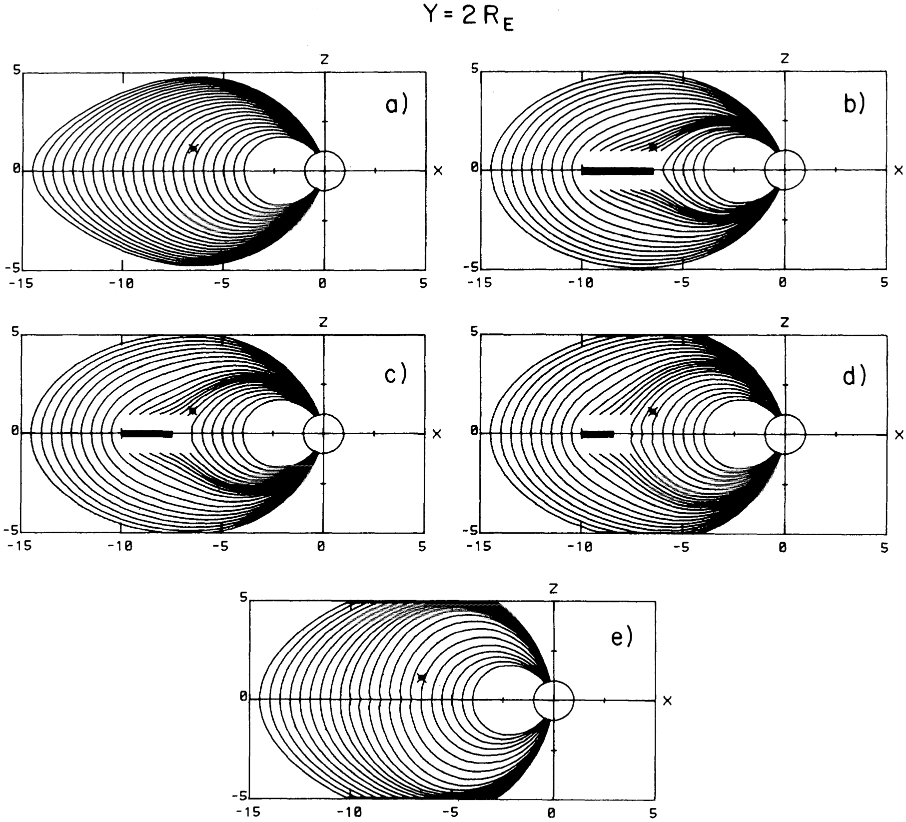


Fig. 4. The effects (b) of adding cross-tail currents to a quiet magnetosphere model (a) and diverting increasing portions to the ionosphere (Kaufmann, 1987).

during the period before the discovery of the radiation belts (Van Allen *et al.*, 1958). Perhaps even H. Alfvén was diverted by some aspects of the prevailing paradigms. After showing that his perturbation method (the guiding center method) was valid for 100 keV electrons almost out the Moon's orbit, he states (Alfvén, 1950):

> "For protons with the same energy the region of validity is restricted to about 7 times the Earth's radius, but such particles cannot be expected to be of any importance in the physics of the Earth."

Protons in just this energy region are now known (Krimigis *et al.*, 1985; Williams, 1985, 1987) to be the primary ion in the ring current region and are so numerous that they carry a large fraction of the total particle energy stored in the Earth's magnetosphere.

The development of the guiding center method was, of course, a monumental achievement and his conformance with the prevailing view about the absence of energetic protons is of minor importance. Alfvén's pre-space-age ring current theories were in fact fundamentally sound and many aspects of these theories help form the basis of our present day concepts of ring current physics. Indeed, even before the first Sputnik was launched, Singer (1957) had his interest in magnetic storms

> "aroused by some stimulating and controversial lectures given at the Universty of Maryland by E. H. Vestine and H. Alfvén during the spring of 1954."

He (Singer, 1957) applied Alfvén's guiding center theory to show that the observed magnetic perturbations and decay time could be produced by the entrapment of about 1 cm^{-3} of protons with velocities of 2×10^8 cm s^{-1} (21 keV). The spatial distribution that he surmised was rather similar to the distributions found by direct measurements (Akasofu and Chapman, 1961; Williams, 1987). His injection mechanism was an inward leakage of solar wind particles during magnetic field deviations (but he did not invoke the assistance and acceleration that Alfvén's solar electric field would have provided). While the leakage mechanism now appears to be relatively unimportant, it will be interesting to see if any of the mechanisms now in fashion are better at standing the test of time.

The paper 'Evolution of Ideas in Solar-Terrestrial Physics' by Akasofu (1983) should be consulted for a more thorough discussion of the historical development of ring current concepts.

17. Conclusion

The motion of charged particles in even the simplest models of the Earth's magnetic and electric fields is exceedingly complex. Unfortunately, the time and space dependencies of the real fields in the Earth's magnetosphere are probably as complex as the weather in the Earth's lower atmosphere. Studies of the motions of an ensemble of particles in any realistic model thus tax the abilities of the fastest computers and the nimblest human minds. This remains true even when the particles are non-interacting test particles. Surely, we can forgiven if some of our conjectures based upon fragmentary data turn out to be false and misleading.

The arrival of each new generation of computers should lead to successively greater insight into active plasma processes, but in the end, we must agree with Parker (1984) that:

> "Only a 'supercomputer' could combine the interacting effects to mock up the total situation. As a matter of fact, nature has already done that for us, the results of which we read out through our observations."

Acknowledgements

The assistance of R. W. Fillius is acknowledged. Partial funding was provided by NASA Grant NGL 05-005-007.

References

Akasofu, S.-I.: 1983, *Geophys. J.R. Astron Soc.* **74**, 257.
Akasofu, S.-I. and Chapman, S.: 1961, *J. Geophys. Res.* **66**, 1321.
Akasofu, S.-I., DeForest, S., and McIlwain, C.: 1974, *Planetary Space Sci.* **22**, 25.

Alfvén, H.: 1939, *Kungl. Sv. Vet.-Akademiens Handl. III* **18**, No. 3, 1, and No. 9, 1.

Alfvén, H.: 1950, *Cosmical Electrodynamics*, Claredon Press, Oxford.

Alfvén, H.: 1986a, *IEEE Trans. Plasma Sci.* **PS-14**, 629.

Alfvén, H.: 1986b, *IEEE Trans. Plasma Sci.* **PS-14**, 779.

Chapman, S. and Bartels, J.: 1940, *Geomagnetism,* Oxford University Press, London, p. 880.

Chapman, S. and Ferraro, V. C. A.: 1933, *Terrest. Mag. Atmos. Elect.* **38**, 79.

Chappell, C. R., Moore, T. E., and Waite, J. H., Jr.: 1987, *J. Geophys. Res.* **92**, 5896.

DeForest, S. E. and McIlwain, C. E.: 1971, *J. Geophys. Res.* **76**, 3587.

Dessler, A. J.: 1983, *Geophys. Monograph* **28**, 22.

Dessler, A. J. and Wilcox, J. M.: 1970, *EOS Trans. AGU* **51**, 180.

Eather, R. H., Mende, S. B., and Judge, R. J. R.: 1976, *J. Geophys. Res.* **81**, 2805.

Fleming, J. A.: 1932, *Terrest. Mag. Atmos. Elect.* **37**, 234.

Greenspan, M. E., Williams, D. J., Mauk, B. H., and Meng, C. I.: 1985, *J. Geophys. Res.* **90**, 4079.

Fälthammar, C.-G.: 1986, *IEEE Trans. Plasma Sci.* **PS-14**, 616.

Johnson, T. G.: 1979, *Rev. Geophys.* **17**, 696.

Kaufmann, R. L.: 1984, *Space Sci. Rev.* **37**, 313.

Kaufmann, R. L.: 1987, *J. Geophys. Res.* **92**, 7471.

Krimigis, S. M., Gloeckler, G., McEntire, R. W., Potemra, T. A., Scarf, F. L., and Shelley, E. G.: 1985, *Geophys. Res. Letters* **12**, 329.

Mauk, B. H.: 1986, *J. Geophys. Res.* **91**, 13423.

Mauk, B. H. and Zanetti, L. J.: 1987, *Rev. Geophys.* **25**, 541.

McIlwain, C. E.: 1969, *Ann. IQSY,* Vol. IV, *Solar Terrestrial Physics: Solar Aspects,* Vol. 302, M.I.T. Press, Cambridge.

McIlwain, C. E.: 1972, in B. M. McCormac (ed.), *Earth's Magnetospheric Processes*, D. Reidel Publ. Co., Dordrecht, Holland, pp. 268–279.

McIllwain, C. E.: 1974, in B. McCormac (ed.), *Magnetospheric Physics*, D. Reidel Publ. Co., Dordrecht, Holland, pp. 143–154.

McIlwain, C. E.: 1975, in B. Hultqvist and L. Stenflo (eds.), *Physics of the Hot Plasma in the Magnetosphere*, Plenum Publ. Corp., New York, pp. 91–112.

McIlwain, C. E.: 1985, *Results of the ARCAD 3 Project and the Recent Programmes in Magnetospheric and Ionospheric Physics*, Cepadues-Editions, Toulouse, France, pp. 275–280.

McIlwain, C. E. and Whipple, E. C.: 1986, *IEEE Trans. on Plasma Sci.* **PS-14**, 874.

National Research Council: 1988, *Space Science in the Twenty-First Century,* National Academy Press, Washington.

Newell, P. T. and Meng, C.-I.: 1987, *J. Geophys. Res.* **92**, 7519.

Parker, E. N.: 1984, in E. W. Hones Jr. (ed.), *Geophys. Monograph* **30**, 32.

Parker, R. L.: 1987, *Rev. Geophys.* **25**, 17.

Quinn, J. M. and McIlwain, C. E.: 1979, *J. Geophys. Res.* **84**, 7365.

Rostoker, G., Vallance Jones, A., Gattinger, R. L., Anger, D. D., and Murphree, J. S.: 1987, *Geophys. Res. Letters* **14**, 399.

Schmidt, A.: 1924, *Z. Geophys.* **1**, 1.

Shepherd, G. G., Anger, D. D., Murphree, J. S., and Vallance Jones, A.: 1987, *Geophys. Res. Letters* **14**, 395.

Singer, S. F.: 1957, *Trans. Am. Geophys.* **38**, 175.

Störmer, Carl: 1930, *Terrest. Mag. Atmos. Elec.* **35**, 193.

Van Allen, J. A., Ludwig, G. H., Ray, E. C., and McIlwain, C. E.: 1958, *Jet Propulsion* **28**, 588.

Whipple, E. C.: 1977, *J. Geophys. Res.* **82**, 1525.

Williams, D. J.: 1985, *Space Sci. Rev.* **42**, 375.

Williams, D. J.: 1987, *Rev. Geophys.* **25**, 570.

CHARGED DUST IN THE EARTH'S MAGNETOSPHERE

I. *Physical and Dynamical Processes*

M. HORANYI

Supercomputer Computational Research Institute, Florida State University, Tallahassee, Florida, U.S.A.

H. L. F. HOUPIS

Space Physics Laboratory, Department of Atmospheric and Oceanic Science, The University of Michigan, Ann Arbor, U.S.A.

and

D. A. MENDIS

Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, U.S.A.

(Received 12 November, 1987)

Abstract. We have considered the electrodynamic effects on small Al_2O_3 spherules dumped into the Earth's magnetosphere in large quantities during solid rocket propellant burns. The charges acquired by these grains in all regions of the terrestrial environment (plasmasphere, magnetosphere, and solar wind) are modest. Consequently electrodynamic effects are significant only at the lower end of the dust size spectrum ($R_g \lesssim 0.1$ μm). In that case, the electrodynamic forces conspire with solar radiation pressure to eliminate the grians from the magnetosphere in a comparatively short time. Although not studied here in detail, we anticipate a similar fate for fine micrometeoroids entering the Earth's magnetosphere, with the electrodynamic effects playing an even more important role.

1. Introduction

Small (0.1–10 μm) sized Al_2O_3 spherules are dumped into the Earth's ionosphere/magnetosphere during solid rocket motor burns used for transfer of satellites from low Earth to geosynchronous orbit. The flux from one such burn (No. of impacts m^{-2} yr^{-1}) could exceed the natural micrometeroid flux in that size (Mueller and Kessler, 1985). These authors also discussed the dynamic and orbital evolution of such particles in considerable detail taking into account the effects of gravity, gas drag, and solar radiation pressure. The one aspect that was not considered by them is the role of electromagnetic (Lorentz) forces that are experienced by these grains (which are necessarily electrically charged in the ambient plasma and radiative environment) as they move relative to the terrestrial magnetic field.

In this paper we will study the physical and dynamical processes associated with this electrostatic charging process. The time evolution of the grain orbits, and spatial distribution, as well as their magnetospheric residence times will be considered in detail in a subsequent paper.

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 215–229.

2. Dynamics

In the Earth-centered inertial frame, the motion of a charged dust grain of mass m, is governed by the equation

$$m\ddot{\mathbf{R}} = \mathbf{F}_{\mathrm{G}} + \mathbf{F}_{\mathrm{LP}} + \mathbf{F}_{\mathrm{L}}\,, \tag{1}$$

where the three terms included on the right-hand side of Equation (1) are the gravitational, light-pressure, and Lorentz forces. In the region of interest, around the geosynchronous orbit, we may neglect the neutral gas and plasma (Coulomb) drags on the dust (Peale, 1966). Furthermore, we regard the interparticle distances to be larger than the Debye shielding length λ, so that interparticle Coulomb interactions are negligible.

To estimate the relative importance of these forces, we calculate their numerical values for a spherical grian with bulk density $\rho = 1\ \mathrm{g\ A}^{-3}$ and radius $r_g = 1$ (and 0.1) μm, in the equatorial plane at geosynchronous altitude with zero velocity having an electrostatic potential ϕ of 10 V with respect to the ambient plasma. In this case, the gravitational force $F_G = 9.4 \times (10^{-11}\text{–}10^{-14})$ dynes.

The light pressure force is given by $F_{LP} = \pi r_g^2 J_0 Q_{pr}/c$, where J_0 is the solar energy flux at 1 AU ($\approx 1.36 \times 10^6$ ergan^{-2} s^{-1}), Q_{pr} is the light scattering efficiency (assumed to be unity for both particles sizes) and c is the speed of light. For the two grains in question $F_{LP} = 1.42 \times (10^{12}\text{–}10^{-14})$ dynes.

The Lorentz force is given by $\mathbf{F}_L = Q(\mathbf{E} + \dot{\mathbf{R}} \times \mathbf{B}/c)$, where Q is the charge on the grain. For a small ($r_g \ll \lambda$) 'isolated' ($\lambda \ll d$, where d is the intergrain distance) grain $Q = \phi r_g$. Since we assume $\dot{\mathbf{R}} = \mathbf{0}$, we only need to consider the effect of $\mathbf{E}$, which has two components; $\mathbf{E}_{\text{co-rotation}}$ and $\mathbf{E}_{\text{cross-tail}}$. If we assume rigid co-rotation, $\mathbf{E}_{\text{co-rotation}} = -c^{-1}(\boldsymbol{\Omega} \times \mathbf{R}) \times \mathbf{B}_0(R_E/R)^3$ where Ω ($= 7.272 \times 10^{-5}$ rad s^{-1}) is the angular velocity of the Earth, B_0 ($= 0.31\Gamma$) is the equatorial surface magnetic field, and we have assumed the magnetic field to be a simple dipole field aligned with the rotatin axis. With the above values, $\mathbf{E}_{\text{co-rotation}} \approx 3.36 \times 10^{-6}$ V cm^{-1} at the geosynchronous orbit, and is pointed towards the Earth; since $\boldsymbol{\Omega}$ and $\mathbf{M}$ (the magnetic moment of the Earth) are antiparallel. The cross-field $\approx 2.5 \times 10^{-6}$ V cm^{-1}, pointing in the dawn to dust direction. The maximum electric field clearly occurs at the dawn meridian, where $\mathbf{F}_L \approx 6.3 \times (10^{-14}\text{–}10^{-15})$ dynes.

Comparing these numerical values, we conclude that for 1 μm grains, the Lorentz force is a minor perturbation, but for grains with radii ≈ 0.1 μm or smaller, the Lorentz force becomes comparable with the other two forces. It will thus play an important role in shaping the trajectories of these very small grains.

In order to follow the trajectory of a grain, one integrates Equation (1) with the current balance equation

$$\dot{Q} = \Sigma I_i\,, \tag{2}$$

where I_i are the various charging currents which will be discussed in the next section. Here it must be noted that due to the finite capacitance of the grain, the charge Q at any point depends not only on the local plasma parameters and radiation field but also on its previous charging history.

The optical and physical parameters of these Al_2O_3 spherules are not too well known. However, we will consider two essentially extreme cases. In one case (type 1), we will assume the grain to have the scattering efficiency Q_{pr} of conducting magnetite with $\rho \approx 3.2 \text{ g cm}^{-3}$, whereas in the other case (type 2) we will assume it to have the scattering efficiency of dielectric olivine with $\rho \approx 2.2 \text{ g cm}^{-3}$. Other material constants of these two types of grains (such as the photoelectron and secondary emission yields) will also be substantially different. These will be discussed in the next section.

3. Charging Currents

The right-hand side of Equation (2) contains all the charging currents to and from the surface of the grain. In the present study, we have the electron and ion (proton) thermal currents, secondary electron currents by electron and proton impact, backscattered and photoelectron currents. In the environment we consider, thermoionic and field emission currents can be shown to be negligible.

Below we summarize the expressions for the charging currents in the orbital motion limited approximation (Lafromboise and Parker, 1973). The thermal electron (e) and ion (i) collection currents are

$$I_{e,i} = eA \int_{\max(0,\, \pm e\phi)}^{\infty} \left[1 \pm \left(\frac{-e\phi}{E}\right)\right]\left(\frac{\mathrm{d}f_{e,i}}{\mathrm{d}E}\right) \mathrm{d}E\,, \tag{3}$$

where $A(4\pi r_g^2)$ is the grain surface area; e, the electron charge; $f_{e,i}$, the differential energy flux to the surface; E, the energy; and the $-$, $+$ signs correspond to electrons and ions, respectively. If we assume a Maxwellian distribution, Equation (3) can be integrated to yield

$$I_{e,i} = eAn_0\left(\frac{kT_{e,i}}{2\pi m_{e,i}}\right)^{1/2} \begin{cases} \exp(-\Psi_{e,i})\,, & \Psi_{e,i} > 0\,, \\ 1 - \Psi_{e,i}\,, & \Psi_{e,i} < 0\,; \end{cases} \tag{4}$$

where $\Psi_{e,i} = \mp e\phi/kT$; k being the Boltzmann factor and $T_{e,i}$ is the temperature parameter in the Maxwell distribution.

This assumption is valid only when the velocity of the grain is much smaller than the thermal velocity of both the electrons and the ions. Otherwise one has to consider a drifting Maxwellian in the frame of the grain, and a more complicated expression for $I_{e,i}$, containing also the grain speed is obtained (see Whipple, 1981; Mendis, 1981). In the present calculations, it is appropriate to use Equation (4) to calculate the electron and ion collection currents both within the plasmasphere and the magnetosphere. It is also appropriate to calculate the electron collection current in the solar wind, but the more general expression for I_i (Mendis, 1981) is required for the ion collection current in the solar wind. Consequently, the so-called radial gyro-phase drift associated with the orbital velocity modulation of the ion collection current by a finite capacitance grain (Hill and Mendis, 1980; Northrop and Hill, 1983) occurs only when the grain is in the solar wind, and is found to be a rather small effect even in this case.

The currents due to secondary electron emission by electron or proton impact, as well as backscattered electrons, can be calculated by incorporating the relevant yield function into Equation (3) as

$$I_s^k = eA \int_{\max(0,\ \pm e\phi)}^{\infty} \left[1 \pm \left(\frac{-e\phi}{E}\right)\right] \delta^k(E \pm (-e\phi)) \left(\frac{\mathrm{d}f_k}{\mathrm{d}E}\right) \mathrm{d}E\,, \tag{5}$$

where δ^k is the appropriate yield function associated with the process (k). The argument of δ^k is shifted due to the fact that the incoming particle looses or gains energy in approaching the grain through the potential field of the charged grain.

Equation (5) needs a further correction, because the secondary electrons will have a certain energy distribution, and in the case of an attractive potential only particles above the potential threshold will escape from the surface. In this calculation, we assume that the energy distribution of the secondary electrons is also a Maxwellian with a characteristic temperature parameter T^k. In this case, the correction factor for the attractive potential is $(1 - \Psi_s^k)\exp(\Psi_s^k)$, where $\Psi_s^k = -e\phi/kT_s^k$.

For secondary electrons by electron impact, we use (Sternglass, 1954) the equation

$$\delta^{SEE}(E) = 7.4\left(\frac{E}{E_M}\right)\delta_M^{SEE} \exp\left[-2\left(\frac{E}{E_M}\right)^{1/2}\right], \tag{6}$$

where δ_M and E_M are material-dependent parameters. For particles of type 1, we use $\delta_M^{SEE} = 1.5$ and $E_M^{SEE} = 250$ eV, and for type 2, we use $\delta_m^{SEE} = 2.4$ and $E_M^{SEE} = 400$ eV (Whipple, 1981). For both types we used $kT_s^{SEE} = 3$ eV (Whipple, 1981; Katz *et al.*, 1977).

For secondary electrons by proton impact we use,

$$\delta^{SEP}(E) = 2\delta_M^{SEP} \frac{\left(\dfrac{E}{E_M^{SEP}}\right)^{1/2}}{\left(\dfrac{1 + E}{E_M^{SEP}}\right)}\,, \tag{7}$$

where δ_M, E_M are the relevant material-dependent parameters. Due to the lack of laboratory data for both types of grains, we used $\delta_M = 4.3$ and $E_M = 40$ keV, and for the energy distribution of the secondary electrons produced, we used $kT_s^{SEP} = 3$ eV (Whipple, 1981; Katz *et al.*, 1977).

For the secondary electrons by electron and proton impact, the yield function δ^k depends on the angle of incidence θ as well. The angular dependence is approximately $\sec(\theta)$, which when averaged over an isotropic distribution is just $\langle \delta^k(\theta) \rangle_\theta = 2\delta^k$

($\theta = 0$), where δ^k ($\theta = 0$) is the yield for normal incidence discussed above (Draine and Salpeter, 1979). This assumption of isotropic impact is valid for electrons in the plasmasphere, magnetosphere, and the solar wind, and it is also valid for ions in the plasmasphere and magnetosphere, but not in the solar wind. In that case the coefficient z multiplying δ^k has to be replaced by $\propto$, where $1 < \propto < 2$.

For the backscattered electrons, we use a single step-function approximation

$$\delta^{BS}(E) = \begin{cases} 0\,, & E < E_{\min}\,; \\ 0.3\,, & E \geq E_{\min}\,; \end{cases} \tag{8}$$

which is in reasonable agreement with the analytic expression given by Katz *et al.* (1977). Taking $E_{\min} = 100$ eV for the energy distribution of the backscattered electrons, we use (cf. Prokopenko and Lafromboise, 1980) the equation

$$kT_s^{BE} = (0.45 + 2 \times 10^{-3} z)\,(kT_e + e\phi_s)\,, \tag{9}$$

where $(kT_e + e\phi_s)$ is the average incident electron energy and z, the atomic number of the target.

For all the secondary emission processes considered, we have neglected the possible decrease of secondary yield when the penetration depth of the bombarding particle becomes comparable to the size of the target, Given the average energies of the electrons and protons in the magnetosphere, this decrease is indeed small.

Finally, the photoelectron current is given by

$$I_{ph} = \begin{cases} \pi r_g^2 e f_1\,, & \phi < 0\,; \\ \pi r_g e f_1 \exp\left(\dfrac{-e\phi}{kT_s^P}\right), & \phi > 0\,; \end{cases} \tag{10}$$

where kT_s^P (≈ 2 eV) is the average energy of the assumed Maxwellian distribution of photoelectrons and $f_1 \approx 2.5 \times 10^{10}\,\chi\,\text{s}^{-1}$, with $\chi \approx 1$ for type 1 grains and $\chi = 0.1$ for type 2 grains (Mendis, 1981).

4. The Plasma Environment

In order to evaluate the various currents, we need to know the parameters of the plasma in which the grains are immersed. To describe the plasma environment of the Earth, we have followed Hill and Whipple (1985) with the modification that we include the local time-dependence of the plasma pause. Inside the plasma pause we assume thermal equilibrium between the ions and the electrons, with their density given by the empirical formula

$$n(L) = 10^{(15-L)/3.5}\ \text{cm}^{-3}\,, \tag{11}$$

where L denotes the magnetic shell, whose equatorial radius is L Earth radii. The temperature is likewise given by the empirical relation

$$T(L) = 0.09293L^{2.7073}\ \text{eV}\,, \tag{12}$$

when the temperature of the plasma given by (12) exceeds 1 eV, the plasma is considered to have two components, a hot one with T given by Equation (12) and $n = 1\ \text{cm}^{-3}$, and a cold one with $T = 1$ eV and density $1\ \text{cm}^{-3}$ less than that given by Equation (11).

The boundary of the plasmapause is given (cf. Lyons and Williams, 1984) by

$$L_{pp} = 7.6R_E\left(\frac{\sqrt{1 + \sin\psi} - 1}{\sin\psi}\right), \tag{13}$$

where Ψ is the local hour angle. Equation (13) defines a raindrop-like shape with $L_{pp} = 3.8$ at 12 and 24 hours, and $L_{pp} = 3.1$ and 7.6 toward 6 and 18 hous local time, respectively. Between the plasmapause and the magnetopause, we used the analytic expression fitted for the measured moments of the energy distribution of electrons and ions (Garrett and DeForest, 1979). The plasma is ocnsidered to have a bi-Maxwellian distribution. The data set is from the ATS-5 and ATS-6 satellites at geosynchronous orbit. Towards 18 hours local time, the plasmapause extends beyond the geosynchronous radius, but it leaves no obvious signature on the above satellite data. Perhaps this is because the detectors had a higher energy threshold, and it is for this reason that we kept the local time-dependence of the plasma pause. The bi-Maxwellian energy distribution, which is different for electorns and ions, depends on the daily geomagnetic activity index A_p and on local time (Garrett and DeForest, 1979). In this study we kept A_p constant at its average value of 120. The magnetospheric plasma density is of the order of $1\ \text{cm}^{-3}$, with characteristic energies of about 2 keV for electrons and about 5 keV for protons. The magnetopause which delineates the magnetospheric plasma, is described by a paraboloidal shape whose apex points toward the Sun. The subsolar point of the magnetopause is at $10.5R_E$ and it opens up to $14.5R_E$ at the terminators.

Beyond the magnetopause, we assume solar wind conditions with $n = 5\ \text{cm}^{-3}$ and $T \approx 10$ eV.

5. The Grain Charge

The current potential characteristics are plotted in Figures 1a – c), for the three different plasma regimes. Within the plasma sphere we have high plasma densities and relatively low thermal energies. For negative potentials, the proton collection current, the secondary electron current due to proton impact, and the photoelectron current are the major contributors to the total current. For positive potentials, the electron collection current dominates. The peak in the secondary electron current by electron impact is a result of the competition between two opposing tendencies. On the one hand, as the grain potential increases, more electrons with energies closer to the optimum for secondalry emission impact the surface. On the other hand, it becomes harder for the secondary electrons to leave the surface.

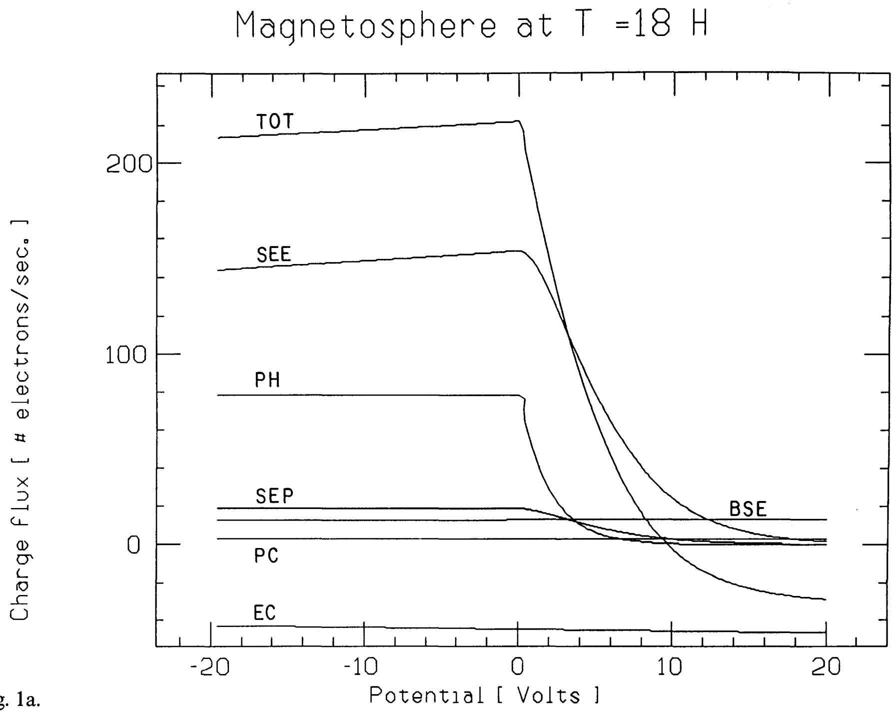

Magnetosphere at T =18 H
TOT
SEE
PH
SEP
PC
EC
BSE
200
100
0
-20
-10
0
10
20
Charge flux [# electrons/sec.]
Potential [Volts]

Fig. 1a.

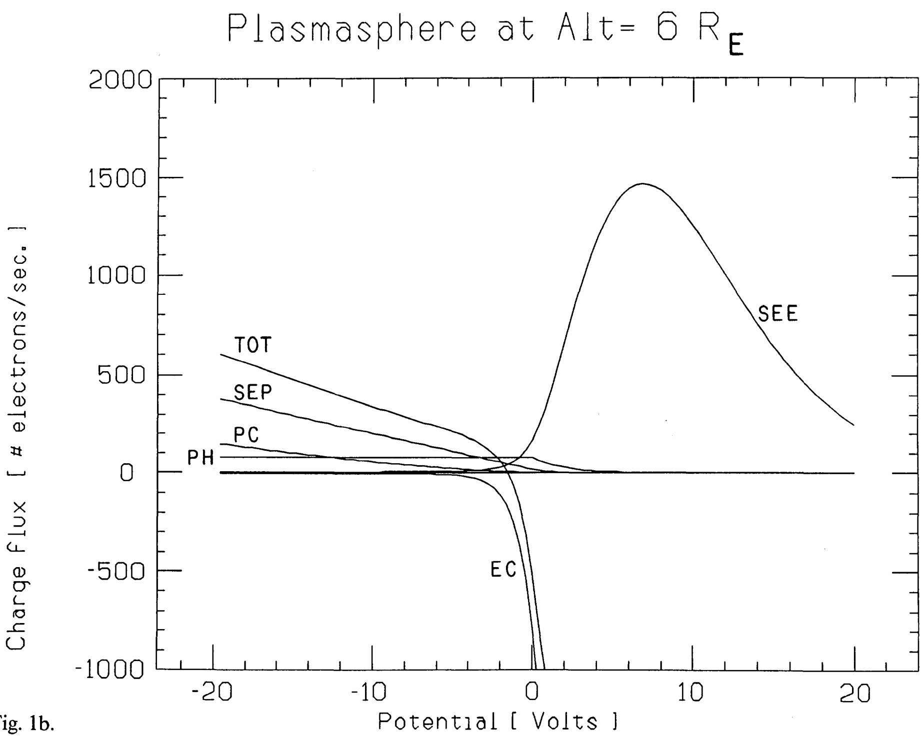

Plasmasphere at Alt= 6 R_E
TOT
SEP
PC
PH
EC
SEE
2000
1500
1000
500
0
-500
-1000
-20
-10
0
10
20
Charge flux [# electrons/sec.]
Potential [Volts]

Fig. 1b.

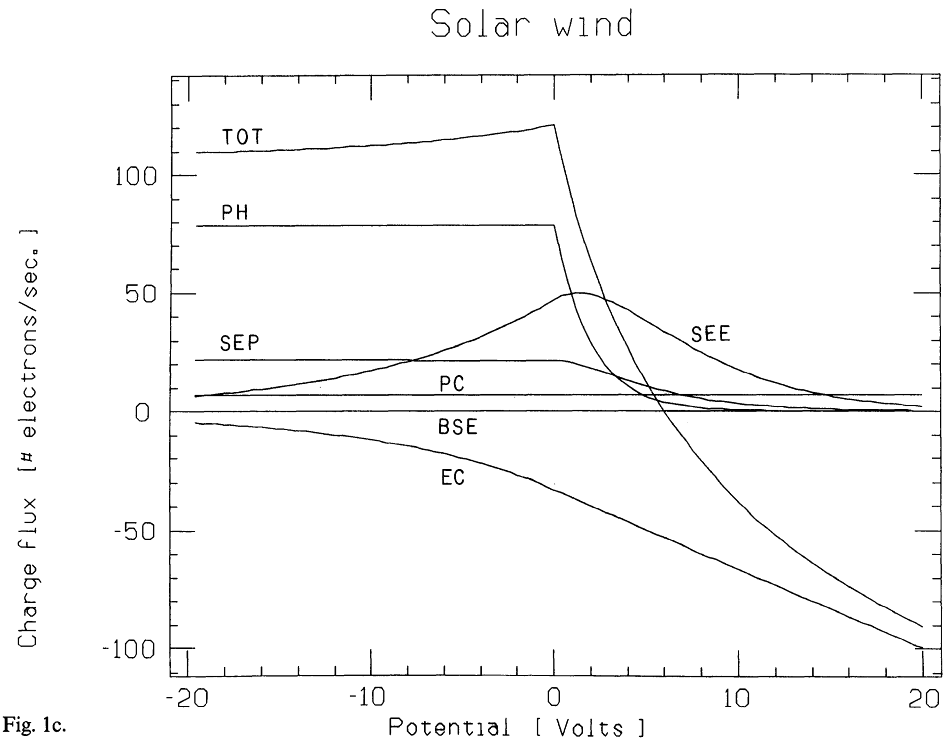


Fig. 1c.

Fig. 1. The current-potential curves the three plasma regimes, for a grain of radius 1 μm. (a) Within the plasma sphere at a geocentric distance of $6R_E$. (b) In between the plasmapause and the magnetopause, at 18 hr local time. (c) In the solar wind. In each case, *EC*, *PC*, and *PH* denotes the electron collection, the proton collection and photo-emission current, respectively, while *SEE*, *SEP*, and *BSE* denotes the secondary electron currents due to electron impact, and ion impact, and the backscattered electron currents. Also the currents are measured in units electron charge per sec, with + sign indicating electrons leaving from and − sign indicating electrons arriving at the grain.

In the magnetosphere, outside the plasmapause, the polasma has low densities and high energies. The addition or subtraction of energy in moving through the grain potential is negligible in comparison with the characteristic particle energies. All currents are almost constant, but secondary electron currents due to electron and proton impact, as well as the photoelectron current rapidly decrease to zero at positive grain potentials due to the fact that the associated emitted electron distribution are soft.

In the solar wind the plasma has both low densities and relatively low temperatures. The major charging currents are due to electron collection, photoemission and secondary electron emission by electron impact.

If we keep a grain long enough in a given plasma and radiative environment, it will achieve an equilibrium potential (and associated charge), when the net current to it is zero. The time history towards this equilibrium potential is shown in Figure 2. Figures

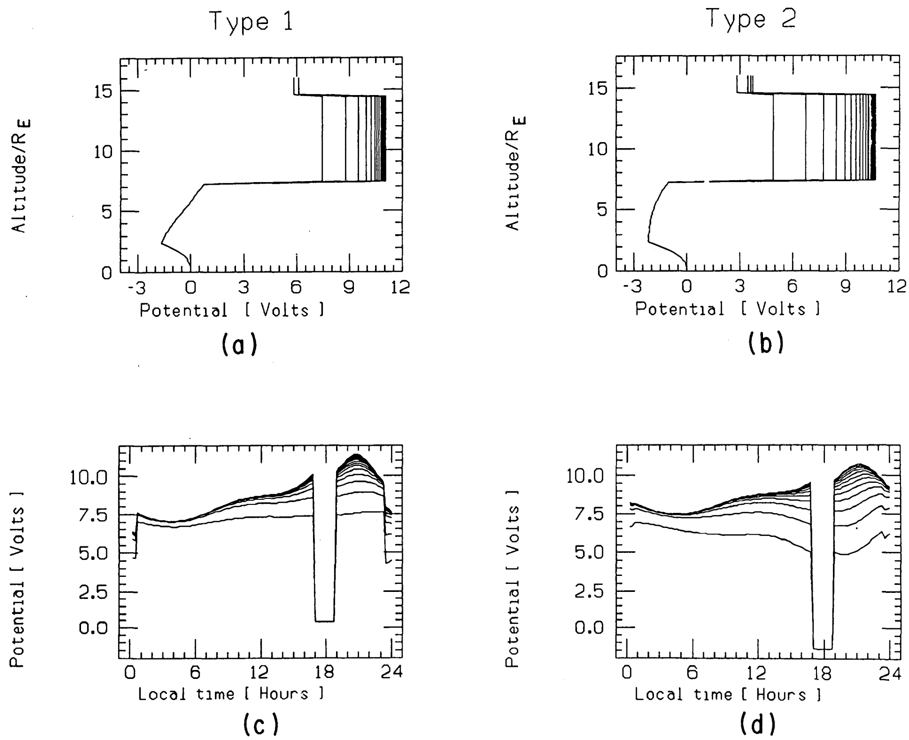


Fig. 2. The time history of the evolution of the potential of a grain of radius 0.1 μm to its equilibrium value: panels (a) and (b) show the variation with altitude at 18 hr local time for grains of type 1 (conductors) and for grains of type 2 (dielectrics), respectively, while panels (c) and (d) show the variation with local time, at geosynchronous orbit for type 1 and type 2 grains, respectively. The individual curves correspond to times 10 min apart.

2(a–b) show this for the two types of grains as a function of altitude, at 18 hours, local time. The individual curves correspond to times 10 min apart. The grain has a small negative potential at low altitudes, decreasing with altitude until the appearance of the second component of the plasma in the plasmasphere. It then begins to increase with increasing altitude, but still remains negative throughout the plasmasphere. Also, within the plasmasphere, due to the high plasma densities, the grain attains its equilibrium potential rapidly (within minutes).

At the altitude of $7.6R_E$ we reach the magnetopause, where it takes almost two hours to reach the equilibrium potential for type 2 grains and a somewhat shorter time for type 1 grains. The solar wind is reached at an altitude of $14.5R_E$ where it takes less than 20 min for type 1 and less than 40 min for type 2 grains to achieve equilibrium potential.

On Figures 2(c–d) are plotted the equilibrium potentials vs local time at geosynchronous orbit. For both type 1 and type 2 grains, the sudden decrease at around 18 hours local time is due to the fact that the geosynchronous orbit cuts into the plasmasphere around that direction. For type 1 grains the photoelectron current is about 10 times larger than that for type 2 grains which explains the jump in the equilibrium potential around midnight when the grains are in the Earth's shadow. (In the present study we used only the geometrical shadow of the Earth, while the optical depth of the atmosphere was neglected.)

It should also be mentioned that we have carefully checked for the possibility of multiple equilibrium potentials as suggested, for example by Meyer-Vemet (1982). When average geomagnetic conditions are used, as we have done, they do not exist. The possibility that they do exist at extreme geomagnetic conditions cannot be ruled out.

As the grains move along their trajectories, they may not spend enough time in a given plasma environment to achieve the local equilibrium potential. It is necessary to integrate Equation (1) simultaneously with Equation (2) to find the actual potential reached at any given point on the orbit. To illustrate this, we consider a grain moving with Kepler speed at the geosynchronous orbit, where the effects of light pressure and the Lorentz force has been turned off in order to have a periodic orbit. The light curve in Figure 3 shows the variation of equilibrium potential, discussed earlier, with local time. The dark curve shows the variation of the actual grain potential with local time once it has achieved a 'steady' periodic potential. It is seen that as the grain dips into the plasmasphere around 18 hours, it achieves its equilibrium potential almost instantly. This is due to the large plasma density there. However, as it emerges from the plasma

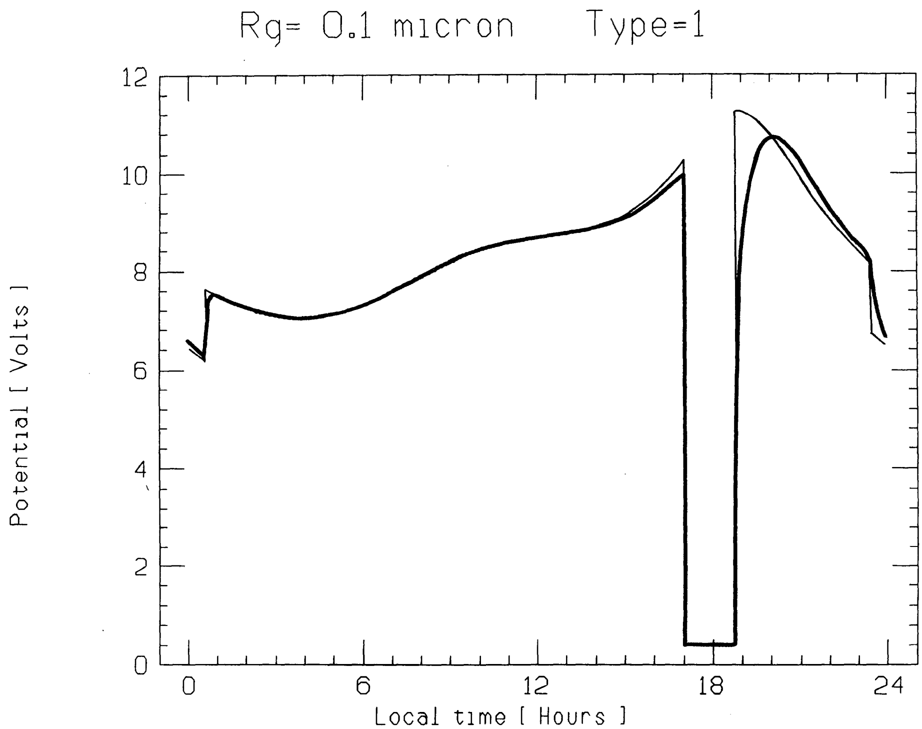


Fig. 3. Variation with local time of the equilibrium potential (light curve) and the actual potential achieved (dark curve) by a type 1 grain of radius 0.1 μm moving in a circular orbit at geosynchronous orbit.

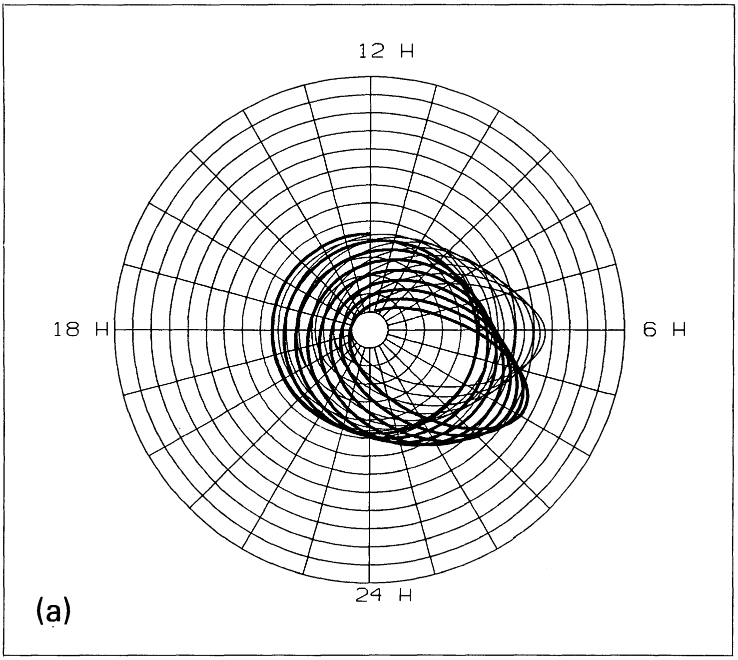
sphere into the magnetosphere (where the plasma density is much smaller) it takes time to approach the local equilibrium potential. Indeed it does so only after it has also entered and re-emerged from the Earth's shadow around 24 hours.

6. Orbital Evolution

Figure 4(a) shows the orbit of a grain of type 2 (dielectric) of radius 0.1 μm injected into the magnetosphere inside the geosynchronous orbit ($R = 0.8R_{gs}$) at noon. The lighter curve corresponds to the case when the Lorentz force is neglected (i.e., $Q = 0$), while the darker curve corresponds to the case when the Lorentz force is included ($Q \neq 0$). The dominant effect of radiation pressure is clear in both cases. The large-amplitude oscillation of ghe grain's orbital eccentricity caused by solar radiation pressure for the case $Q = 0$ is well known to be the dominant effect in the evolution of the orbit (Shapiro, 1963; Peale, 1966). This is also seen to be the main effect in the dynamics of the charged grains. The main difference is in the orientation of the orbits. In the case when $Q = 0$, the peregee and apogee are along the E–W meridian. The evolution continues until a maximum eccentricity is reached after which the process is reversed and repeated, unless the grain hits the Earth's atmosphere and is lost. In the case when $Q \neq 0$, the apogee and peregee are not along this E–W meridian but displaced from it by about 18°. Also, both grains ($Q = 0$ and $Q \neq 0$) hit the Earth in about the same time period; ($T(Q = 0) = 6.01$ days, $T(Q \neq 0) = 5.91$ days), although they do so at different points. The uncharged grain hits the Earth at the evening (18 hr) meridian, whereas the charged hits the Earth around the noon (12 hr) meridian.

Figure 4(b) shows the orbit of a grain of type 2 of radius 0.1 μm, which is injected into the magnetosphere outside the geosynchronous orbit ($R = 1.2R_{gs}$) at noon. As before, the lighter curve corresponds to the case ($Q = 0$) while the darker one corresponds to the case ($Q \neq 0$). The situation is qualitatively similar to the previous case, but due to the larger Lorentz forces experienced by the charged grain, its orbit differs more from the orbit of the charged grain than in the previous case. Note that in this case the apogee and peregee lie on a line displaced from the E–W meridian by about 25°. Also in this case, while the time taken by the uncharged grain to hit the Earth is about 6.08 days (which is in fact slightly larger than that for the unchanged grain launched from $R = 0.8R_{gs}$), the time taken by the charged grain is significantly smaller (4.54 days). Clearly, while the residence times of the grains that are charged are smaller than their uncharged counterparts, the discrepancy in the residence times of charged and uncharged grains (of equal physical and optical properties) launched from the same point will increase as its geocentric distance increases.

Finally in Figure 5 are shown the orbits of type 1 (conducting) grains of radius 0.1 μm injected into the magnetosphere at noon from $R = 0.8R_{gs}$ and from $R = 1.2R_{gs}$. As before, the lighter curves correspond to the uncharged grains, while the darker curves correspond to the charged grains. The radiation pressure effect on these grains are much larger than in the case of the type 1 (dielectric) grains. Consequently, the orbital evolution due to this effect is much larger, and the difference between the orbits of the



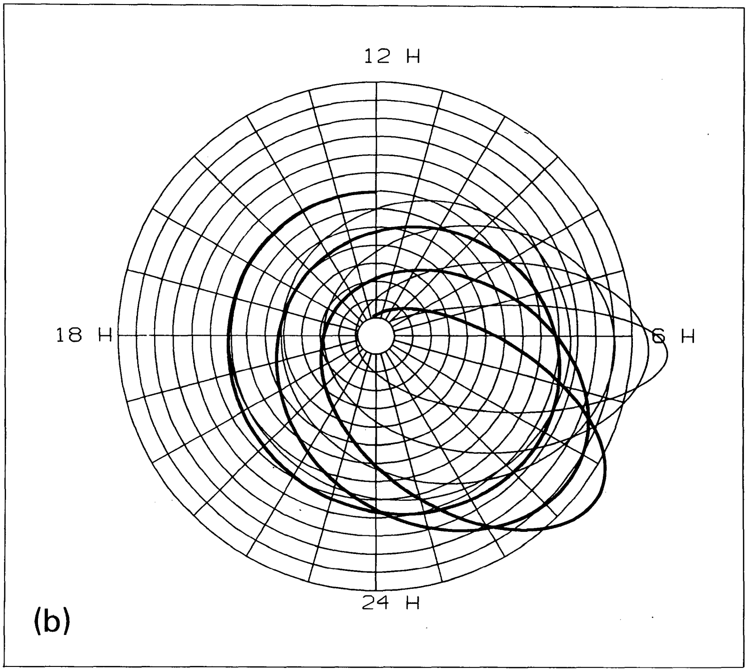


Fig. 4. Orbits of type 2 grains or radius 0.1 μm injected into the magnetosphere at 12 hr local time, (a) inside the synchronous orbit (at $R = 0.8R_s$) and (b) outside the synchronous orbit (at $R = 1.2R_s$). In each case, the light curve denotes an uncharged grain ($Q = 0$), while the dark curve denotes a charge grain.

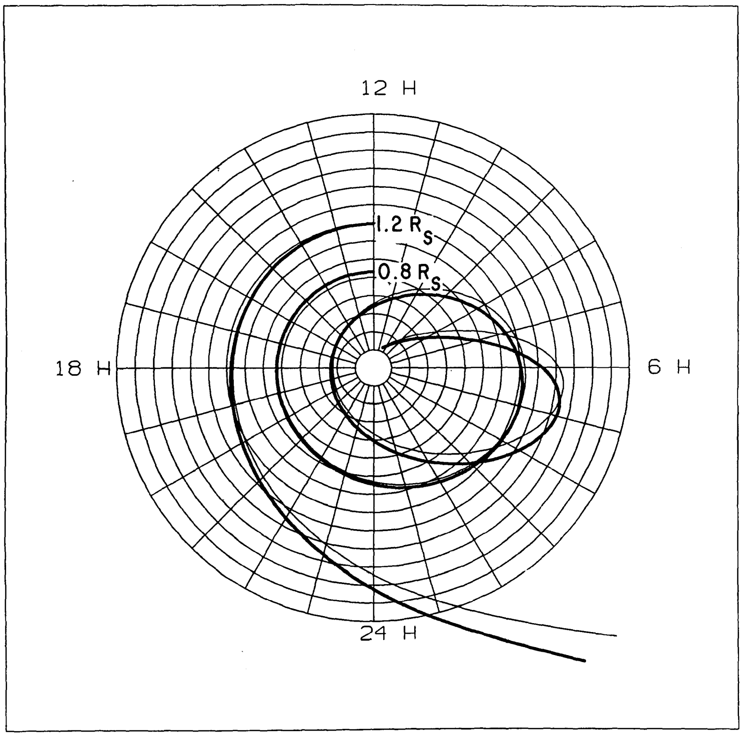


Fig. 5. Same as Figures 4(a–b), but foa a grain of type 1 with radius 0.1 μm. The orbits corresponding to injection both at $0.8R_s$ and at $1.2R_s$ are shown in the figure.

charged and uncharged grains is smaller than in the case of the dielectric grains. Thus, the discrepancy of the residence times between the charged and uncharged grains is also small. Both grains injected at $R = 0.8R_{gs}$ crash to Earth at about the same point, after about the same residence time ($T(Q = 0) = 1.88$ days, ($T(Q \neq 0) = 1.86$ days). Both grains injected at $R = 1.2R_{gs}$ leave the magnetosphere (at $R = 20E_E$) in the very first orbit ($T(Q = 0) = 1.12$ days, ($T(Q \neq 0) = 1.14$ days).

7. Discussion

The central aim of our ongoing study is to assess the role of electromagnetic (Lorentz) forces that are experienced by fine dust particles that are injected into the Earth's magnetosphere by natural (micro-meteoroid flux) and by artificial (solid rocket propellant burns) processes. In the present paper we have confined our attention to the latter case. In order to make some progress in this initial study, we have used a rather simple model of the particles and fields environment of the Earth, and confined our attention to grain orbits in the equatorial plane. We have considered all the important charging currents on to the grains and shown that the electric charge acquired by the grains is rather modest. It varies typically from a few volts (negative) within the plasma pause to about + 10 V in the region between the plasmapause and the magnetopause, to a few volts (positive) in the solar wind. Under these circumstances, the solar radiation pressure-driven oscillation of the orbital eccentricity is still the dominant dynamical feature of the orbital dynamics. However, for the smallest grains ($r_g \approx 0.1$ μm), the electromagnetic (Lorentz) forces associated with the grain motion through the magnetized plasma, significantly changes the nature of the orbits, particularly if the grains are dielectric. The magnetospheric residence times of the grains are also significantly decreased by the charging, the effect being more pronounced for grains injected at larger geocentric distances.

While we recognize that the relevant plasma distributions are highly dependent on the geomagnetic index A_p we have kept A_p constant at its average value of 120 in this calculation. Since increased A_p during disturbed solar wind conditions leads to higher magnetospheric temperatures (Garrett and DeForest, 1979), we expect that the electrodynamic effects on the grains during such times would be larger, leading to even shorter residence times than those corresponding to $A_p = 120$. In a subsequent paper we propose to investigate that effect.

Besides the artificial injection of grains in the vicinity of the synchronous orbit discussed above, there is also the natural injection of micrometeoroids into the magnetosphere. Since these are injected with a relative velocity equal to about the escape velocity and are already charged as they penetrate the magnetopause, we expect that the electrodynamic effects they experience would be quite large (perhaps even comparable to the effects of radiation pressure). Consequently, they may be lost (either by impact with the Earth) or by ejection from the magnetosphere, even faster. We propose to study the question too in the subsequent paper.

There we will study the evolving distribution functions as well as the differential residence times (with grain radius and location of injection) for both the artificial and natural populations. We will also consider the steady-state distributions assuming continuous injection at the appropriate rates.

Another complication that has to be taken into account is the finite inclination of the spin axis of the Earth to the magnetic axis, which necessitates the consideration of 3-D distributions. In this case grains injected in the equatorial plane will be quickly dispersed normal to it by the electrodynamic forces they experience in that direction.

While we prognose a detailed quantitative study of all these in the subsequent paper, it is already apparent qualitatively from the present study that the electrodynamic effects on charged grains in the Earth's magnetosphere appears to conspire with the solar radiation pressure effects to eliminate fine dust from it.

Acknowledgements

One of us (M.H.) acknowledges support from DOE under agreement No. DE-FCOS-85ER25000. The simulations were done on the Cyber 205 at the Supercomputer Computational Research Institute, Florida State University, which donated part of the computation time.

References

Draine, B. T. and Salpeter, E. E.: 1979, *Astrophys. J.* **231**, 77.
Garrett, H. B. and DeForest, S. E.: 1979, *J. Geophys. Res.* **84**, 2083.
Hill, J. R. and Mendis, D. A.: 1980, *The Moon and Planets* **23**, 53.
Hill, J. R. and Northrop, T. G.: 1983, *J. Geophys. Res.* **88**, 1.
Hill, J. R. and Whipple, E. C.: 1985, *J. Spacecraft Rockets* **22**, 245.
Katz, I., Parkes, D. E., Mandell, M. J., Harvey, J. M., Brownell, D. H., Wang, S. S., and Rotenberg, M.: 1977, NASA CR-135256.
Lafromboise, J. G. and Parker, L. W.: 1973, *Phys. Fluids* **16**, 629.
Lyons, L. R. and Wiliams, D. J.: 1984, *Quantitative Aspects of Magnetospheric Physics*, D. Reidel Publ. Co., Dordrecht, Holland, p. 79.
Mendis, D. A.: 1981, in F. D. Kahn (ed.), *Investigating the Universe*, D. Reidel Publ. Co., Dordrecht, Holland, p. 353.
Meyer-Vermet, N.: 1982, *Astron. Astrophys.* **105**, 98.
Mueller, A. C. and Kessler, D. J.: 1985, *Adv. Space Res.* **5** (No. 2).
Northrop, T. G. and Hill, J. R.: 1983, *J. Geophys. Res.* **88**, 1.
Peale, S. J.: 1966, *J. Geophys. Res.* **71**, 911.
Prokopenko, S. M. L. and Laframboise, J. G.: 1980, *J. Geophys. Res.* **85**, 4125.
Shapiro, I. I.: 1963, in M. Roy (ed.), *Dynamics of Satellites*, Springer-Verlag, Berlin, p. 257.
Sternglass, E. J.: 1954, *Sci. Pap. 1772*, Westinghouse Research Laboratory, Pittsburg, PA.
Whipple, E. C.: 1981, *Rep. Progress Phys.* **44**, 1197.

GENERALIZED ADIABATIC THEORY APPLIED TO THE MAGNETOTAIL CURRENT SHEET*

SHELLEY H. ROGERS and ELDEN C. WHIPPLE

Center for Astrophysics and Space Sciences, University of California at San Diego, La Jolla, CA, U.S.A.

(Received 30 June, 1987)

Abstract. We use the generalized first adiabatic invariant, an extension of the magnetic moment for regions of large field gradients, to treat particles in the magnetotail current sheet. The equations of motion can be expressed in terms of drift parameters which vary slowly and smoothly at the drift rate, not at the gyration rate. The analysis leads to boundaries in phase space which form a generalized loss cone and separate particles drifting into and out of the layer from particles trapped within the layer. These boundaries can be used in the moment integrals for densities and currents when the drifting particles differ in temperature, or in other properties, from the trapped population, as has been suggested by observations. We give examples of how different kinds of particle orbits contribute to the spatial profiles of density and current and thus to the field structure of the current sheet. We find that the parallel pressure of the drifting particles must exceed the transverse pressure for self-consistent solutions to exist, and based on this result, we give examples of fully self-consistent solutions using bi-Maxwellian ion and Maxwellian electron distributions. We give a proof, using generalized adiabatic theory, of Cowley's (1978a) theorem that particles trapped in the current layer experience zero net drift.

1. Introduction

Current sheets in the magnetosphere, such as the magnetopause and the tail current sheet, are critical regions in that they are the most important sites of particle energization. They are also frequently boundaries between plasmas with different properties. In these current sheets the magnetic field is small; the particle gyradius can be, therefore, large compared to the scale of the region, and the electric field can consequently play a dominant role in the particle motion. This is especially true at neutral points or neutral lines where the magnetic field vanishes. Attention is being focussed on these regions as possible sites for transformation of magnetic field energy into particle kinetic energy.

A great deal of theoretical effort has been expended in studying plasma behaviour in these current sheets. However, these studies have been hampered by the difficulty that fluid descriptions often neglect important details of the plasma velocity distributions, and in addition fluid descriptions can break down in the vicinity of current sheets where effects of particle trajectories may be important and where the scale for gradients can be on the order of or even smaller than the particle gyroradius. Kinetic descriptions based on the Vlasov equation have been used to treat certain aspects of these problems, but a realistic, self-consistent treatment that takes into account different velocity distributions which may be trapped within and traversing current sheets has been beyond present techniques.

Alfvén (1939) was the first to point out that the magnetic moment of charged particles

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 231–256.

was an adiabatic invariant which could be used to simplify the description of particle motion. We have recently developed a generalized adiabatic treatment of plasma behaviour that can be used in regions of strong field gradients such as occur in current sheets (Whipple *et al.*, 1986; see this for references to earlier work). This development is based on the fact that a generalized first invariant for particle motion – a generalization of the magnetic moment – can still be defined in such regions. Generalized drift velocities and other slowly varying parameters can also be obtained, and the equations of motion can be cast in a form involving these slowly varying drift parameters. In this paper we show how this procedure can be used to describe particle behaviour and current structure in the tail sheet where there is a normal component of the magnetic field across the sheet, and where there are different velocity distributions for the trapped population as compared to the plasma population drifting across the layer.

2. Previous Work

The first treatments of current sheet structure based on the Vlasov equation were made by Grad (1961) and Harris (1962). Grad discussed the one-dimensional problem of a current sheet separating a plasma and a magnetic field and obtained a self-consistent solution. Harris treated a current sheet separating two regions of oppositely directed magnetic fields. They showed how solutions could be obtained by using the Poisson and curl **B** equations after expressing the plasma distribution function in terms of constants of the motion. However, neither treatment included a normal component of **B** across the layer and they both had a single plasma population with the density vanishing at large distances. Harris's solution, with a constant normal component of the magnetic field added to it, has been widely used as a model for the geomagnetic tail.

A number of papers have treated the kinetic structure of the magnetopause where the current sheet separates two different magnetized plasmas (e.g., Sestero, 1964, 1966; Alpers, 1969, 1971; Lemaire and Burlaga, 1976; Lee and Kan, 1979, 1980; Roth, 1978, 1979, 1980). A problem with these solutions is that they are not unique. The choice of distribution functions was made arbitrarily without regard for the question of particle accessibility to the current sheet from source regions. We have shown (Whipple *et al.*, 1984) that consideration of particle accessibility removes this nonuniqueness, and that the restrictions of a plane current sheet with no normal magnetic field component must be relaxed in order to allow particle access to the current sheet via drifts from source regions.

For example, when a normal component of **B** is added to the Harris model for a current sheet, then the particles in a boundary layer can approach the sheet via field-aligned drifts. However, the tangential canonical momenta which were constants of motion in the Harris model are now no longer constants, but vary slowly as the particles drift, thus rendering the Harris model invalid. In addition, there are particles trapped in drifting orbits in the current sheet which must be taken into account. None of the previous treatments of current sheets have been able to properly distinguish between the drifting and trapped plasma components which in general will have different densities, temperatures, etc. An accurate distinction based on the particle orbits is essential in order to obtain the contributions of each of these populations to the particle pressure, density, and current as a function of position across the layer.

Several authors have treated the tail current sheet from Vlasov theory with different approximations (see Dungey, 1972, 1975; and Schindler, 1975, 1979, for reviews). Cowley (1971, 1973) and Eastwood (1972) developed self-consistent models with a cold plasma, but they neglected oscillations of particles in trapped orbits. Thus, they could assume adiabatic behaviour up to the region of field reversal, but they emphasized that adiabatic behaviour must break down in the current layer itself. Eastwood (1974, 1975) extended this work to include a Boltzmann relation for electrons and ion beams at discrete speeds and various pitch angles, and obtained solutions with an iteration procedure. Toichi (1972) and Kan (1973) treated the two-dimensional Vlasov equation (x and z in GSM coordinates) and obtained results which reduce to the Harris solution when the x-dependence is removed, but neither author made a distinction between trapped particles and particles traversing the current sheet. Bird and Beard (1972a, b) used adiabatic theory and the resulting drifts together with the curl $\mathbf{B}$ equation to get at the structure of the magnetotail.

Schindler and his collaborators have carried out several studies of the self-consistent structure of the tail and its stability based on Vlasov theory (Schindler, 1972, 1975, 1976, 1979; Soop and Schindler, 1973; Birn *et al.*, 1975, 1977). Schindler (1975) emphasized that "A satisfactory theoretical description ... is outside the reach of present techniques. The main difficulty lies in the fact that the majority of the tail phenomena required a self-consistent theory... The MHD-approximation ... is not a valid approximation for processes in which details of the particle orbits are essential." They obtained solutions with velocity distributions assumed to have a form such that the particle pressure was isotropic (in the v_x, v_z plane), and as a result particle pressure was constant along a given magnetic field line. Consequently, any trapped population must have the same plasma properties as the field-aligned population – i.e., no loss cone structure is allowed. This work has been extended to three dimensions by Birn *et al.* (1977) and by Birn (1979). Since the boundary conditions were not well known, the free parameters in the model were adjusted to observations.

There have been several studies where magnetic field models have been assumed and particle orbits have been calculated. Speiser (1965a, b, 1967) calculated ion trajectories in a current sheet which included simple magnetic and electric fields and found that particles could be accelerated to large energies in the sheet and then ejected in thin beams. Cowley (1978a) proved that in a one-dimensional current sheet with a normal component of $\mathbf{B}$, there are trapped particles in drifting orbits which exhibit zero net drift (as suggested earlier by Stern and Palmadesso, 1975). Lyons and Speiser (1982) used a simple model of the tail magnetic field together with a constant y-component electric field to calculate particle energization. The distribution of accelerated ions with an assumed mantle-like distribution as the initial condition was found to be very similar to observed plasma sheet distributions. Bird (1975), Propp and Beard (1984), and Beard and Cowley (1985) calculated ion orbits in a two-dimensional model of the tail. In the latter two papers the authors found that adiabatic theory predictions were generally valid in spite of 'violation of the adiabatic limit'. However, Tsyganenko (1982), Popielawska *et al.* (1985), and most recently Basu and Rowlands (1986) have studied the effects of a current sheet in changing the magnetic moment of particles and have defined a 'scattering matrix' which can be used to describe these changes. Chen and Palmadesso (1986) used the Harris magnetic model with a constant cross-tail component and followed particle orbits. They classified three types of orbits: bounded integrable orbits, unbounded stochastic orbits, and unbounded transient orbits, and emphasized the chaotic nature of the stochastic orbits.

There is evidence that as one goes towards the central plasma sheet, the central part is hottest with substantial cooling both above and below the center (Akasofu *et al.*, 1973). Eastman *et al.* (1984, 1985) state that the central plasma sheet has hotter, more isotropic distributions than the plasma sheet boundary layer. This suggests that there may be a trapped population in the plasma sheet with different properties (i.e., a higher temperature) than in the surrounding boundary layer plasma. Francfort and Pellat (1976) pointed out that including a trapped population in the current sheet can make it possible to get a self-consistent (one-dimensional) description. They point out the utility of the generalized first invariant although their discussion is in the context of a geometry where the field reversal occurs at a discontinuity. Cowley and Pellat (1979) showed how trapped particles could be included in an adiabatic formulation of the one-dimensional current sheet.

3. Generalized Adiabatic Theory

As mentioned in the Introduction, we have extended our work on generalized adiabatic theory by applying it to the kind of one-dimensional current sheet that is found in the geomagnetic tail, i.e., one that gives rise to a magnetic field parallel to the sheet and basically antisymmetric about the sheet's center. We describe here how generalized adiabatic theory allows us to distinguish between the free and trapped populations that result when a component of the magnetic field normal to the sheet is present in addition to the characteristic parallel field. This theory offers a flexible approach to obtaining a fully self-consistent description of the tail current, since if we can designate each point in phase space as belonging to either a free or a trapped orbit, we can then attempt a self-consistent solution by giving free and trapped populations different distribution functions in the integrals for density and current. Such a model would thus include two realistic features of the magnetotail not treated in detail by previous models: the geometrical effect of the normal field and the ability to separate two physically distinct populations in phase space.

We use standard GSM coordinates for our model current sheet, with z pointing northward across the sheet, x toward the Sun, and y from dawn to dusk. Figure 1 depicts magnetic field lines representative of the magnetotail. We assume that B_x varies only with z, and that B_z, the normal component, is constant. Figure 2 illustrates two

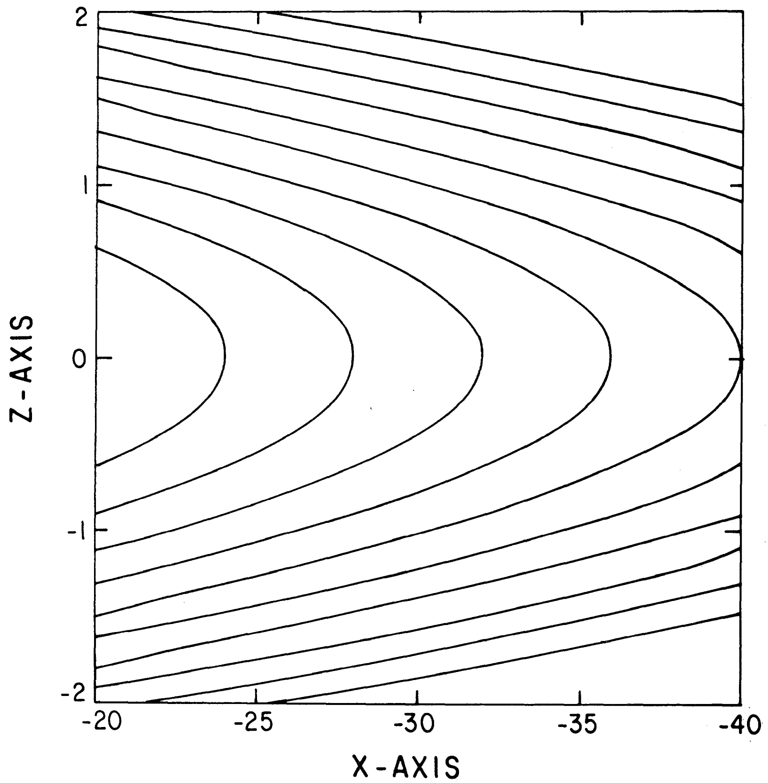


Fig. 1. Typical magnetic field line configuration in the tail. Note the different scales for the x- and z-axes. The units are Earth-radii.

ways B_x might vary with z; though B_x vanishes at the sheet's center, the total field remains finite because of B_z. In addition to the magnetic field, there will in general be an electric field E_z; any tangential component of E, which must be constant and across the layer in a steady state, can be removed by the de Hoffman and Teller (1950) transformation.

In order to apply generalized adiabatic theory we must treat B_z as a small perturbation, henceforth denoted by β. We define a vector potential $A_y(z)$ such that $B_x(z) = -\mathrm{d}A_y(z)/\mathrm{d}z$ and an electric potential $\phi(z)$ such that $E_z = -\mathrm{d}\phi/\mathrm{d}z$. We assume that all fields and potentials vary only with z and are either symmetric (A_y, ϕ, β) or antisymmetric (E_z, B_x) about $z = 0$. The equations of motion then are

$$\begin{aligned} \mathrm{d}v_x/\mathrm{d}t &= (q/m)(v_y\beta)\,, \\ \mathrm{d}v_y/\mathrm{d}t &= (q/m)(v_zB_x - v_x\beta)\,, \\ \mathrm{d}v_z/\mathrm{d}t &= (q/m)(E_z - v_yB_x)\,. \end{aligned} \tag{1}$$

The Kruskal (1962) procedure, on which generalized adiabatic theory is based, uses constants of the motion of the zeroth-order equations (i.e., Equation (1) with $\beta = 0$) as

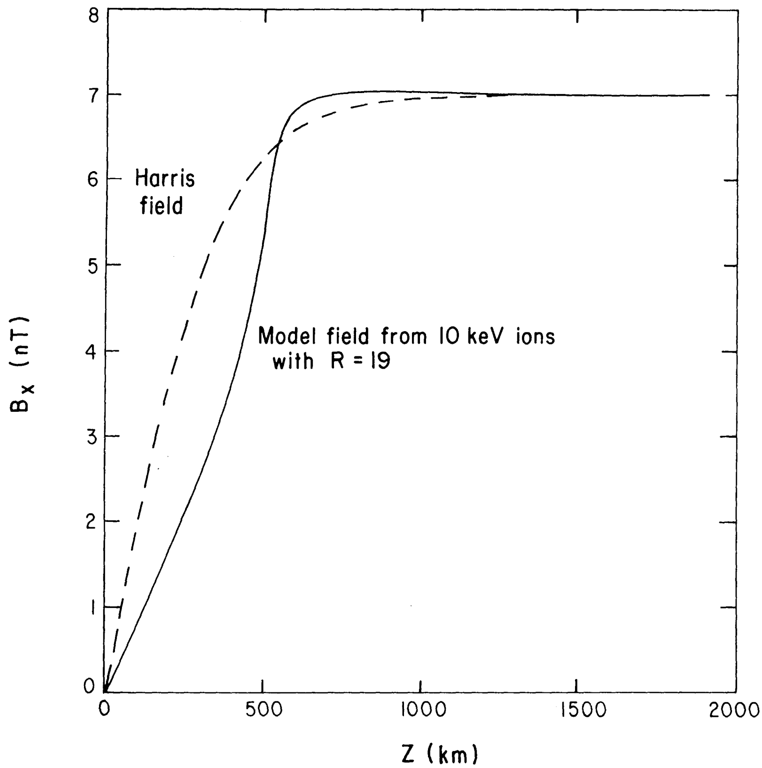


Fig. 2. Possible current-sheet fields B_x, as a function of distance from the center of the sheet. The dashed curve shows the Harris field. The solid curve shows the model field generated by an ion delta function distribution in E and J (see Section 5).

starting points for constructing the 'Z-variables' that are useful when β is finite. In our case these constants are the generalized transverse energy H, the canonical y-momentum P, and the x-component of velocity V_x, where

$$
\begin{aligned}
H &= (m/2)(v_y^2 + v_z^2) + q\phi, \\
P &= mv_y + qA_y, \\
V_x &= v_x .
\end{aligned}
\tag{2}
$$

With the perturbation included, these quantities are no longer constant but vary with time derivatives of order β – i.e.,

$$
\begin{aligned}
\mathrm{d}H/\mathrm{d}t &= -\beta q v_x v_y, \\
\mathrm{d}P/\mathrm{d}t &= -\beta q v_x, \\
\mathrm{d}V_x/\mathrm{d}t &= \beta q v_y/m .
\end{aligned}
\tag{3}
$$

The motion of the particles consists basically of a gyration about the magnetic field superposed on a drift. When a particle is far away from the current sheet, its drift is primarily along a magnetic field line, but as the particle approaches the sheet it develops a drift in the y-direction as well. The gyration aspect of a particle's motion can be analyzed by considering $\psi(P, z)$, an effective potential for the unperturbed motion in the z-direction given by

$$\psi(z) = (P - qA(z))^2/(2m) + q\phi(z) . \tag{4}$$

Since $A_y(z)$ and $\phi(z)$ are symmetric functions about $z = 0$, $\psi(z)$ is symmetric also. A typical plot of $\psi(P, z)$ vs z for a particle in the current sheet is shown in Figure 3 for $(P/q) < 0$. There are two equal minima on opposite sides of $z = 0$, and the height of the central maximum, which depends on P and $\phi(0)$, is given by $H_c = P^2/(2m) + q\phi(0)$. If a particle's H-value is less than H_c, it gyrates in one of the side-wells, while if $H > H_c$ it oscillates back and forth across the central plane. For a particle far from the current sheet, where the electric potential and the magnetic field gradient are small, the side-wells are well-separated and the particle gyrates in one of them in an approximately circular orbit. As the particle drifts toward the sheet, H, P, and V_x vary on both gyration and drift scales and the shape of the well gradually changes, so that H finally becomes

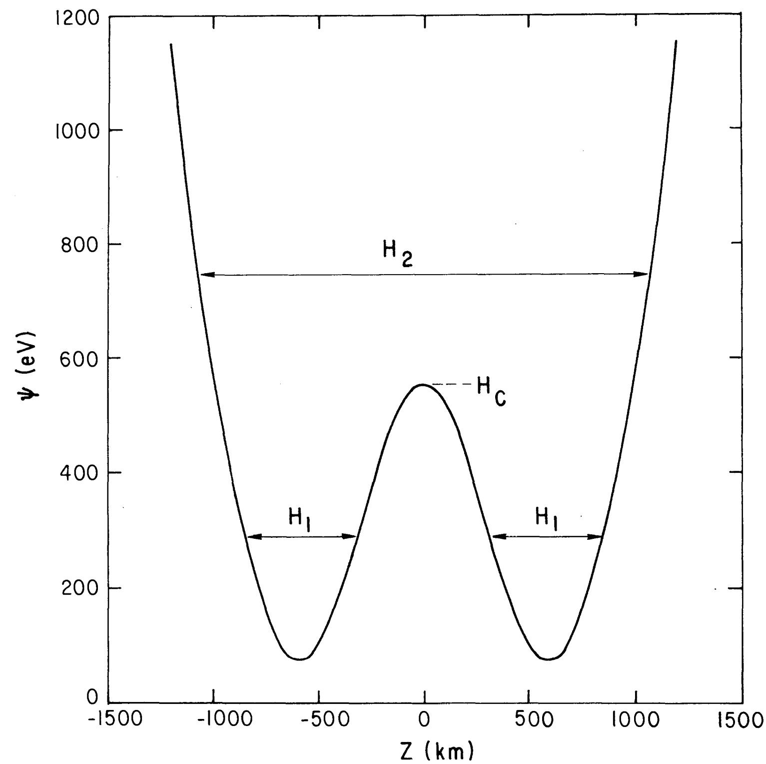


Fig. 3. An example of the effective potential for particle motion in the z-direction. As the particle drifts, its value of $\mathscr{H}$ changes and the shape of the well also changes.

greater than H_c and the particle oscillates about $z = 0$. The particle's H-value may then pass below and above H_c a number of times before the particle finally escapes from the sheet. A trapped particle undergoes similar motion but does not have enough energy to escape in the absence of collisions. For $(P/q) > 0$, the effective potential ψ in our example has only one minimum, which occurs at $z = 0$.

We can see this kind of behaviour in individual orbits obtained by integrating the equations of motion of a particle in the modified Harris field, $B_x = B_0 \tanh(z/L), B_z = \beta$. Figure 4(a) shows one such orbit projected into the x–z plane, while Figure 4(b) shows the same orbit projected into the $y - z$ plane. As seen in the $x - z$ plane, the particle spirals into the current region on a field line (gyrating in a side-well) until it reaches the sheet's center, at which point it begins oscillating about $z = 0$ (i.e., gyrating in the central

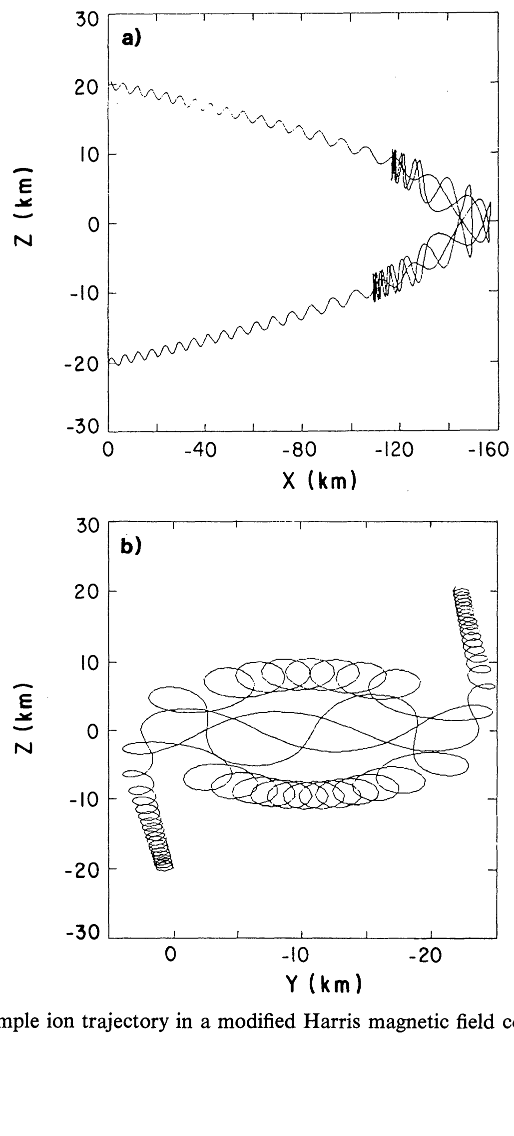


Fig. 4. A sample ion trajectory in a modified Harris magnetic field configuration.

well). It then moves back into a side-well, mirrors, re-enters the central well, moves to the other side-well, mirrors again, and finally escapes from the sheet, displaying at large distances roughly the same value of the magnetic moment it started with. Figure 4(b) is significant in that it shows typical particle motion in the y-direction (the direction of electric current) and illustrates how particles describing various orbits can reinforce or counter the prevailing sheet current. As the particle spirals in on a field line, it gradient-B drifts in y contributing a current opposite to the Harris current. When it enters the central well, its direction dramatically changes and it 'snakes' along the y-axis as it oscillates in z, contributing a current that reinforces the sheet current. When it moves back into a sidewell, it again experiences strong gradient-B drift that opposes the sheet current. It is this interplay of drift currents that determines the net current profile that particles on a given orbit contribute to the overall current (see Section 5).

The quantities H, P, and V_x, which are called 'Y-variables' in Kruskal's terminology, can be improved by the Kruskal procedure to obtain the 'nice Z-variables', which have the property that they do not oscillate at the gyrofrequency as do the Y-variables. We denote these improved quantities by $(\mathscr{H}, \mathscr{P}, \mathscr{V}_x)$ and by following the procedure described in Whipple *et al.* (1986) find them to be

$$\begin{aligned} \mathscr{H} &= H - \beta q \mathscr{V}_x \int_0^\nu \frac{(\bar{v}_y - v_y)\,\mathrm{d}\nu}{\mathrm{d}\nu/\mathrm{d}t}\,, \\ \mathscr{P} &= P\,, \\ \mathscr{V}_x &= V_x + (\beta q/m) \int_0^\nu \frac{(\bar{v}_y - v_y)\,\mathrm{d}\nu}{\mathrm{d}\nu/\mathrm{d}t}\,. \end{aligned} \tag{5}$$

The time-derivatives of the Z-variables are non-oscillatory and of order beta: i.e.,

$$\begin{aligned} \mathrm{d}\mathscr{H}/\mathrm{d}t &= -\beta q \mathscr{V}_x \bar{v}_y\,, \\ \mathrm{d}\mathscr{P}/\mathrm{d}t &= -\beta q \mathscr{V}_x\,, \\ \mathrm{d}\mathscr{V}_x/\mathrm{d}t &= \beta q \bar{v}_y/m\,; \end{aligned} \tag{6}$$

where $\bar{v}_y$ is the gyro-averaged y-component of velocity, and ν is an appropriately defined phase angle for the gyromotion. The advantage of these new quantities is that they are strictly drift parameters, varying slowly and smoothly as the particle drifts in the z-direction. The particle's motion across the layer can be described in terms of these drift parameters and the generalized invariant, as we now explain.

4. The Generalized Invariant

To analyze the drift motion we must introduce the generalized invariant J, which through first order in β is constant throughout a particle's orbit, even in the current sheet region where B_x and E_z vary considerably over a gyro-radius. In our case J is a function of the Z-variables $\mathscr{H}$ and $\mathscr{P}$ and is given by the action integral (Whipple *et al.*, 1986)

$$J(\mathscr{H},\mathscr{P}) = m \oint v_z(\mathscr{H}, \mathscr{P})\, \mathrm{d}z\,. \tag{7}$$

The invariant J and the total energy E, an exact constant of the motion, allow us to follow the particle's drift path and in particular to assign to the particle the status of 'trapped' if its orbit is confined to the current sheet or 'free' if its orbit originates and ends at $|z| = \infty$. To illustrate the latter point, we use (7) to plot curves of constant J in the ($\mathscr{H}$, $\mathscr{P}$) plane (hereafter referred to as 'J curves') and note that as a particle drifts, its Z-variables $\mathscr{H}$ and $\mathscr{P}$ are constrained to move along one such curve (Figure 5). A further constraint is imposed on a particle's $\mathscr{H}$-value by its total energy, since $\mathscr{H}$ can never be greater than the total energy $E = \mathscr{H} + m\mathscr{V}_x^2/2 + o(\beta^2)$. We see that in the absence of collisions a particle with (J, E_1) will be trapped in the current sheet indefinitely, while a particle with (J, E_2) will drift into the current sheet from afar and eventually escape. A particle with (J, E_3) can be either trapped or free, depending on which $\mathscr{P}$ regime it finds itself in, and E_4 is the boundary energy between trapped and free orbits all characterized by the invariant J.

One key to whether a particle is free or trapped is in its allowed range of $\mathscr{P}$, since $\mathscr{P}$ indicates the location of the particle's (generalized) guiding center z_c via the expression

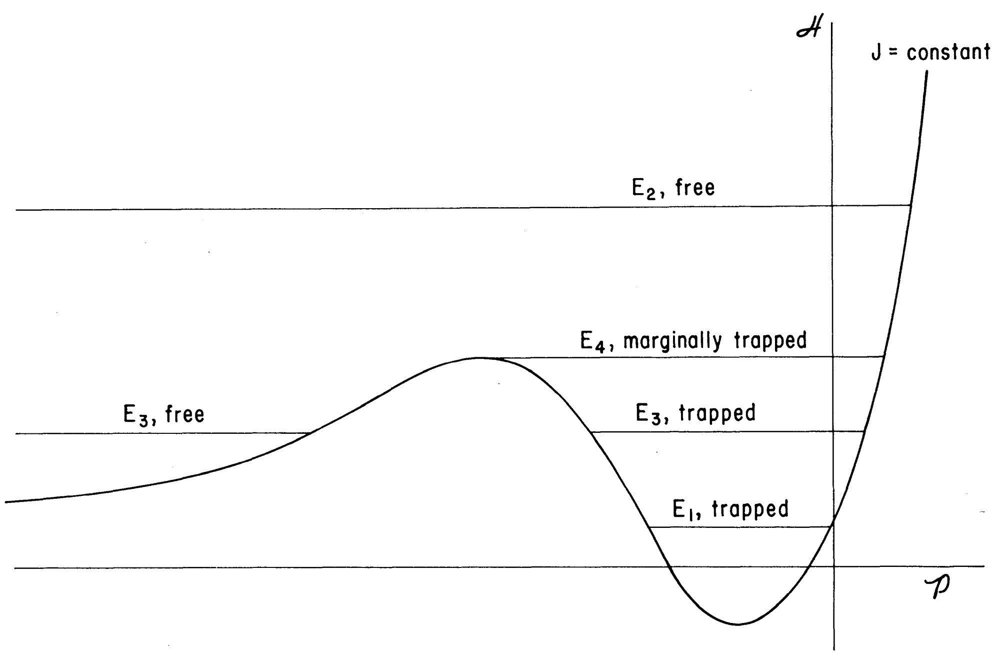


Fig. 5. A typical $J(\mathscr{H}, \mathscr{P})$ = constant curve in the ($\mathscr{H}$, $\mathscr{P}$)-plane. The particle's total energy E determines whether the particle is free (i.e., can escape from the current layer) or trapped in the current layer. The particle with a given (constant) E can be thought as moving back and forth on the $J(\mathscr{H}, \mathscr{P})$ curve and reflecting when it reaches the points where $\mathscr{H} = E$.

$$\left.\frac{\mathrm{d}\psi}{\mathrm{d}z}\right|_{z=z_c} = 0 \quad \text{or} \quad \mathscr{P} - qA_y(z_c) = \frac{mE_z(z_c)}{B_x(z_c)} \;. \tag{8}$$

If we set $A_y(0) = 0$, then $A_y(z) \leq 0$ for B profiles like those in Figure 2. Thus a large, negative $\mathscr{P}/q$ corresponds to a distance far from the current sheet where the particle executes true guiding-center motion, while (for $E_z/B_x \geq 0$) a smaller negative $\mathscr{P}/q$ or any positive $\mathscr{P}/q$ corresponds to motion through the sheet. For trapped particles, the $\mathscr{P}$-values at the intersections of E = constant with the J-curve indicate the extent to the particles travel from the sheet via Equation (8) and are turning points of the drift motion.

In terms of Z-variables the boundary between free and trapped orbits can be expressed as

$$E = (1/2)m\mathscr{V}_x^2 + \mathscr{H} \quad \begin{array}{ll} < \mathscr{H}_b & \text{for trapped orbits}\,, \\ > \mathscr{H}_b & \text{for free orbits}\,; \end{array} \tag{9}$$

where $\mathscr{H}_b$ (Figure 6) is essentially a 'barrier' value and for a given particle species depends only on J and the electromagnetic field configuration. We note an exception to (9) in that for some J-values a small fraction of the particles may drift in from far away with $E < \mathscr{H}_b$ and 'reflect' before penetrating the sheet (e.g., $E = E_3$ in Figure 5). Solving (9) for $\mathscr{V}_x$ when $E = \mathscr{H}_b$ we obtain

$$\mathscr{V}_{xb} = \pm\,[(2/m)(\mathscr{H}_b - \mathscr{H})]^{1/2}\,, \tag{10}$$

where $\mathscr{H}_b$ is a function of a particle's J. When we do the moment integrals over velocity space to get the density and current, we can transform from velocity variables to Z-variables, fix a pont in the $(\mathscr{H}, \mathscr{P})$ plane, and use (10) to differentiate between trapped and drifting orbits in the integral over $\mathscr{V}_x$. In other words, all particles with a given pair of values $(\mathscr{H}, \mathscr{P})$ have an invariant whose value is $J(\mathscr{H}, \mathscr{P})$, but some move in trapped orbits and some in free orbits depending on whether $|\mathscr{V}_x|$ is $<$ or $>$ $|\mathscr{V}_{xb}|$.

The boundary between trapped and free orbits in the $\mathscr{H} - \mathscr{P}$ plane can be determined by referring again to the J-curve (Figure 6(a)). For particles of fixed J, the part of the J-curve between $\mathscr{P}_a$ and $\mathscr{P}_b$ (region 1) corresponds to both trapped and free orbits, while the remainder (region 2) corresponds to free orbits only. In region 1, whether an orbit is free or trapped is determined by the particle's $\mathscr{V}_x$, as discussed above. We must translate the boundaries between these regions into limits for the $\mathscr{H}$ and $\mathscr{P}$-integrations: if we integrate over $\mathscr{P}$ first, for instance, we need to find the least and greatest $\mathscr{P}$-values attainable by trapped particles of a given $\mathscr{H}$. In other words, if we observe all particles with a certain $\mathscr{H}$, we will always find that for $\mathscr{P}/q >$ some finite $\mathscr{P}_b/q$, no trapped orbits can exist for any J. Similarly, we will find no trapped orbits for $\mathscr{P}/q < \mathscr{P}_a/q$, though this lower bound can be indefinitely negative depending on the shape of the J-curve.

In Figure 6(b) we illustrate how to obtain these $\mathscr{P}$ extrema for a particular fixed $\mathscr{H} > 0$. We start by constructing that J-curve whose barrier energy $\mathscr{H}_b$ equals our fixed $\mathscr{H}$ and then proceed to investigate the J-curves slightly below and above it, J_1 and J_2.

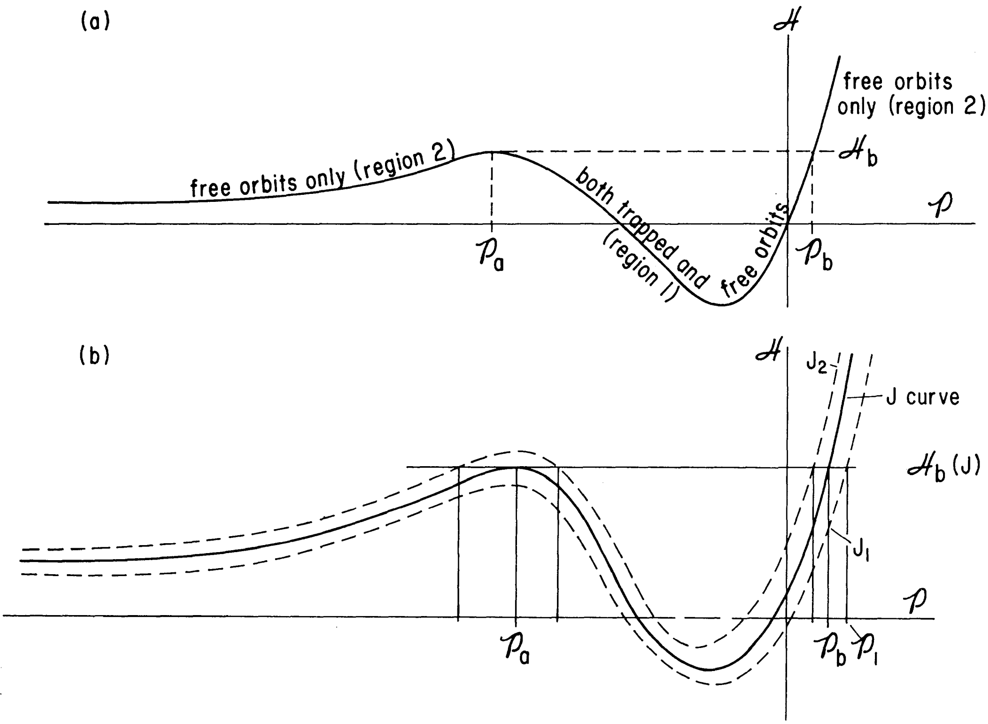


Fig. 6. (a) Types of orbits allowed on various parts of a $J(\mathcal{H}, \mathcal{P})$ = constant curve. (b) Illustration of how $\mathcal{P}$ boundaries between regions 1 and 2 for fixed $\mathcal{H}$ (to be used in the moment integrals) can be determined from the J curve whose $\mathcal{H}_b = \mathcal{H}$.

The J_1-curve has a barrier energy slightly below $\mathcal{H}_b$, while the J_2-curve has one slightly above it. We note that the J_1-curve passes through $\mathcal{H}$ at only one $\mathcal{P}$-value ($\mathcal{P}_1$), with the point ($\mathcal{H}$, $\mathcal{P}_1$) falling in region 2: in other words, no trapped particles with invariant J_1 and this $\mathcal{H}$ exist. On the other hand, the J_2 curve passes through $\mathcal{H}$ at three $\mathcal{P}$-values, with two of the ($\mathcal{H}$, $\mathcal{P}$) points falling in region 1 and the other in region 2. Hence, trapped particles with invariant J_2 and $\mathcal{H}$ exist, and their $\mathcal{P}$-values are given by the intersection of $\mathcal{H}$ = constant with the J_2-curve. But these values cannot be the extrema because, from the figure, as J_2 approaches J there are trapped orbits with $\mathcal{H}$ whose $\mathcal{P}$-values extend beyond these values for $\mathcal{P}$ at these intersections. The J-curve whose barrier $\mathcal{H}_b$ is $\mathcal{H}$ is thus a limiting curve and provides the boundaries beyond which no trapped particles exist for a given $\mathcal{H}$. The points ($\mathcal{H}$, $\mathcal{P}_A$) and ($\mathcal{H}$, $\mathcal{P}_B$) thus characterize a marginally trapped orbit for a particle with $E = \mathcal{H}_b$.

The limits of the $\mathcal{H}$ integration are straightforward. Since the moment integrals are performed at fixed z, $\mathcal{H}$ can range from $q\phi(z)$ to infinity.

The transformation from the velocity-space variables (v_x, v_y, v_z) to the Z-variables ($\mathcal{H}$, $\mathcal{P}$, $\mathcal{V}_x$) in the density and current integrals is accomplished via the following

Jacobian, derived from Equations (5) as

$$W(\mathcal{H}, \mathcal{P}, z) = \frac{1}{m^2 v_z} = \frac{1}{m^2[(2/m)(\mathcal{H} - \psi(z, \mathcal{P}))]^{1/2}} . \tag{11}$$

Note that this Jacobian puts a further restriction on the limits of the integration since $\psi(z, \mathcal{P}) \leq \mathcal{H}$, so that both the Jacobian condition and the $\mathcal{P}$-extrema condition just discussed must be used to determine the limits. Particle distributions satisfying Liouville's equation (e.g., distributions that are functions only of the dynamical constants of motion) can now be chosen separately for those particles drifting into the current sheet from outside (the boundary plasma) and those particles confined to the sheet.

The trapped particles, which 'snake' back and forth across the current sheet and gradient-B drift along it as discussed earlier, have finite excursions in all three dimensions and, therefore, remain in a finite volume of space. In the Appendix we present another proof of this behaviour which was first pointed out by Stern and Palmadesso (1975) and proved by Cowley (1978a). An examination of the expression for $\mathcal{H}$ (see Equations (2) and (5)) reveals that the trapping is due, in general, to both magnetic and electrostatic mirroring. Using the invariant description we have essentially formulated a generalized loss cone which includes electrostatic effects and can be used in regions of large electric and magnetic field gradients.

5. The Effects of Delta-Function Distributions

As a step towards our goal of obtaining fully self-consistent solutions where both Poisson's equation and the curl **B** equation are satisfied, in this section we examine the effects of individual particle orbits on the current sheet structure by considering distributions containing only a single value of the adiabatic invariant J and a single value of the total energy E. Since J and E remain constant throughout a particle's orbit, such a distribution represents a collection of particles describing the same trajectory but located at all points along that trajectory. In terms of the Kruskal picture, at any instant the particles are 'spread out' along that portion of the J-curve available to them through their energy E (see Figure 5). The current and density at a distance z from the center of the sheet will thus arise from particles of all $\mathcal{P}$-values that are dynamically possible at z.

To find the charge and current densities $\rho(z)$ and $i(z)$ resulting from particles moving along identical orbits in a current-sheet field, the product of delta-functions in E and J is used for the phase density F and introduced into the moment equations (12), of the form

$$\begin{aligned} \rho(z) &= q \int\int\int \mathrm{d}\mathcal{P}\, \mathrm{d}\mathcal{H}\, \mathrm{d}\mathcal{V}_x\, F(E, J) W(\mathcal{H}, \mathcal{P}, z) , \\ i(z) &= q \int\int\int \mathrm{d}\mathcal{P}\, \mathrm{d}\mathcal{H}\, \mathrm{d}\mathcal{V}_x\, v_y(\mathcal{P}, z) F(E, J) W(\mathcal{H}, \mathcal{P}, z) . \end{aligned} \tag{12}$$

Because of the delta-function in E, the $\mathscr{V}_x$ integral is easily evaluated, yielding

$$\int_{-\infty}^{\infty} \mathrm{d}\mathscr{V}_x\,\delta(E-E') = \frac{2}{[2m(E-\mathscr{H})]^{1/2}}\,, \quad \mathscr{H}<E\,; \qquad = 0\,, \quad \mathscr{H}>E\,. \tag{13}$$

Though the delta-function in J leads to a simple form for the $\mathscr{H}$ integral evaluated at fixed $\mathscr{P}$, the transformation from an integral over $\mathscr{H}$ to one over J, which introduces $(\mathrm{d}J/\mathrm{d}\mathscr{H})_{\mathscr{P}}$, and the function $\mathscr{H}(\mathscr{P})_J$ appearing in the denominator (i.e., the J-curve of Figure 5), necessitates its numerical evaluation. A further complication arises in that the limits of the final $\mathscr{P}$ integration are given by the intersection of the parabola condition required by the Jacobian and the J-curve, so that the $\mathscr{P}$ limits must also be evaluated numerically. The limits of the $\mathscr{H}$ integration range from $q\phi(z)$ to the total energy E. Figure 7 shows graphically the result of the $\mathscr{H}$ integral for the current at z produced by ions on a particular free orbit. Because of the delta-function in J, contributions to the $\mathscr{H}$ integral can arise only from points on the J-curve; hence, for each fixed value of $\mathscr{P}$ (represented by the vertical dashed line) exactly one point on the J-curve contributes.

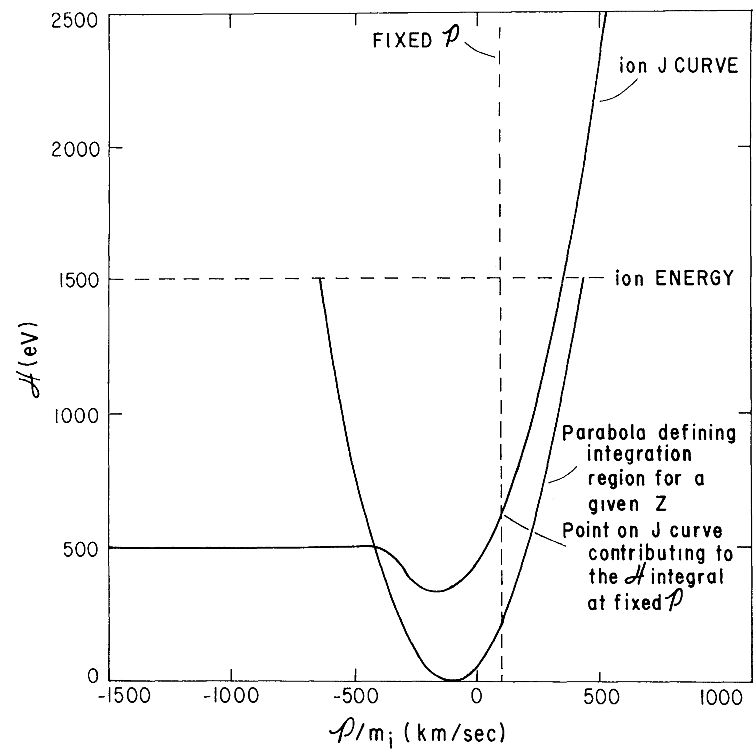


Fig. 7. Illustration of how both the parabola from the Jacobian in (11) and the J-curve are involved in the contribution to the $\mathscr{H}$ integral at fixed $\mathscr{P}$ for a free ion orbit.

Finally, the remaining integral over $\mathscr{P}$ has the form

$$\int_{\mathscr{P}_L}^{\mathscr{P}_R} \frac{d\mathscr{P}}{[E - \tilde{\mathscr{H}}]^{1/2}} \frac{1}{(dJ/d\mathscr{H})_{\mathscr{P}}} \frac{K(\mathscr{P}, z)}{[\tilde{\mathscr{H}} - \psi(\mathscr{P},z)]^{1/2}}, \tag{14}$$

where $\tilde{\mathscr{H}}$ is the function $\mathscr{H}(\mathscr{P}, J)$ depicted in Figure 7 and $K(\mathscr{P}, Z)$ is $\mathscr{P} - qA(z)$ for the current integral and 1 for the density integral. The limits $\mathscr{P}_L$ and $\mathscr{P}_R$ are either the left and right intersections of the Jacobian parabola with the J-curve, or minimum and maximum $\mathscr{P}$-values imposed by the condition $\mathscr{H} \leq E$. Figure 8 depicts graphically the $\mathscr{P}$ integration procedure at two z locations for a free-electron orbit. At z_1, the integration picks up the steep rise of the J-curve that corresponds to the 'S-shaped' motion of the electrons back and forth across the sheet's central plane, while at z_2 the integration picks up none of this motion but rather the gradient-B drift of particles spiraling around field lines at moderate distances from the sheet's center.

Using the delta-function distributions, we have investigated the contributions of single ion and electron orbits to the charge and current densities in a current sheet. Our approach was to specify a current-sheet field configuration, select an orbit by choosing

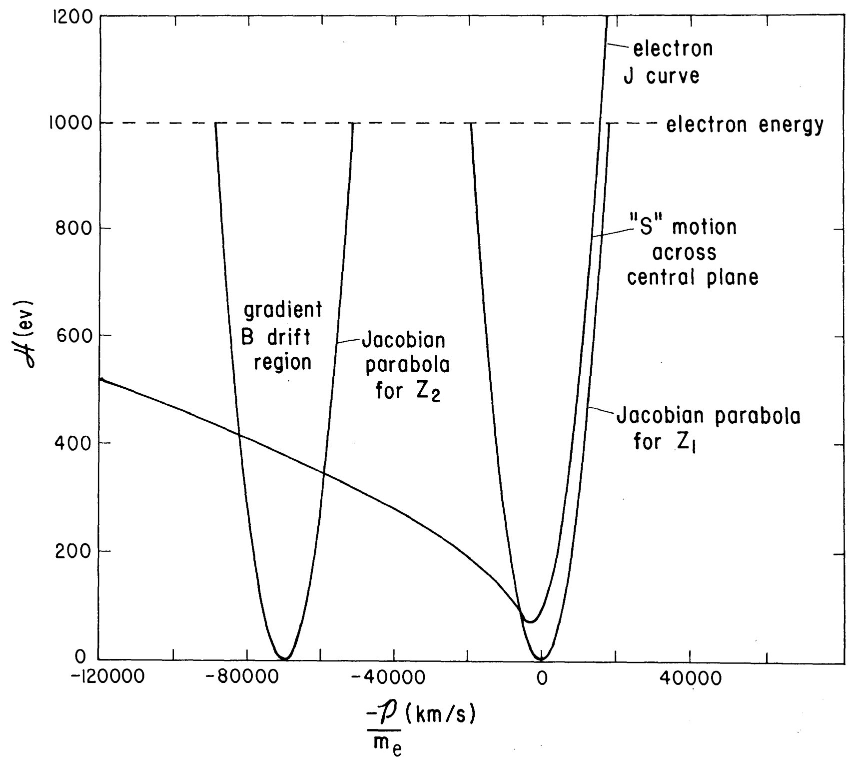


Fig. 8. Illustration of how the $\mathscr{P}$ integral at two different z-locations for free-electron orbits is determined by the intersection of the Jacobian parabola with the J-curve.

J and E, and compute the resulting $J(\mathscr{H}, \mathscr{P})$-curve. We then performed the current and density integrals outlined above to see how particles on a particular orbit contribute to the overall field.

For the current-sheet field we used the configuration that is 'magnetically self-consistent' with an ion delta-function distribution where $E = 10$ keV, $J = (0.5 \text{ keV})(4\pi m/eB_0)$, and the asymptotic field $B_0 = 7$ nT (the model field of Figure 2). By 'magnetically self-consistent' we mean that the magnetic field is consistent with the ion motion and currents. Since the motion assumes no electric field in spite of large amounts of space charge, the model is not electrostatically self-consistent. A useful way to characterize the invariant J is by the transverse kinetic energy that the

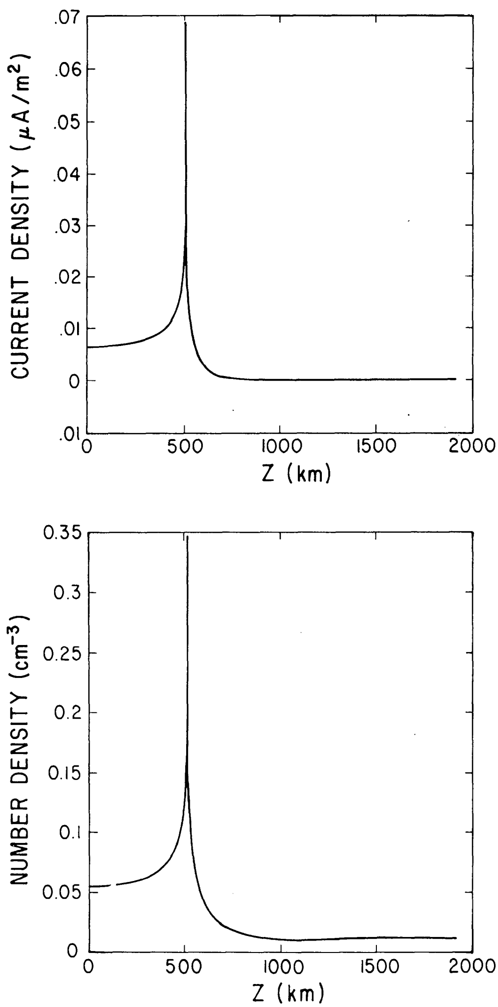


Fig. 9. Current and number densities for 10 keV ions in the model magnetic field ($\mathscr{H}_\infty = 0.5$ keV).

particle has far from the current sheet (0.5 keV in this case). We denote this energy by

$$\mathscr{H}_\infty(J) = (eB_0 J)/(4\pi m)\,. \tag{15}$$

This relation follows from the asymptotic limit of (7) where the invariant equals the magnetic moment.

Figure 2 shows the magnetic field produced by 10 keV ions with $\mathscr{H}_\infty(J) = 0.5$ keV, while Figure 9 shows the associated current and charge densities. Both the current and charge densities are sharply peaked, producing a field whose shape is somewhat more structured than the modified Harris field (Figure 2). Figure 10 shows the magnetic fields produced by ion trajectories with the same $\mathscr{H}_\infty(J)$ but lower total energies orbiting in the same model field of Figure 2. A useful parameter for characterizing these orbits is R, the asymptotic ratio of parallel to transverse kinetic energy given by $(E - \mathscr{H}_\infty(J))/\mathscr{H}_\infty(J)$. ($R$ is the square of the cotangent of the pitch angle at infinity.) From the figure it is apparent that the fields produced by 'low R' orbits cannot alone support the current-sheet structure. In fact, only those free orbits with R of 1 or greater appear capable of reinforcing the current-sheet field, as has been suggested by Rich *et al.* (1972) and by Cowley (1978b).

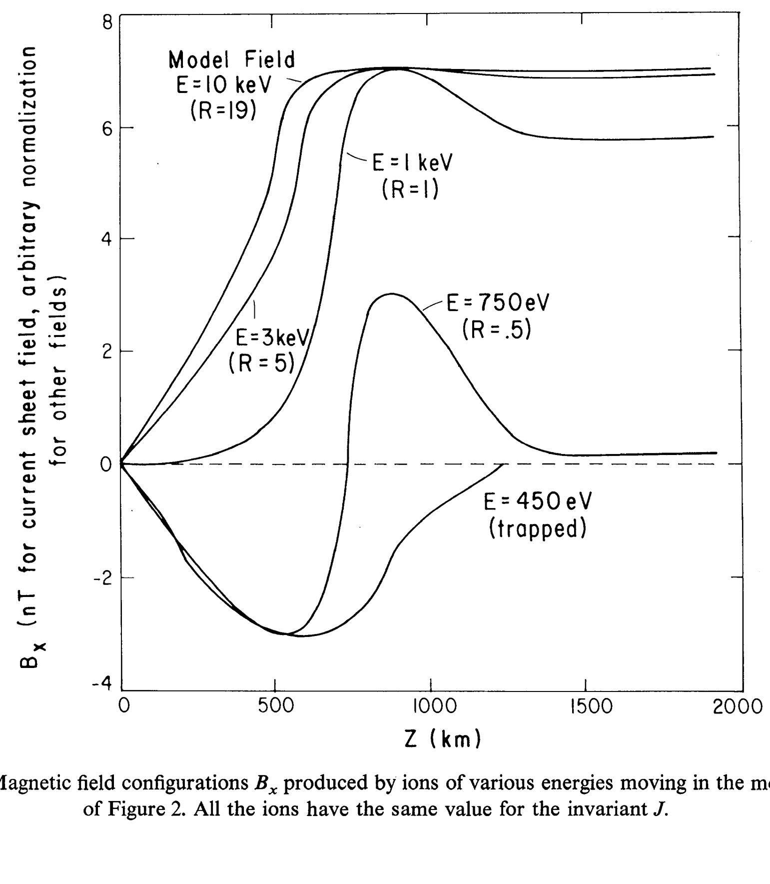


Fig. 10. Magnetic field configurations B_x produced by ions of various energies moving in the model field of Figure 2. All the ions have the same value for the invariant J.

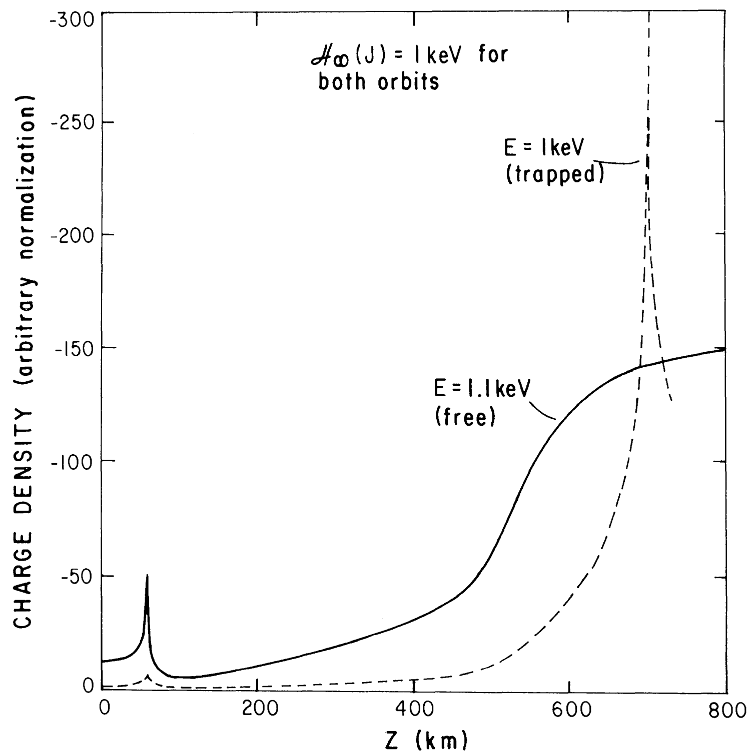


Fig. 11. Charge density configurations produced by free and trapped electrons orbiting in the model field.

Finally, Figures 11, 12, and 13 show the densities, currents, and magnetic fields for 1.1 keV and 1 keV electrons orbiting in the field of Figure 2. Both orbits have $\mathcal{H}_\infty(J) = 1$ keV, so that the 1.1 keV orbit is marginally free while the 1 keV orbit is marginally trapped by $\mathcal{H}_b$. Although the energies are close, the densities, currents, and fields look very different for the trapped and free orbits. Since the trapped electrons are spatially confined and the current is an even function of z, the trapped current forms closed loops on both sides of the current sheet and thus produces no magnetic field outside the region of confinement (Cowley, 1978a). The free electrons, on the other hand, are present infinitely far from the current sheet. As z approaches ∞ their current drops continuously to 0, their density approaches an analytically calculable limit, and in the course of their orbits they can, unlike trapped particles, experience a net displacement in y. This net displacement produces the net charge transport that reinforces the current-sheet field when R is large; when R is small, the net displacement produces either an inconsequential reinforcing current or a contrary current.

As is apparent from the figures, the currents and densities produced by the delta-function distributions are extremely spiky, and though they are qualitatively alike for ions and electrons, the distance and velocity (and, hence, current) scales for ions and

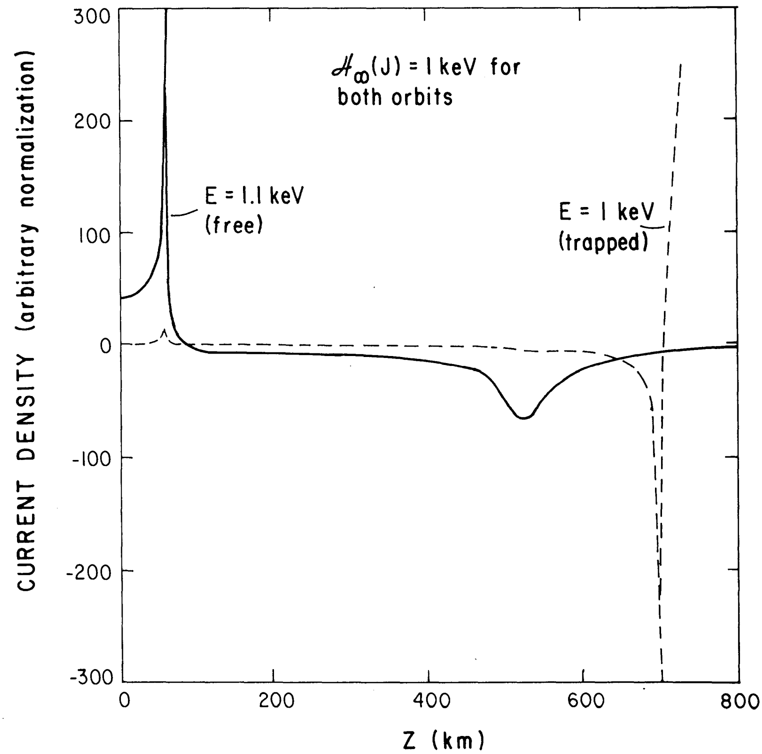


Fig. 12. Current density configurations produced by free and trapped electrons orbiting in the model field.

electrons of comparable energy are markedly different because of their mass discrepancy. A self-consistent solution, therefore, cannot be found by superposing small numbers of delta-function orbits. This situation is not surprising, since a real current sheet consists of particles with a continuous range of energies and invariants, and a realistic self-consistent solution must be some superposition of all the orbits present.

6. Fully Self-Consistent Solutions Using a Bi-Maxwellian Ion Distribution

Equipped with the invariant J and guided by the results of the preceding section, we illustrate in this section how to obtain fully self-consistent solutions by considering a simple situation in which the sheet current is supported entirely by free ions while the space charge is neutralized by free and trapped electrons. Such solutions are illustrative and fairly easy to obtain.

Since only those orbits with large asymptotic ratios of parallel to perpendicular energy can support a current-sheet structure (Section 5), we consider an ion distribution that far from the sheet is bi-Maxwellian with 'parallel' and 'perpendicular' temperatures equal to 10 keV and 0.5 keV – those values of E and $\mathcal{H}_\infty(J)$ that produced our model

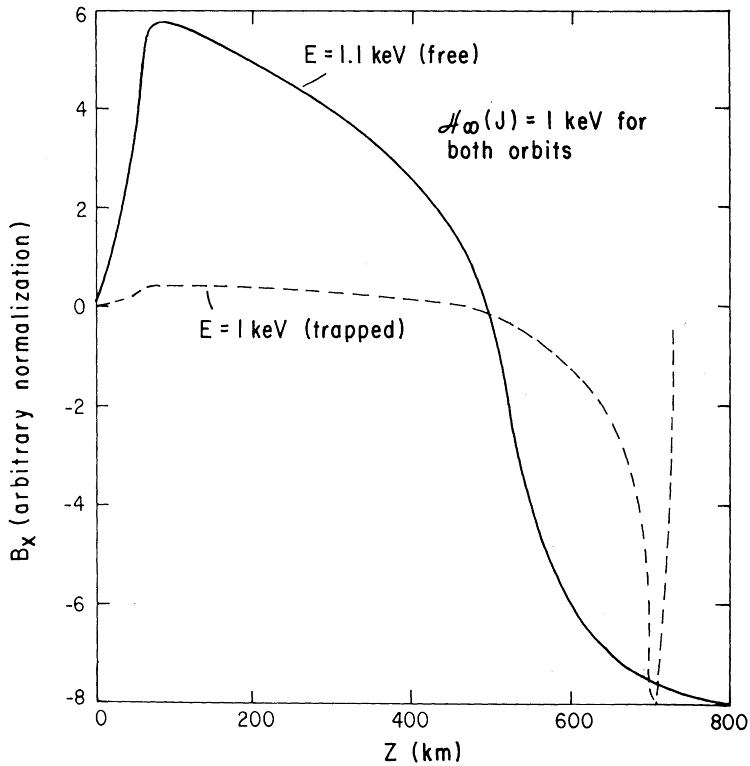


Fig. 13. Magnetic field configurations B_x produced by free and trapped electrons orbiting in the model field.

field of Section 5. Because of the proportionality between $\mathcal{H}_\infty(J)$ and J (Equation (15)), we can express such a distribution in terms of J and the total energy E as

$$F_i(J, E) = N_0(M_i/2\pi kT_\perp)\,(M_i/2\pi kT_\parallel)^{1/2} \times$$
$$\times \exp(-(E - \mathcal{H}_\infty(J))/kT_\parallel)\exp(-\mathcal{H}_\infty(J)/kT_\perp)\,. \tag{16}$$

Since J and E are constant during a particle's orbit, this F_i satisfies Liouville's equation. In the vicinity of the sheet, F_i is no longer a simple velocity-space bi-Maxwellian since $\mathcal{H}$ and J are no longer simply related by (15) but by the more complex expression (7). Though F_i generally includes particles on both trapped and free orbits, we exclude the trapped orbits by integrating over only that part of phase space corresponding to free orbits, as outlined in Section 4. A great advantage of bi-Maxwellian over delta-function distributions is that the spikiness apparent in Figures 7 and 8 disappears when contributions from orbits of all R-values are added. The resulting densities and currents are then smooth functions that decrease monotonically away from the sheet's center.

For the electron distribution we assume an isotropic Maxwellian in all parts of phase space, so that

$$F_e(E) = N_0(M_e/2\pi kT_e)^{3/2} \exp(-E/kT_e). \tag{17}$$

This F_e, which contains both free and trapped orbits, is convenient because it always yields

$$i_e(z) = 0 \quad \text{and} \quad \rho(z) = -eN_0 \exp(e\phi(z)/kT_e),$$

for the current and charge densities, respectively, and these quantities can, therefore, be obtained analytically once $\phi(z)$ is specified. In fact, if there is a 'magnetically self-consistent' solution for F_i of (16) as for the large-R delta-function distributions of Section 5, this solution is fully self-consistent in the cold electron limit (T_e approaches 0) since the electrons then exactly neutralize the ion space charge without contributing current.

Because the J curves are functionals of the fields B_x and E_z, we suspect that iteration on these fields is the easiest way to obtain solutions. Fortuitously, the apparent lack of sensitivity of large-R orbits to the detailed structure of $B_x(z)$ resulted in $B_x(z)$ converging after several iterations. The electrostatic potential converged quickly for low T_e ($\sim$ 1 keV), but convergence became touchy as T_e was increased to 5 keV. As our starting functions we used the model **B** field of Figure 2 and $\phi = 0$. We show the fully self-consistent currents, densities, magnetic fields, and potentials for electron temperatures of 0, 1, and 5 keV in Figure 14. In the solution for T_e = 5 keV, the ϕ convergence is not perfect; we have, therefore, concluded that 5 keV is about as hot as the electrons can get and still neutralize the ion space charge. We note that the self-consistent magnetic fields are nearly identical to Harris fields, but we hesitate to attach significance to this apparent coincidence.

Two remarks about these solutions are in order. First, while the boundaries between regions 1 and 2 in the $\mathscr{H}$, $\mathscr{P}$-plane (Section 4) are easy to obtain when B_x is smooth and $\phi = 0$, they generally become more complicated for finite ϕ. Because of the large anisotropy in the ion distribution, however, the self-consistent magnetic fields, in conjunction with positive (or zero) potentials, give rise to J-curves that for most ions have very shallow minima. As a consequence the integral over all ion orbits, both free and trapped, yields essentially the same numerical result as the integral over free orbits alone. This circumstance greatly simplifies the computation since we need not trace boundaries for each iteration; we can still claim, however, that our model current sheet consists entirely of free ions – an important statement about its physical character.

Secondly, we emphasize how inadequate a guiding-center description is for our particular model sheet. If B_x and ϕ did not vary much over a Larmor radius, then the invariant J would always be proportional to $\mathscr{H}/B_x(z_c)$, where z_c is the ion guiding center. As the ion moved toward the sheet its $\mathscr{H}$-value would decrease as $B_x(z_c)$ until in the central plane $\mathscr{H}$ would be very small. In the guiding center description, therefore, an ion does not experience the sharp rise in $\mathscr{H}$ (i.e., $\mathscr{H}$ approaches infinity for small and positive $\mathscr{P}$ in Figure 5) that corresponds to the S-shaped, positively y-drifting motion of the ion across the sheet in the y – z-plane. Whereas the Kruskal invariant describes

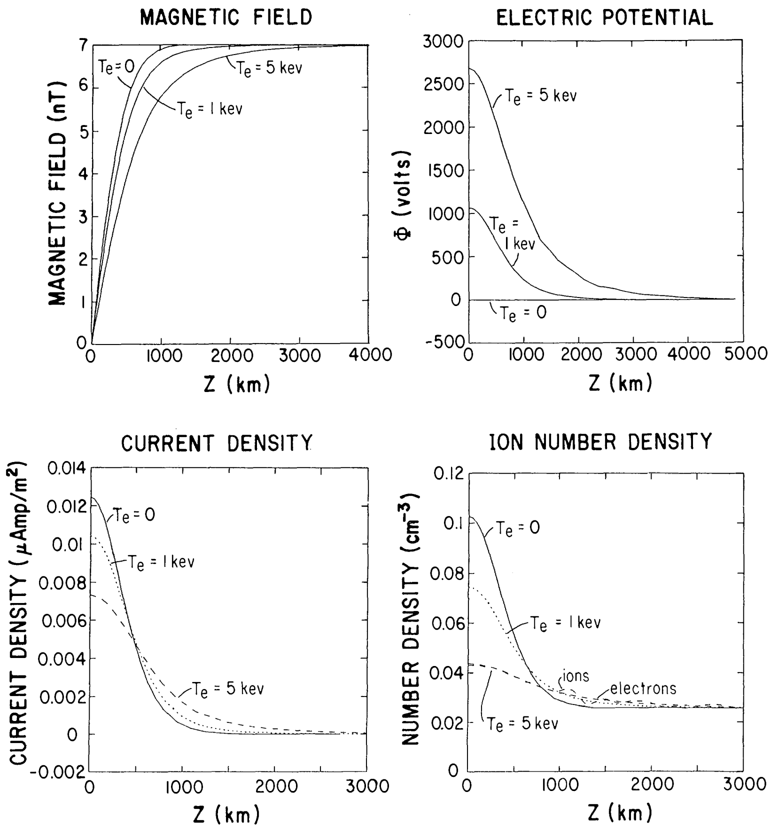


Fig. 14. Self-consistent magnetic fields, potentials, currents, and densities for bi-Maxwellian ions and Maxwellian electrons. Both ion and electron densities are shown for the $T_e = 5$ keV case where the ϕ convergence is not perfect.

how at the sheet's center all the ion's energy is converted into the perpendicular energy $\mathscr{H}$ that generates the sustaining current, the guiding center invariant does not contain this crucial feature of the orbits. To demonstrate the S-shaped motion for ions in our example, we introduced the self-consistent field solutions for $T_e = 1$ keV into our trajectory code; an illustrative part of one resulting free ion orbit is shown in Figure 15.

Finally, we comment on the invariant approach in the light of recent work that finds that current sheet orbits can behave chaotically. When the particle reenters a side-well from the central well, the particular side-well it enters cannot be predicted from the drift parameters alone and most likely depends on the particle's gyrophase (the generalized guiding center parameter determines only $|z|$, not z itself, through Equation (8)). This behaviour indicates that we have washed out the gyrophase-dependent stochastic elements that affect real orbits, as described by Chen and Palmadesso (1986).

Stochastic effects of greater importance probably arise, or are amplified, when the invariant approach breaks down. Kruskal's theory is a perturbation procedure whose

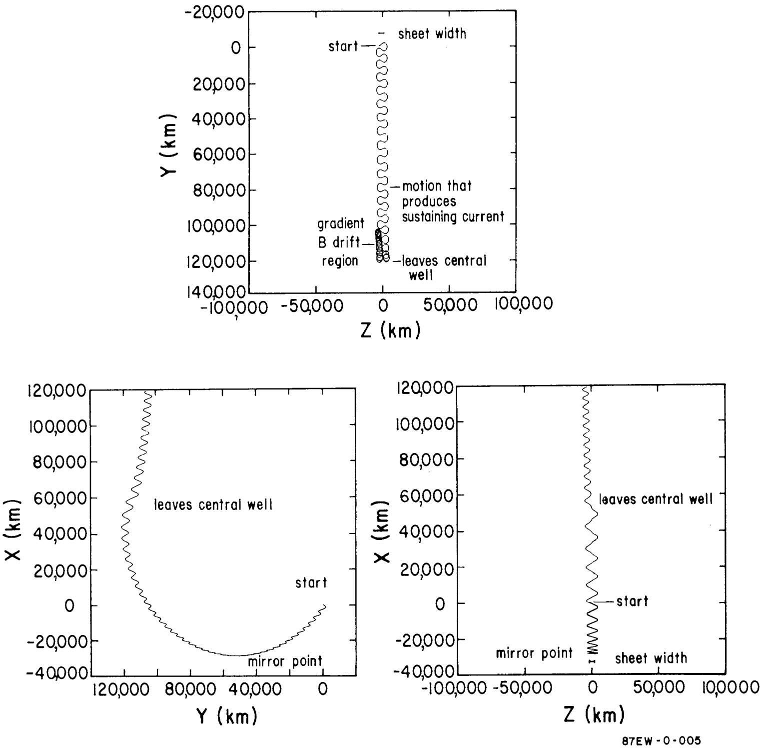


Fig. 15. Part of a 13 keV ion trajectory in the self-consistent current sheet of Figure 14 with $T_e = 1$ keV.

validity depends on the parameter β being small. Though we have not rigorously determined when and how badly the theory breaks down, we have traced enough orbits to offer some observations. For a β/B_0 of 0.05, where B_0 is the magnitude of B_x at the edges of the current sheet, the invariant approach describes the 'real' situation (e.g., Figure 4) quite well. The computed J seems indeed to be invariant, and the particles gyrate in and out of the sidewells in accordance with expectations, even though the trajectories themselves are complex. For a β/B_0 as small as 0.1, however, there are clearly orbits on which J is not invariant. Which and what portion of the orbits do not conserve J for a given β, how greatly the J's vary, and how the departures affect the validity of the theory are questions we have not addressed. Instead, we regard our work as a 'first-pass' quantitative method for formulating self-consistent solutions that include the normal field component and allow for distinctions between trapped and free plasmas in the magnetotail.

Acknowledgements

We especially thank T. G. Northrop for many helpful discussions during the course of this work. We have benefitted also from the encouragement and stimulus of. H. Alfvén who has been extremely generous in giving of his time in personal interactions, and gracious in his comments in his weekly seminars at UCSD. Both authors wish to thank him personally for his support and extend to him our congratulations on this occasion. Part of this work was performed while E.W. was supported at Goddard Space Flight Center by NASA grant NGR 21–002–316 with the University of Maryland. The work at UCSD was supported by NSF Grant ATM83–18445 and also by NASA Grants NSG 7623 and NGL 05–005–007.

Appendix. Proof of Cowley's Theorem that Trapped Particles Experience no Net Drift

In this Appendix we use our generalized adiabatic treatment to prove Cowley's (1978a) theorem that particles trapped in the current layer have finite excursions in all three dimensions and, therefore, remain in a finite volume of space. By trapped particles we mean those indicated in Figure 5 by the trajectories labelled E_1 or E_3. These particles oscillate back and forth in their drift motion. In addition they are performing gyro-oscillations on a smaller scale. Our proof consists of a demonstration that the average drift velocity (averaged over a drift oscillation) vanishes. We use bar notation to indicate averages over the gyro-oscillation and brackets to indicate averages over the drift motion – i.e.,

$$\bar{v}_x = \frac{1}{T_G} \oint v_x \, \mathrm{d}t \,, \text{ etc.}, \quad \text{and} \quad \langle v_x \rangle = \frac{1}{T_D} \oint \bar{v}_x \, \mathrm{d}t \,, \tag{A1}$$

where T_G is the gyroperiod and T_D is the drift period. The circle on the first integral indicates an integral over the gyro-motion (with $\mathscr{H}$, $\mathscr{P}$, and $\mathscr{V}_x$ held constant), and the box on the second integral indicates an integral over the drift motion (with E and J held constant). We make use of the following relations which can be obtained from Equation (7):

$$\left(\frac{\partial J}{\partial \mathscr{H}}\right)_{\mathscr{P}} = T_G \,, \qquad \left(\frac{\partial J}{\partial \mathscr{P}}\right)_{\mathscr{H}} = -\bar{v}_y T_G \,, \qquad \left(\frac{\partial \mathscr{H}}{\partial \mathscr{P}}\right)_J = +\bar{v}_y \,. \tag{A2}$$

Now the x-component of drift velocity $\bar{v}_x$ is equal to the Kruskal Z-variable $\mathscr{V}_x$. Therefore, we can write

$$\langle v_x \rangle = \frac{1}{T_D} \oint \frac{\mathscr{V}_x \, \mathrm{d}\mathscr{P}}{(\mathrm{d}\mathscr{P}/\mathrm{d}t)} = \frac{1}{T_D} \oint \frac{\mathscr{V}_x \, \mathrm{d}\mathscr{P}}{(-\beta q \mathscr{V}_x)} = \frac{-1}{\beta q T_D} \oint \mathrm{d}\mathscr{P} = 0 \,. \tag{A3}$$

The substitution for $(\mathrm{d}\mathscr{P}/\mathrm{d}t)$ comes from (6) and the last step follows from the fact that the net change in $\mathscr{P}$ for the closed drift orbit is zero.

Similarly,

$$\langle v_y \rangle = \frac{1}{T_D} \oint \bar{v}_y \, \mathrm{d}t = \frac{+1}{T_D} \oint \frac{(\mathrm{d}\mathscr{H}/\mathrm{d}\mathscr{P})_J \, \mathrm{d}\mathscr{P}}{(\mathrm{d}\mathscr{P}/\mathrm{d}t)} =$$

$$= \frac{-(m/2)^{1/2}}{\beta q T_D} \oint \frac{\mathrm{d}\mathscr{H}}{\sqrt{[E - \mathscr{H}]}} = 0 , \tag{A4}$$

with the substitution for $\bar{v}_y$ given by (A2), and $\mathscr{V}_x$ has been expressed in terms of E and $\mathscr{H}$. Finally,

$$\langle v_z \rangle = \frac{1}{T_D} \oint \bar{v}_z \, \mathrm{d}t = \frac{1}{T_D} \oint \frac{\mathrm{d}\bar{z} \, \mathrm{d}t}{\mathrm{d}t} = \frac{1}{T_D} \oint \mathrm{d}\bar{z} = 0 , \tag{A5}$$

where we have used the relation between the particle guiding position $\bar{z}$ and $\mathscr{P}$ given by (8) to write $\bar{v}_z = \mathrm{d}z/\mathrm{d}t$. The last step follows from the one-to-one correspondence between $\bar{z}$ and $\mathscr{P}$, and the fact that there is no net change of $\mathscr{P}$ in the drift oscillation.

References

Akasofu, S.-I., Hones, E. W., Bame, S. J., Asbridge, J. R., and Lui, A. T. Y.: 1973, *J. Geophys. Res.* **78**, 7257.
Alfvén, H. A.: 1939, *J. Kungl. Svenska Vet. Akad. Handl., 3* **18**, No. 3.
Alpers, W.: 1969, *Astrophys. Space Sci.* **5**, 425.
Alpers, W.: 1971, *Astrophys. Space Sci.* **11**, 471.
Basu, S. and Rowlands, G.: 1986, *Planetary Space Sci.* **34**, 631.
Beard, D. B. and Cowley, S. W. H.: 1985, *Planetary Space Sci.* **33**, 773.
Bird, M. K.: 1975, *Planetary Space Sci.* **23**, 27.
Bird, M. K. and Beard, D. B.: 1972a, *J. Geophys. Res.* **77**, 4864.
Bird, M. K. and Beard, D. B.: 1972b, *Planetary Space Sci.* **20**, 2057.
Birn, J.: 1979, *J. Geophys. Res.* **84**, *J. Geophys. Res.* **84**, 5143.
Birn, J., Sommer, R. R., and Schindler, K.: 1975, *Astrophys. Space Sci.* **35**, 389.
Birn, J., Sommer, R. R., and Schindler, K.: 1977, *J. Geophys. Res.* **82**, 147 (note a correction which appeared in *J. Geophys. Res.* **82**, 1964; 1977).
Chen, J. and Palmadesso, P. J.: 1986, *J. Geophys. Res.* **91**, 1499.
Cowley, S. W. H.: 1971, *Cosmic Electrodyn.* **2**, 90.
Cowley, S. W. H.: 1973, *Cosmic Electrodyn.* **3**, 448.
Cowley, S. W. H.: 1978a, *Planetary Space Sci.* **26**, 539.
Cowley, S. W. H.: 1978b, *Planetary Space Sci.* **26**, 1037.
Cowley, S. W. H. and Pellat, R.: 1979, *Planetary Space Sci.* **27**, 265.
de Hoffman, F. and Teller, E.: 1950, *Phys. Rev.* **80**, 692.
Dungey, J. W.: 1972, 'Theory of Neutral Sheets', in B. M. McCormac (ed.), *Earth's Magnetospheric Processes*, D. Reidel Publ. Co., Dordrecht, Holland, p. 210.
Dungey, J. W.: 1975, *Space Sci. Rev.* **17**, 173.
Eastman, T. E., Frank, L. A., and Huang, C. Y.: 1985, *J. Geophys. Res.* **90**, 9541.
Eastman, T. E., Frank, L. A., Peterson, W. K., and Lennartsson, W.: 1984, *J. Geophys. Res.* **89**, 1553.
Eastwood, J. W.: 1972, *Planetary Space Sci.* **20**, 1555.
Eastwood, J. W.: 1974, *Planetary Space Sci.* **22**, 1641.
Eastwood, J. W.: 1975, *Planetary Space Sci.* **23**, 1.
Francfort, P. and Pellat, R.: 1976, *Geophys. Res. Letters* **3**, 433.
Grad, H.: 1961, *Phys. Fluids* **4**, 1366.
Harris, E. G.: 1962, *Nuovo Cimento* **23**, 115.

Kan, J. R.: 1973, *J. Geophys. Res.* **78**, 3773.

Kruskal, M.: 1962, *J. Math. Phys.* **3**, 806.

Lee, L. C. and Kan, J. R.: 1979, *J. Geophys. Res.* **84**, 6417.

Lee, L. C. and Kan, J. R.: 1980, *J. Geophys. Res.* **85**, 37.

Lemaire, J. and Burlaga, L. F.: 1976, *Astrophys. Space Sci.* **45**, 303.

Lyons, L. R. and Speiser, T. W.: 1982, *J. Geophys. Res.* **87**, 2276.

Popielawska, B., Szalinska-Piechota, E., and Tsyganenko, N. A.: 1985, *Planetary Space Sci.* **33**, 1433.

Propp, K. and Beard, D. B.: 1984, *J. Geophys. Res.* **89**, 11013.

Rich, F. J., Vasyliunas, V. M., and Wolf, R. A.: 1972, *J. Geophys. Res.* **77**, 4670.

Roth, M. A.: 1978, *J. Atmospheric Terrest. Phys.* **40**, 323.

Roth, M. A.: 1979, 'A Microscopic Description of Interpenetrated Plasma Regions', in B. Battrick and J. Mort (eds.), *Proceedings of Conference on Magnetospheric Boundary Layers,* ESA SP-148, ESTEC, Noordwijk, The Netherlands.

Roth, M. A.: 1980, 'La structure interna de la magnetopause', Ph.D. Thesis, Belgisch Instituut voor Ruimte-Aeronomie, Brussels, Belgium.

Schindler, K.: 1972, 'A Self-Consistent Theory of the Tail of the Magnetosphere', in B. M. McCormac (ed.), *Earth's Magnetospheric Processes,* D. Reidel Publ. Co., Dordrecht, Holland, p. 200.

Schindler, K.: 1975, *Space Sci. Rev.* **17**, 589.

Schindler, K.: 1976, 'Magnetotail Model', in B. M. McCormac (ed.), *Magnetospheric Particles and Fields*, D. Reidel Publ. Co., Dordrecht, Holland, p. 79.

Schindler, K.: 1979, *Space Sci. Rev.* **23**, 365.

Sestero, A.: 1964, *Phys. Fluids* **7**, 44.

Sestero, A.: 1966, *Phys. Fluids* **9**, 2006.

Soop, M. and Schindler, K.: 1973, *Astrophys. Space Sci.* **20**, 287.

Speiser, T. W.: 1965a, *J. Geophys. Res.* **70**, 1717.

Speiser, T. W.: 1965b, *J. Geophys. Res.* **70**, 4219.

Speiser, T. W.: 1967, *J. Geophys. Res.* **72**, 3919.

Stern, D. P. and Palmadesso, P.: 1975, *J. Geophys. Res.* **80**, 4244.

Toichi, T.: 1972, *Cosmic Electrodyn.* **3**, 81.

Tsyganenko, N. A.: 1982, *Planetary Space Sci.* **30**, 433.

Whipple, E. C., Hill, J. R., and Nichols, J. D.: 1984, *J. Geophys. Res.* **89**, 1508.

Whipple, E. C., Northrop, T. G., and Birmingham, T. J.: 1986, *J. Geophys. Res.* **91**, 4149.

PARTICLE DYNAMICS AND CURRENTS IN THE MAGNETOTAIL*

R.-L. XU

Institute of Space Physics, Chinese Academy of Sciences, Beijing, China

(Received 12 June, 1987)

Abstract. This paper is a brief summary of our studies of magnetotail phenomena, based on the motion of charged particles and on electric currents. Theory and satellite data have been used together to obtain quantitative understanding.

A guiding center theory similar to that of the usual Alfvén's guiding center theory has been developed. In the region where the magnetic field is small or equal to zero, the circular motion becomes an '8'-shaped motion, and the magnetic moment is equivalent to two opposite magnetic moments.

An analytical trajectory of the particle motion has been found in the neutral sheet magnetic field. In the more general case, in the neutral line magnetic field, a particle computer simulation has been used, studying the accelerations, the pitch angular scattering and the propagation of a point source plasma in the magnetotail.

Some current descriptions in the magnetotail have been discussed. The plasma current sheet was found from the solution of a time-independent Vlasov equation, and the relation between the particle motion and the magnetic field is a selfconsistent system. We have derived an expression for a 3-dimensional function of a field-aligned current. Moreover, some simple model of the filamentary current found by the satellite data in the magnetotail lobe is discussed.

Finally, we have analysed the particle and magnetic field data observed by IMP-7, IMP-8, and ISEE-1, and found the dawn-dusk asymmetry of the particles in the neutral sheet, the magnetic current sheet, and the filamentary current in the magnetotail lobe, which are comparable to the theoretical results.

1. Introduction

The dynamics and currents in the magnetotail have been studied by many investigators in the recent years (papers and references in Akasofu, 1980; Carovillano and Forbes, 1982; Lyons and William, 1984; Potemra, 1984; and Nishida, 1982). The transport and energization of the particles in the magnetotail were carried out by active, tracer element released in the distant magnetotail in the AMPTE program (Krimigis *et al.*, 1982), which provides informations on the convective motion within the magnetotail.

In describing basic quantitative aspect of the magnetotail, a detailed mathematical description of magnetotail phenomena based on the motion of charged particles and on current has been studied in our group in the recent years. Theory and data have been used together to obtain quantitative understanding.

In order to study the physical process in the magnetotail, it is necessary to understand the particle dynamics in the magnetotail. The usual guiding center theory is not valid within the neutral sheet of the tail, since the magnetic field is small or equal to zero (Alfvén, 1950). A guiding center theory, similar to that of the usual Alfvén's guiding center theory, has been developed, when the magnetic field is a neutral sheet (Xu, 1981c, d). In the more general case, when the magnetic field is a neutral line, a simple particle computer simulation has been used in studying the acceleration (Xu *et al.*,

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 257–277.

1986b), pitch angular scattering, (Xu and Gu, 1983a; Xu *et al.*, 1986b) and propagation of particles (Xu *et al.*, 1986b) in the magnetotail.

Since most magnetospheric phenomena are of electromagnetic nature, and many magnetospheric processes are directly and indirectly related to electric currents, current description is now no doubt a tendency for magnetospheric studies (Alfvén, 1981; Potemra, 1984; Lyons and William, 1984). In the recent years, we have discussed the neutral sheet current in the magnetotail (Xu, 1981e, 1982), the field-aligned current in the night sector of the earth (Xu, 1981; Xu and Gu, 1982; Xu *et al.*, 1986b), and the filamentary current in the magnetotail lope (Li *et al.*, 1986). All of these will be summarized below.

2. Particle Dynamics of the Magnetotail

2.1. Model

The transport and energization of a point source plasma in the magnetotail which has been studied in the AMPTE Program are a very complex problem (Krimigis *et al.*, 1982). We study this problem by means of computer simulation. Since the problem is a 3-dimensional one, as the first step of our study we begin with the motion of a lot of particles starting from the point source without considering the electromagnetic field caused by the motions (Xu and Gu, 1983; Xu and Jin, 1984; Xu *et al.*, 1986b, c).

Let the origin of the rectangular coordinate system x, y, z, be at the middle point of the magnetotail, which is twenty to thirty Earth radii away from the Earth's center, as shown in Figure 1. In this rectangular coordinate, the x-axis points to the Earth, the y-axis points to the dawn-dusk direction, and the z-axis to the north. The $x - y$ plane coincides with the neutral sheet of the magnetotail. The electromagnetic field of the magnetotail can be looked upon approximately as a neutral magnetic field and the dawn-dusk electric field is given by

$$\mathbf{B} = \frac{2z}{a^2}\mathbf{e}_x + \frac{2x\varepsilon}{a^2}\mathbf{e}_z \tag{1.1}$$

and

$$\mathbf{E} = E\mathbf{e}_y\,, \tag{1.2}$$

where

$$\varepsilon = \frac{a^2}{b^2} \tag{1.3}$$

is the parameter of the field. When $\varepsilon = 0$, the magnetic field is a neutral sheet magnetic field, and it is a linear reversal one with field lines parallel with the $z = 0$ plane. In the magnetotail B_z is not equal to zero but is very small, $\varepsilon \ll 1$.

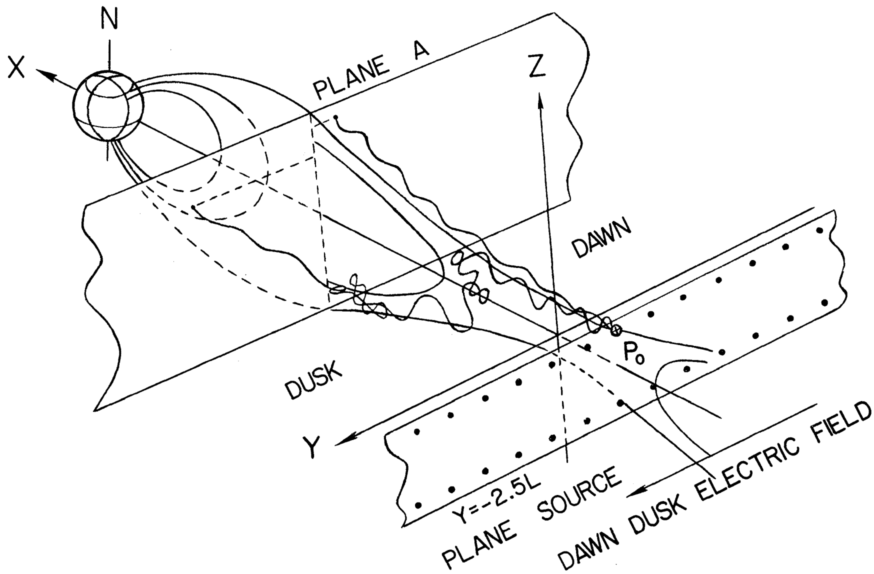


Fig. 1.

Let us suppose that plasma is released from a point source $P_0(X_0 = -2.5, Y_0 = 0.0, Z_0 = 1.0)$ near the neutral line of the magnetotail, as shown in Figure 1. Since the process is collisionless, we can study it by means of the particle trajectories, which are evaluated numerically in terms of the standard Runge–Kutta method. The initial pitch angular distribution of the plasma source is supposed to be isotropic, with an angle interval $\Delta\phi = 5°$ and $\Delta\lambda = \Delta\phi/\sin\phi$, so that 856 particles start from this fixed point. Suppose $\varepsilon = \frac{1}{25}$ and $T = 0.0$, 0.2, and 0.4 for different electric field intensities, where $T = eE/mv_0$ is the electric fielkd parameter. In a $50L$ wide magnetotail, $T = 0.2$ and 0.4 correspond to potentials of 240 and 480 kV across the whole tail. In order to study the propagation of the particles starting from the point source, particles are collected in a series of time which correspond to $s = 2.5$, 5.0, 7.5, 10.0, and 12.5, where $s = v_0 t$ (Xu *et al.*, 1986c). To study spatial distribution, particles are also collected in a series of $X = X_n$ planes, supposing $X_n = 14L$ (Xu and Gu, 1983; Xu and Jin, 1984; Xu *et al.*, 1986b).

2.2. The trajectories of the particles

The motion of charged particles in the neutral magnetic field is a very complex problem (Xu *et al.*, 1980, 1983a; Xu, 1981a, b). Near the neutral line, the magnetic field is very small, so the Alfvén's guiding-center theory cannot be used, and the problem must be studied by the equation of motion of charged particles given by

$$\frac{\mathrm{d}v_x}{\mathrm{d}s} = \frac{4\varepsilon x v_y}{L_0^2}\,,$$

$$v_y = \frac{2v_0(Z^2 - \varepsilon x^2)}{L_0^2} - \frac{2v_0(Z_0^2 - \varepsilon x_0^2)}{L_0^2} + \frac{eE}{m}t + v_{y0}\,, \tag{1.4}$$

$$\frac{\mathrm{d}v_z}{\mathrm{d}s} = \frac{4zv_y}{L_0^2}\,;$$

where

$$L_0 = \sqrt{\frac{e}{2cma^2v_0}} \tag{1.5}$$

is the unit length, t, c, m, v, and e correspond to time, velocity of light, mass, velocity, and charge of the particle, respectively. Parameters with a subscript 'o' are the initial values of the corresponding parameters. Let ϕ be the initial angle between v_0 and the x-axis, and γ be the initial angle between the component of v_0 in yz-plane and the y-axis. In the neutral sheet magnetic field ϕ is the initial pitch angle and γ is the initial direction of the particle in the plane perpendicular to the magnetic field.

According to Equation (1.4), a lot of particle trajectories with different initial conditions are evaluated numerically by means of the standard Runge–Kutta method. Figure 2 is a group of trajectories in unit of L_0, with initial condition $x_0 = -2.5$, $y_0 = 0.0, z_0 = 1.0, \gamma = 0.0$ for different initial direction ϕ. We suppose that $\varepsilon = 0.01$ and $T = 0.2$. Figure 2(a) is the projection of the trajectories on the xz-plane, which corresponds to the center meridian plane. Figure 2(b) is the projection on the cross-section of the tail (yz-plane), and Figure 2(c) is the projection on the equatorial plane when looking upward from the south (xy-plane).

The trajectories of the particles are very complicated. According to Figure 2, particles with small pitch angle ($\phi = 10$, 20, and 30°) do not pass through the neutral sheet. The Alfvén's perturbation method can be used. The motion can be divided into two kinds of motion. One is a circular motion around its guiding center, and the other is the motion of its guiding center. We call these particles perturbation (P) particles. When the pitch angles of the particle are larger than 30°, the particles pass through the neutral sheet. The Alfvén's perturbation method cannot be used, since the magnetic field is very small or equal to zero. We call the region near the neutral sheet where the perturbation method cannot be used the non-perturbation (N) region, and the particles non-perturbation (N) particles. For N particle the circular motion around its guiding center becomes an '8' shape or an oscillationy motion, which correspond to $\phi = 40°$ or 60° and 70°, respectively, in Figure 2(a).

Moreover, according to Figure 2(b) and 2(c), we have found that there is a dawn-dusk separation between N and P particles. As a result of the gradient B drift, P particles move

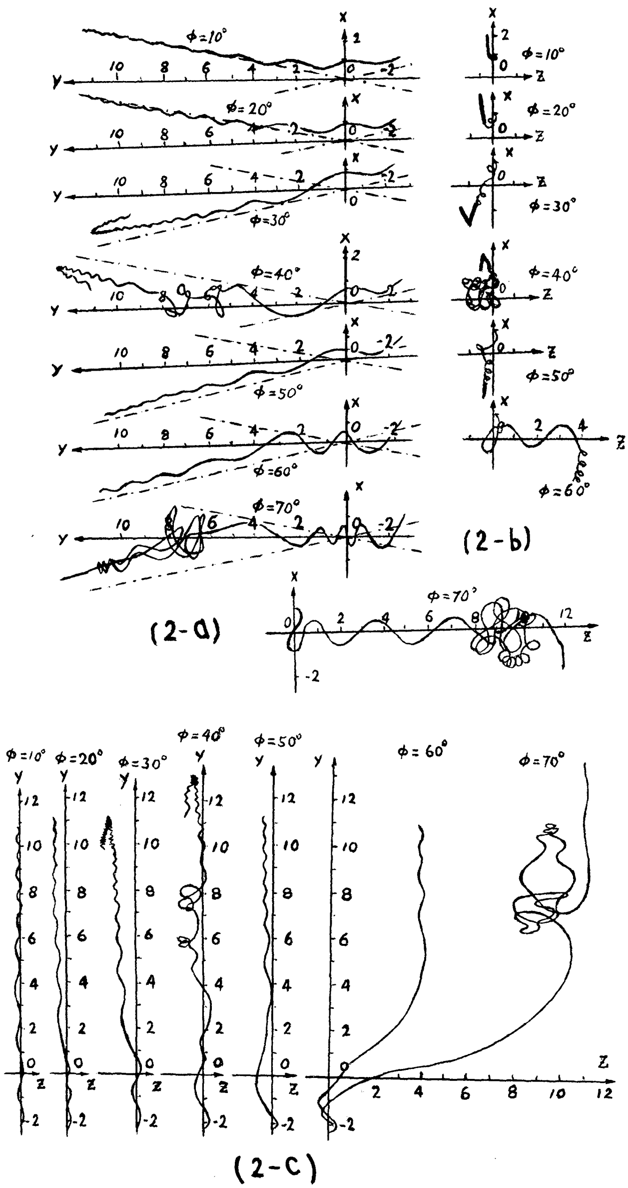


Fig. 2.

along the dusk-dawn direction, but N particles drift along the dawn-dusk direction in the N region. These results will be discussed in more details in the following section.

2.3. Analytical trajectory of the particle motion

The analytical solution of the equation of the motion (Equation (1.4)) of a charged particle moving in the neutral magnetic field can be found by means of the perturbation technique when the field is a neutral magnetic sheet ($\varepsilon = 0$) (Xu, 1981c, d). The analytical solution of Equation (1.4) can be written as

$$\mathbf{R}(t) = \mathbf{V}t + \mathbf{R}'(t) + \mathbf{R}_c =$$
$$= (V_\perp \mathbf{e}_\perp + V_\| \mathbf{e}_\|)t + (Z'\mathbf{e}_Z + y'\mathbf{e}_y) + Z_c\mathbf{e}_Z\,, \tag{1.6}$$

where $\mathbf{R}_c(t)$ is the position of the center of the trajectory. $\mathbf{R}'(t)$ is the closed oscillatory motion, $V_\perp$ and $V_\|$ are the component of the velocity of $\mathbf{R}_c$ which are perpendicular and parallel to the magnetic field. The expressions of (1.6) in the perturbation (P) region and nonperturbation (N) region are different from one another. In the N region

$$z' = Z_a(1 - \beta - \beta^2 + \cdots)\cos\Omega t + Z_a\beta\cos 3\Omega t + Z_a\beta\cos 5\Omega t + \cdots\,,$$
$$y' = Y_a(1 - 3\beta^2 + \cdots)\sin\Omega t + Y_a\beta\sin 4\Omega t + \beta^2\sin 6\Omega t + \cdots\,, \tag{1.7}$$

where

$$Y_a = \frac{Z_a^2}{4\sqrt{2}\,\omega} < 1\,,$$

$$\beta = \frac{Z_a^2}{32\,\omega^2} < 1\,,$$

$$\Omega = 2\sqrt{2}\,v\omega\sin\phi_0\,,$$

$$\omega^2 = \frac{2 - Z_a^2 + \sqrt{4 - 4Z_a^2 - Z_a^4/2}}{8}\;;$$

where Ω is the frequency of the closed oscillatory motion. The closed oscillatory motion is an '8'-shaped motion, as shown in Figure 3. In Figure 3, curves ①, ②, ③, ④, and ⑤ correspond to $Z_a = 0.2$, 0.4, 0.6, 0.8, and 1.0, respectively. The center of the '8'-shape move at a drift velocity of

$$V_\perp = v\sin\phi[1 - Z_a^2(1 + 2\beta)]\,. \tag{1.8}$$

Its direction is perpendicular to the magnetic field and parallel to the neutral line. The particle moves along the magnetic field at a velocity of

$$V_\| = v\cos\phi\,. \tag{1.9}$$

In the P-region, the closed oscillatory motion is a circular motion. The velocity of the center of the circular motion is equal to the gradient B drift, especially when the position of the center is far away from the neutral sheet. The velocity along magnetic field is the same as Equation (1.9).

Consequently, we have found a complete analytical form of trajectories, which agree quite well with the numerical trajectories evaluated by the computer, except for a slight deviation near the boundary of the P and N region.

The above results are applicable to protons (Xu, 1981e, 1982). For electrons the

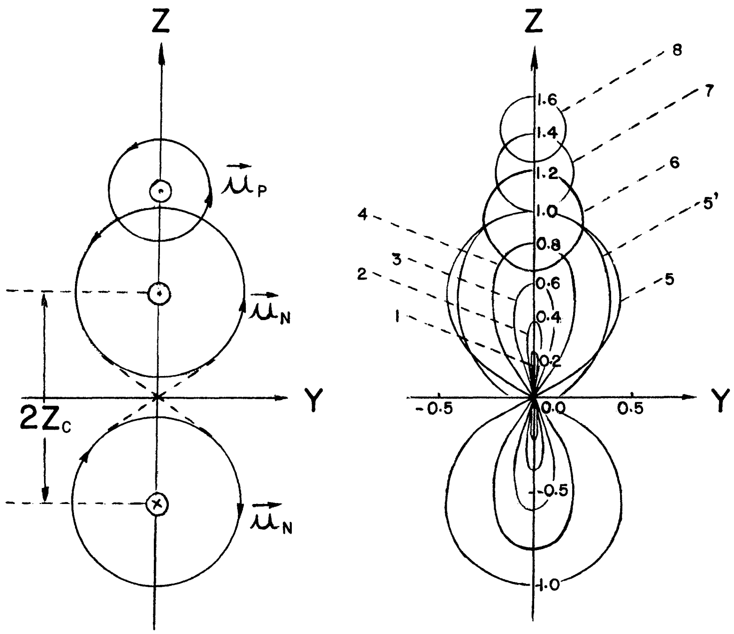


Fig. 3.

results are almost the same, except for

$$\gamma_- = \gamma_+ + \pi\,, \qquad v_{z-} = -v_{z+}\,, \qquad V_- = -V_+\,.$$

The physical quantities referring to protons and electrons are denoted by subscripts '+' and '−'.

2.4. The constants of motion

The time-average of the magnetic fields produced by a particle which moves along a large number of loops is equivalent to that produced by circular current. Hence, the particle motion is equivalent to a magnet with a magnetic moment μ. μ can be determined from the analytical form of the circular current, the area enclosed by the circular motion, and its time period. The '8'-shaped motion can be divided into two kinds 'circular' motion, equivalent to two magnets in opposite directions separated from each other by a distance equal to $2Z_c$ (Xu, 1981e, 1982b).

According to the analytical solution of the particle motion, the first approximations of the magnetic moments corresponding to the P and N particles are given by

$$\boldsymbol{\mu}_P = -\frac{\frac{1}{2}mv_0^2\sin^2\phi}{H_c^2}\mathbf{H}_c\,,$$

$$\boldsymbol{\mu}_N = -\frac{2e^2Z_c^2}{3\pi mc^2}\mathbf{H}_c\,;$$

where $\mathbf{H}_c = (2Z_c/a^2)\mathbf{e}_x$ is the magnetic field at the centre of the trajectory.

In P region μ_P is a constant of motion, and is the same as the magnetic moment derived from the guiding centre theory. In N region H_c is very small or equal to zero, but here we also find a magnetic moment, which is constant during the motion. Other constants of motion are the center of the trajectory Z_c, kinetic energy of the particle, and the velocity parallel to the magnetic field.

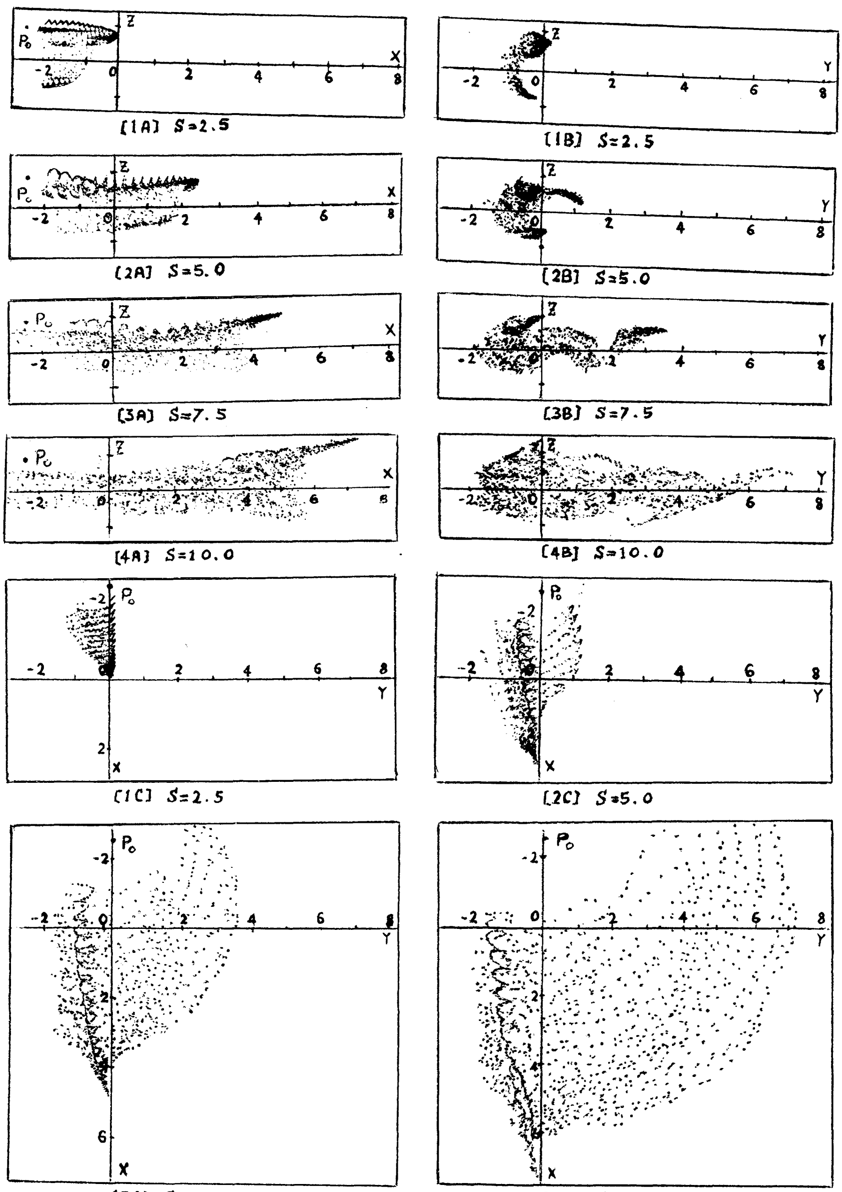


Fig. 4.

2.5. Propagation of a Point Source Plasma in the Magnetotail

The propagation of a point source plasma in the magnetotail is very complicated. Figure 4 is the spatial distribution of 856 particles released from a point source at P_0 (Figure 1), which are collected in a series of time, correspond to s = 2.5, 5.0, 7.5, 10.0, and 12.5, respectively. When $T = 0.2$ (Xu *et al.*, 1986c). Figure 4(a) is the spatial distribution projected on the meridian plane while Figure 4(b) is on the equatorial plane.

According to Figure 4(a) particles released from P_0 propagate toward the Earth (x-direction) inside the bounds of the magnetic field lines. The north-south asymmetry of the distribution is due to the different motion of P and N particles. P particles do not move through the neutral sheet and remain in the northern part of the magnetotail. The distribution of N particle is nearly symmetrical. Moreover, we find a double layer along the magnetic field line, caused by the latitude separation of P and N particles. This double layer structure is more obvious in the region $X > 5L$, which is near the Earth, when $s = 12.5$.

With regard to the propagation of particles projected on the equatorial plane (Figure 4(b)), particles move along an anticlockwise direction in the equatorial plane. They have both dawn-dusk and Earthward propagation. The Earthward plasma flow becomes more obvious when the dawn-dusk electric field is greater.

2.6. Spatial Distribution of Particles in the Magnetotail (Xu and Jin, 1984; Xu *et al.*, 1986a)

According to Figures 2 and 5, particles starting from a point source P_0 (Figure 1) will undergo a dawn-dusk and latitude separations. In order to study the spatial distribution of particles in the magnetotail, particles which start continuously from a point source P_0 are collected in a series of $X = X_n$ planes. Figure 5 shows the spatial distribution of these particles in $X_n = 14L$ plane, where Figures 5(a) and 5(b) are the spatial distributions along the dawn-dusk direction (Y-axis) and latitude (Z-axis). When $T = 0.2$, we have found that there is a dawn-dusk separation in the N region, the degree of which depends on the intensity of the electric field. These separations will undergo both acceleration and pitch angular scattering. Spatial distributions in different values of T will be reported elsewhere.

The spatial distribution of the particles along Z-axis (Figure 5(b)) shows that the particles separate into the northern and southern part of the magnetotail, when they leave the N region. In each part, the distribution has two peaks, which are more obvious when the electric field is high. Some of these particles with small pitch angle penetrate into the Earth's ionosphere, and have a latitude separation near the Earth, corresponds to these two peaks.

Actually, the particle source in the magnetotail is not a point source, but distributed uniformly in the neutral sheet. We assume that there are 44 particle sources P_i ($i = 1, 2, 3, \ldots, 43, 44$) distributed uniformly over the $X = -2.5L$ plane, as shown in Figure 1. The positions of these sources are $X_n = -2.5L$, $Y_n = nL$, $Z_n = \pm 1L$, respectively, where $n = 1, 2, 3, \ldots, 21, 22$. Again we suppose that the initial pitch angular

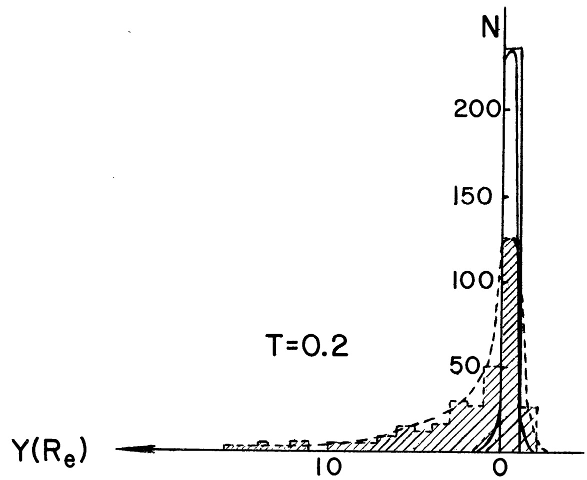


Fig. 5a.

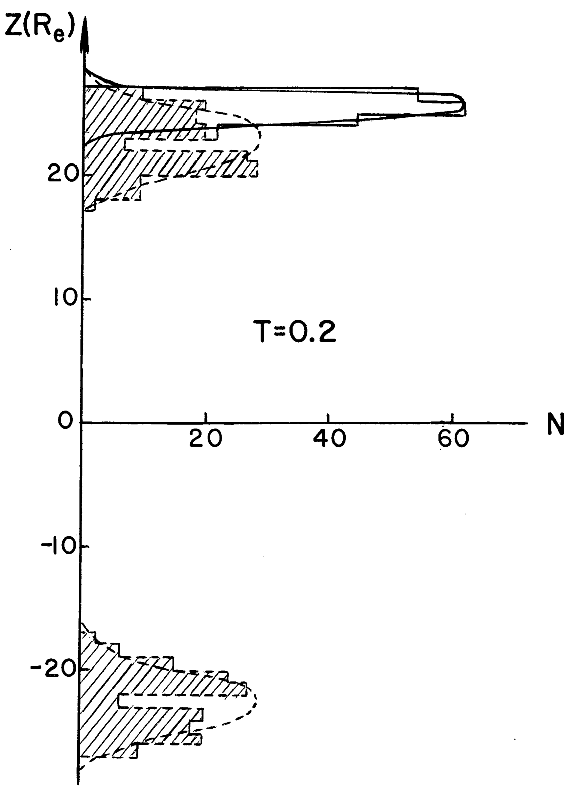


Fig. 5b.

distribution of all particle sources are isotropic. We collect all these particles which start from the plane source in the $X = 14L$ plane with different electric field intensities $T = 0.0$, 0.2, and 0.4. For low-energy particles, the spatial distribution along the dawn-dusk direction is uniform. But for higher energy ranges there is an obvious asymmetry along the dawn-dusk direction, and the degree of asymmetry increases with the energy. The trajectories of the negatively charged particles are similar to those of positively charged particles. So there is also a dawn-dusk asymmetry, but its direction is reverse.

Finally, bursts of charged ions and electron observed by IMP 7 in the magnetotail have been analysed. The ions are P1 ($Q \geq 6$, 37–55 keV/Q), P2 ($Q \geq 2$, 65–100 keV/Q), and P4 ($Q \geq 1$, 160–230 keV/Q). The energy range of electron is 200–360 keV. We have found that there is a dawn-dusk asymmetry in the spatial distribution of both electron and ion bursts, except that the ion bursts are observed more frequently in the dusk sector than in the dawn sector, whereas the asymmetry for the electron bursts is reversed. These results agree with the theoretical results.

2.7. Acceleration of Particles in the Magnetotail (Xu *et al.*, 1980; Xu, 1981b, 1986c; Xu and Gu, 1982)

According to Figure 2, particles with small initial pitch angle do not move through the N region. The guiding center of the particle motion moves along the magnetic field line-of-force toward the Earth, and drifts along the opposite direction of y-axis, owing to the magnetic gradient drift. When the initial pitch angle is large, the particles move through the N region. The motion along the y-axis in the N region is opposite to the gradient drift, and they move along the dawn-dusk electric field and are accelerated. The acceleration depends on the intensity of the electric field and the time interval of the particle in the N region, or the deviation along the dawn-dusk direction. Particles with small pitch angles do not move through the N region, so there is no acceleration. Medium initial pitch angle particles stay in the N region for a very short time, so that the acceleration is not obvious. When the initial pitch angles are large, they spend a long time in the N region so that the acceleration is obvious.

The acceleration process of particles starting from a point source at different times are shown in Figure 6, when $T = 0.2$. Each figure corresponds to $s = 2.5$, 5.0, 7.5, 10.0, and 12.5, respectively. The shaded regions represent the energy spectrum of P particles, whereas the dark histogram line represents the energy spectrum of N particles. We find that the mean energy spectrum of P and N particles decreases with time due to the gradient B drift. The two types of particles moving along the opposite directions of the dawn-dusk electric field. The deviation of the spectrum from its mean energy are due to the circular motion of the particles, where the positions of the particles oscillate along the direction of the electric field. For $s = 5.0$, some of N particles which move through and stay in the N region are accelerated. The acceleration depends on the time interval of the particles in the N region. When $s = 12.5$, the energies of a few of N particles become five times greater than their initial values.

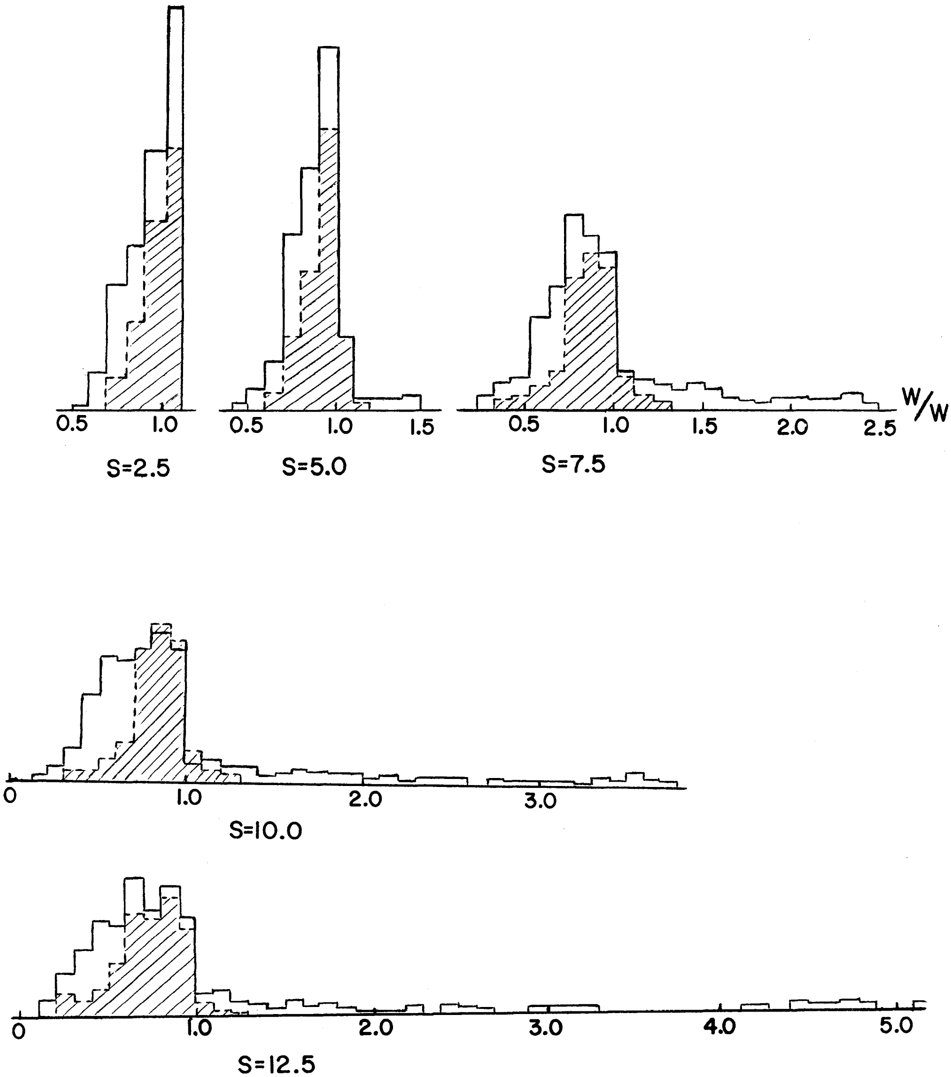


Fig. 6.

2.8. Pitch angular scattering of particles in the magnetotail (Xu and She-fen, 1983a; Xu *et al.*, 1986b, c)

According to Figure 2(a), the trajectories of particles with small pitch angles (ϕ = 10 and 20°) do not pass through the N region, satisfy the guiding center theory, and the variation of the pitch angle depends on the magnetic moment invariant. For particles which move through the N region, some of them are reflected many times in the N region, as shown in Figure 1(a) when ϕ = 40 and 70°. The magnetic moment is no longer constant. Especially when ϕ = 70°, when it leaves the N region, its pitch angle becomes very small. This means that there is a pitch angular scattering in the N region.

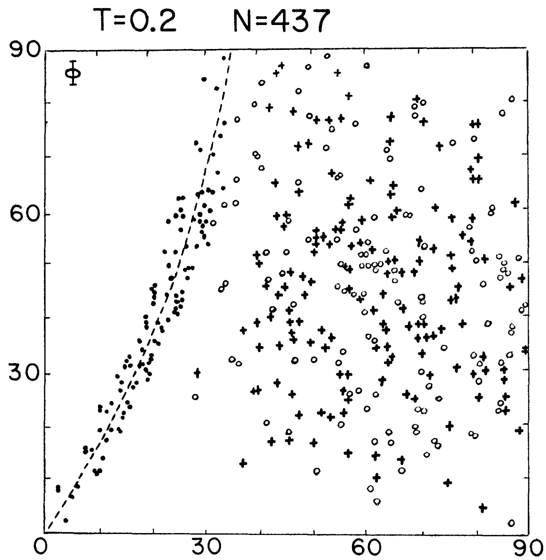


Fig. 7.

We have studied the correlation between the initial pitch angle ϕ_0 of an isotropic point source plasma and the pitch angle ϕ collected in $X = 14L$ plane for different electric field intensities T = 0.0, 0.2, and 0.4, respectively. Figure 7 is the correlation between ϕ and ϕ_0, where the dashed line denotes the relation between ϕ and ϕ_0, derived from the μ invariant. Particles near the dashed line are denoted by 'o' are P particle. N particle is denoted by '•'. Due to the pitch angular scattering of N particle in the N region, the relation between ϕ_0 and ϕ of N particles is a random one, some particles with large ϕ_0 have small ϕ when they reach $X = 14L$ plane. More than 48% of the pitch angles become smaller. When the electric field is absent (T = 0.0), particles in $X = 14L$ plane are almost P particles, and the relation between ϕ_0 and ϕ satisfies the μ invariant. In a large electric field (T = 0.4), most of the particles are N particles, and the scattering process is more obvious.

3. Currents in the Magnetotail

3.1. Magnetotail current sheet

The magnetotail current sheet can be inferred from the deviation of measured magnetic fields from the fields predicated from magnetic field model. However, since the magnetic field measured by the satellites vary with time and space, it is rather difficult to calculate the current source of the magnetic data obtain on one satellite. Here we calculated the current source of the magnetic field by means of the curl of magnetic fields observed on one satellite with a simple model (Xu *et al.*, 1984). The data we used are from the magnetometer of University of California at Los Angeles on the ISEE-1 satellite.

In the magnetotail coordinate (X_{mt}, Y_{mt}, Z_{mt}), where X_{mt} and Y_{mt} are the same as in the magnetospheric coordinate (Xu *et al.*, 1983a, b), and $X_{mt}\,Y_{mt}$ plane in the magnetotail coincides with the neutral sheet, B_x is much larger than B_y and B_z, and the variation of B_x is mainly in the Z_{mt}-direction. That is, the magnetic field can be approximated by

$$\mathbf{B} = B_x(z)\,\mathbf{e}_x\,, \tag{2.1}$$

and the current density is given by

$$\mathbf{J} = \frac{c}{4\pi}\nabla\mathbf{B} = \frac{c}{4\pi}\left[\frac{B_x(Z+\Delta Z) - B_x(Z)}{\Delta Z}\right]\mathbf{e}_y\,. \tag{2.2}$$

These equations can only be used in the magnetotail region, which is far away from the earth. In this region the dipole field B_p can be omitted. Near the Earth, B_p is comparable to or larger than the magnetic field of the magnetotail, so that the total magnetic field does not satisfy Equation (2.1), and B_p must be removed from the data, here we supposed that B_p is a tilted dipole field.

The fine structures of a magnetic field data used for the analysis were reduced by $\frac{5}{7}$ hour running average. Reorganized B_x as a function of Z_{mt}, we have Figure 8, which corresponds to the data on March 7, 1978. The variation of B_x with Z_{mt} shows that the magnetotail has a form of magnetic sheet with a dimension L_s about $3R_e$. Outside the

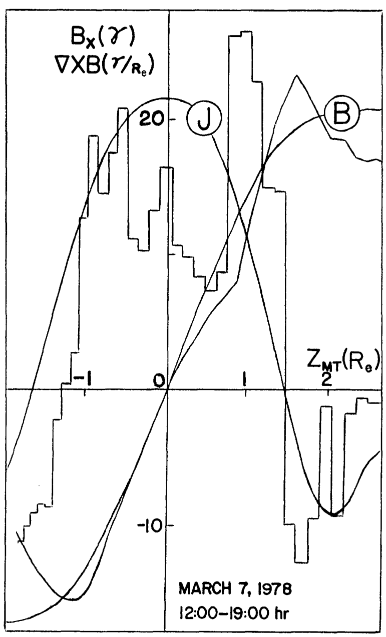


Fig. 8.

sheet, where $|Z_{mt}| > L_s/2$, the magnetic field is nearly constant, but in the sheet, where $|Z_{mt}| < L_s/2$, it is similar to a linear field which is equal to zero near $Z_{mt} = 0$. The curl of B or the distribution of the current sheet density calculated from Equation (2.2) is shown as a histogram in Figure 3, where it is shown that the electric current is confined in the neutral sheet and has a valley near the centre of the sheet. The current decreases to zero near the boundary.

The character of the magnetic field and current in the neutral sheet of the magnetotail can also be studied theoretically by means of the particle motion (Xu, 1981e; 1982). The electric current of the particles moving in a magnetic field is given under adiabatic conditions by

$$\mathbf{J} = \pm e n_{\pm} \langle \mathbf{V}_{\pm} \rangle + c\nabla \times \mathbf{M}, \tag{2.3}$$

where n, $\langle V \rangle$, and M correspond to other density, average drift, and total magnetic moment of the trajectory center, respectively. The first term on the right-hand side of Equation (2.3) is the current due to the drift motion, and the second term is due to the circular motion. Since we have found the drift velocity and magnetic moment of the '8'-shaped motion, we extended the guiding center theory to the region where the magnetic field is zero. The quantities n, $\langle V \rangle$, and M can be determined by the distribution function, which is the solution of Vlasov equation. The standard procedure for obtaining the solution of a time-independent Vlasov equation is to express the distribution function f as a function of the constant of motion, so that f can be written as

$$f = n_0 \left(\frac{m}{2\pi KT} \right)^{1/2} \exp\left[-\lambda Z_c^2 - \frac{m}{2KT}(v_\parallel^2 + v_\perp^2) \right], \tag{2.4}$$

where K is the Boltzmann constant, T is the temperature of the particle, and λ is the parameter of f. According to the distribution function f, we can find M and $\langle V \rangle$, and by substituting then in Equation (2.3), we find the formula for the current sheet J. The variation of J with Z_c is shown in Figure 8 by the line J, when $\lambda = 5$, and $Z_a = L_s$. The magnetic field caused by this current can be evaluated from the current, and its variation with Z_c is shown by the line B.

The variation of B with Z_c shows that the magnetic field has a form of a magnetic sheet with a dimension of $2Z_s$. Outside the sheet where $|Z_c| > Z_s$, the magnetic field is similar to a uniform one. But in the sheet where $|Z_c| < Z_s$, it is similar to a linear field, which is equal to zero at $Z_c = 0$.

Since the total current in the N region is mainly caused by the 'reverse drift' current, due to the '8'-shaped motion, we find that there is only a slight deviation on J and B for different density distribution with different value of λ ($\lambda = 0$, 1, and 5). We have also studied the relation between charged particle motions and the magnetic field, and derived a self-consistent magnetic field.

Finally, the variations of electric current and magnetic field calculated from the particle motion in a neutral magnetic sheet are in agreement with the results obtained by analysing the data observed by ISEE-1 in the magnetotail on March 7, 1978. Here we supposed that $Z_s = Z_a$, as shown in Figure 8.

3.2. The field-aligned current

The field-aligned current of the Birkeland current during substorms has been confirmed by many recent observations and studies (papers and references in Akasofu, 1980; Corovillano and Forbes, 1982). According to the regular pattern of charged particles moving in the neutral magnetic field, we have derived an expression for a three-dimensional function of field-aligned current, which is related to the scattering process and spatial distribution of charged particles in the magnetotail (Xu, 1981a; Xu and Gu, 1982a; Xu *et al.*, 1986b).

The same model as in Figure 1 has been used. In the region far away from the Earth, the magnetic field of the magnetotail can be assumed to be a neutral line magnetic field; near the Earth, it is a dipole plus a neutral sheet magnetic field. The common boundary of the magnetic field in these two region is at the $X = 14L$ plane. The source of the field-aligned current is supposed to be a plane source ($X = 2.5L$) as shown in Figure 1. This plane source contains 44 point sources of plasma, and corresponds to a uniform Earthward plasma flow in the neutral sheet during substorms. According to the spatial distribution of particles in $X = 14L$, we can find the field-aligned current in the night side of the Earth (Xu *et al.*, 1986b).

The field-aligned current can be written as

$$J_{\|i} = +ef_{yi}(Y)f_{zi}(Z)\int_0^v \int_0^{\phi_c} f_{vi}(v)f_{\phi i}(\phi)\,\mathrm{d}\phi\,\mathrm{d}v\,, \tag{2.5}$$

where parameters with a subscript i ($i = P$ and N) correspond to the parameters of P and N particle, and ϕ_c is the critical pitch angle of the particle in $X = 14L$ plane. In the $X = 14L$ plane, particles with pitch angles larger than ϕ_c will be reflected by the magnetic mirror before they reach the ionosphere, but will enter into the ionosphere and form a field-aligned current when $\phi < \phi_c$. In Equation (2.5) $f_y(Y)$, $f_z(Z)$, $f_v(v)$, and $f_\phi(\phi)$ are the spatial distribution function, the velocity distribution function, and the pitch angular distribution function of particles on the $X = 14L$ plane. We can find them from the results of the computer simulation by a functional approximation. The pitch angular distribution function can be found from the correlation between the initial pitch angle ϕ_0 of the plasma source and the pitch angle ϕ collected on $X = 14L$. When $T = 0.2$, the correlation between ϕ_0 and ϕ is shown in Figure 7. According to Figure 7, we find the distribution of ϕ_0 and ϕ. Since the source is isotropic, the initial distribution can be written as

$$f_{\phi_0}(\phi_0) = N_0 \sin\phi_0\,, \tag{2.6}$$

where the angular interval of the angular distribution is $5°$ and $N_0 = 64$.

For P particles, $f_{\phi P}$ can be derived from the μ invariant and is given by

$$f_{\phi P}(\phi) = N_0\left(\frac{v}{v_0}\right)\left(\frac{B_0}{B}\right)^{1/2}\frac{\cos\phi\sin\phi}{\sqrt{1-\left(\frac{v}{v_0}\right)^2\left(\frac{B_0}{B}\right)\sin^2\phi}}\,, \tag{2.7}$$

with

$$\phi < \phi_m\,,$$

where ϕ_m is the critical pitch angle of particles which can reach $X = 14L$ plane. For different electric field intensity, $T = 0.0$, 0.2, and 0.4, ϕ_m is equal to 90, 65, and 0°, respectively. This distribution function is in agreement with the computer results.

The approximating functions of the pitch angular distribution function of N particles, the velocity distribution function, and the spatial distribution function calculated from the computer can be assumed as

$$f_A(A) = a(A - A_0)^c \exp[-b(A - A_0)^n]\,, \tag{2.8}$$

where $A = \phi$, v, Y, and Z for the different distribution functions and the coefficients a, b, c, n, and A_0 are different and can be determined from the computer results. Substituting Equations (2.7) and (2.8) to Equation (2.5), and projecting them to the Earth along the magnetic field line, we can derive the expression for a three-dimensional function of the field-aligned current system near the polar cap. The outflow current is due to the electrons precipitating into the polar cap (Xu *et al.*, 1986b), and the inflow current due to protons. Due to the dawn-dusk and altitude separation of the particles in $X = 14L$ plane, electrons penetrate in the dusk sector with higher latitude, and in the dawn sector with lower latitude. But protons are reverse. They penetrate in the dawn sector with higher latitude, and in dusk sector with lower latitude. This complex current structure agrees with the observational results.

3.3. Filamentary current (Li *et al.*, 1986)

Filamentary structures are often observed in space plasma. There seems to be a continuous transition from the filamentary structures to sheet structures. The filamentary structure of the magnetotail has been studied using particles bursts and magnetic field disturbance observed on ISEE-1 in the magnetotail during the time January 25 to May 1, 1978. The particle data are from Ultra-Low Energy Wide Angle Telescope (ULEWAT) of University of Maryland/Max-Planck-Institute. They are electrons for four different ranges (75–115, 115–300, 300–400, 440–1300 keV), protons (170–400 keV), alpha particles (120–2000 keV), and ions ($Z \geq 1$, 120–200 keV). The magnetic field data are from the magnetometer of the University of California, Los Angeles.

As a natural consequence, most of the particle bursts concentrate within the neutral sheet, but during the time January 25 to May 1, 1978, we have found 45 examples of isolated bursts in the magnetotail lobe, all of them correlate well with the magnetic field

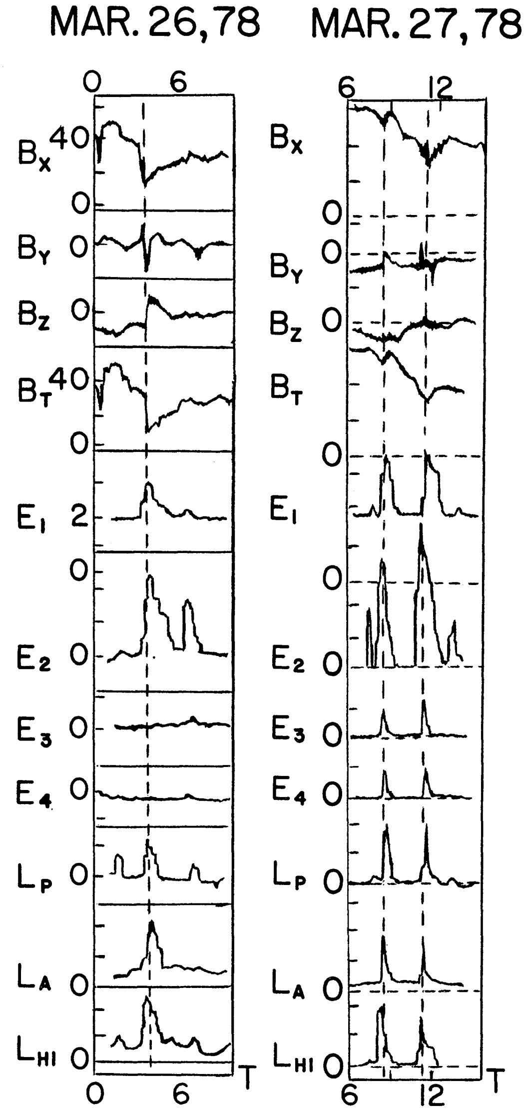


Fig. 9.

disturbance. Two typical examples of these events are shown in Figure 9, where B_x, B_y, B_z, and B_t correspond to the components of the magnetic field along x-, y-, z-axis, and the total magnetic field. E_1, E_2, E_3, and E_4 are log (count rate) $(1/s)$ of electrons correspond to an energy ranges of 75–115 keV, 115–300 keV, 300–400 keV, and 440–1300 keV. LP, LA, and LH correspond to protons (170–400 keV), alpha particles (120–2000 keV), and ions ($Z \geq 1$, 120–200 keV). The morphology of B_x, B_y, and B_z disturbances show that these particle bursts can be explained as a filamentary current of the magnetotail lobe.

The filamentary current in the magnetotail lobe has the following characters:

(1) Most of the directions of these currents are along the magnetic field.

(2) The scale of these currents can determined from the time scale of the particles bursts, and its magnitude is about 10^3 km.

(3) The composition of the filamentary current is mostly electrons, and also some protons, alpha particles, and ions.

Finally, the morphologies of the magnetic field observed from a satellite passing through a filamentary current model has also been studied. The trajectories of the satellite have different directions and distances from the center line of the filamentary current. We have found some trajectories in which the morphology of the magnetic field is comparable with the magnetic disturbance observed on ISEE-1.

4. Summary

The main results of this paper can be summarized as follows:

(1) The neutral sheet of the magnetotail can be divided into a perturbation (P) region and nonperturbation (N) region. In the N region, the Alfvén's perturbation method cannot be used, since the magnetic field is very small or equal to zero. According to the analytical solution of the particle's motion in N region, the motion can also be divided into closed oscillatory motion around its guiding center and the drift of the guiding center. Difference with particle's motion in P region, the closed oscillatory motion of particle in N region is not a circular motion, but is an '8'-shaped motion. The '8'-shaped motion is equivalent to two magnets in opposite direction, separate from each other. We also find expression of the magnetic moment and other constant of motion of N particle.

(2) We have studied the propagation of a point source plasma in the magnetotail, by means of a computer simulation using test particles. We find that particles propagate toward the Earth inside the bounds of magnetic lines. The propagation of the particles projected on the equatorial plane shows that there are both dawn-dusk and Earthward propagation. The Earthward plasma flow become more obvious when the dawn-dusk electric field is large during the substorm. Due to the influence of the space charges, some of the equi-electric potential planes deviate, and become convex planes. The peak of the convex planes propagate toward the Earth, and decrease with time. The deviation depend on the density of the particle source.

(3) The spatial distribution of particles in the magnetotail, starting from a uniform plane source of plasma in the neutral sheet shows that, for low-energy particles, the spatial distribution along the dawn-dusk direction is uniform. But for higher energy range there is an obvious asymmetry along the dawn-dusk direction. The degree of the asymmetry increases with the energy. The motions of negatively charged particles are similar to those of the positively charged particles; there is also a dawn-dusk asymmetry, but the direction is reverse. These results simulate the observation of ions and electron bursts on IMP-7 and -8 in the magnetotail.

(4) We determined the current sheet of the magnetotail by means of calculating the curl of the magnetic field data observed on ISEE-1. We found that the current is

confined in the neutral sheet, the thickness of the current sheet is about 3 Earth radii, and the current density is about $1.5 \times 10^{-2}\,\mathrm{km}^{-2}$.

The electric current in the neutral sheat is the summation of the current due to the drift motion and the closed oscillation motion. We find the analytical expression of the current by means of the solution of a time-independent Vlasov equation, in which we can express the distribution function of the plasma as a function of the constant of motion. Moreover, we can calculated the magnetic field caused by this current, and find that the particle motion and the magnetic field is self-consistent.

The theoretical results are in agreement with the results obtained by analysing the data observed by ISEE-1 in the magnetotail.

(5) According to the regular pattern of the charged particles moving in the neutral magnetic field, we have derived an expression for a three-dimensional field-aligned current, which is related to the scattering process and spatial distribution of charged particles in the magnetotail.

(6) We found that most of the isolated particle bursts in the magnetotail lobe correlated well with the magnetic disturbances. The morphology of the magnetic disturbances shows that these particle bursts can be explained as a filamentary current along the magnetic field.

Acknowledgements

This paper is a summary of results obtained by a number of my colleagues in the research group at the Institute of Space Physics, Beijing, and by students of the University of Sciences and Technology, Hefei. I wish to thank Wan Zhen-fu, Ma Fu-sheng, Gu She-fen, Zheng Liu, Wu Zhen, Jin Zhong-hai, Wu Ji-ping, Li Ya-qi, and Zhang Dong-mei for their contributions.

I an greatly indebted to Professors H. Alfvén, D. A. Mendis, and Dr E. C. Whipple of the University of California at San Diego, and Professor C. Y. Fan at the University of Arizona, who invited me to join their groups in the U.S.A. and let me have the benefit of inspiring discussions and advice.

I thank Professor C. T. Russell of the University of California at Los Angeles for providing the magnetic data, and Professor G. Gloeckler of the University of Maryland and Professor D. Hovestadt of Max-Planck-Institut für Physik and Astrophysik, F.R.G., for providing the particle data.

Finally, I am grateful to Li Zai-kun and Sie Liu-xiang for helping me with the manuscript.

References

Akasofu, S.-I.: 1980, *Dynamic of the Magnetosphere*, D. Reidel Publ. Co., Dordrecht, Holland.

Alfvén, H.: 1950, *Cosmical Electrodynamics*, Oxford University Press, New York.

Alfvén, H.: 1981, *Cosmic Plasma*, D. Reidel Publ. Co., Dordrecht, Holland.

Corovillano, R. L. and Forbes, J. M.: 1982, *Solar-Terrestrial Physics*, D. Reidel Publ. Co., Dordrecht, Holland.

Krimigis, S. M., Haerendel, G., McEntire, R. W., Paschmann, G., and Bryant, D. A.: 1982, *EOS* **62**, 843.

Li, Ya-qi, Xu, R.-L., and Fan, C. Y.: 1986, 'The Filamentary Current in the Magnetotail Lobe Observed on ISEE-1', *Proceedings of the International Symposium on Space Physics*, November 10–14, Beijing, China, pp. 4–63.

Lyons, R. R. and William, D. J.: 1984, *Quantitative Aspects of Magnetospheric Physics*, D. Reidel Publ. Co., Dordrecht, Holland.

Nishida, A.: 1982, *Magnetospheric Plasma Physics*, D. Reidel Publ. Co., Dordrecht, Holland.

Potemra, T. A.: 1984, *Magnetospheric Currents*, Geophysical Monograph 28, American Geophysical Union, Washington DC.

Xu, R.-L., Wan, Zhen-fu, and Ma, Fu-Sheng: 1980, *Acta Geophys. Sinica* **23**, 233 (in Chinese).

Xu, R.-L.: 1981a, in K. Knott (ed.), *Adv. Space Res. Phys. Planetary Magnetosphere* **1**, 197.

Xu, R.-L.: 1981b, in K. Knott (ed.), *Adv. Space Res. Phys. Planetary Magnetosphere* **1**, 283.

Xu, R.-L.: 1981c, *Chinese J. Space Sci.* **1**, 2 (in Chinese).

Xu, R.-L.: 1981d, 'The Analytical Trajextory of the Charged Particle Moving in a Neutral Sheet Magnetic Field', *Proceeding of an International School and Workshop on Plasma Astrophysics*, Verona Como, Italy, pp. 421.

Xu, R.-L.: 1981e, 'A Plasma Current Sheet Studied by Means of Particle Motion', *Proceeding of an International School and Workshop on Plasma Astrophysica*, Verona, Como, Italy, pp. 417.

Xu, R.-L.: 1982, *Chinese J. Space Sci.* **2**, 23 (in Chinese).

Xu, R.-L. and Gu, She-fen: 1982, *Sci. Sinica* **A24**, 644.

Xu, R.-L. and Gu, She-fen: 1983, *Chinese J. Space Sci.* **3**, 23 (in Chinese).

Xu, R.-L. and Zhong-hai, Jin: 1984, 'The Dawn-Dusk Asymmetry of Particles Near the Neutral Sheet of the Magnetotail', *Proc. Conf. Achievements of the IMS*, ESA SP-217, p. 209.

Xu, R. L., Fan, C. Y., Gloeckler, G., and Hovestadt, D.: 1983a, *The Spatial Distribution of Energetic Ions in the Magnetotail*, Chapman Conference on Magnetospheric Currents, Washington DC, April 5–8.

Xu, R.-L., Fan, C. Y., Gloeckler, G., Ma-Sung, L. S., and Hovestadt, D.: 1983b, 'Acceleration of Particles in the Magnetotail', *18th International Cosmic-Ray Conference*, Bangalore, India, MC 5–2.

Xu, R.-L., Zheng, Liu, Wu, Zhen, and Fan, C. Y.: 1984, 'Magnetotail Current Sheet and Energetic Particles Observed on ISEE-1', *Proc. Conf. Achievements on the IMS*, ESA SP-217, p. 167.

Xu, R.-L., Fan, C. Y., Gloeckler, G., and Hovestadt, D.: 1986a, *Planetary Space Sci.* **35**, 57.

Xu, R.-L., Xie, Liu-xiang, and Jin, Zhong-hai: 1986b, *Scattering of Charged Particles in the Magnetotail and the Field-Aligned Current*, presented at the International Symposium on Solar Terrestrial Physics June 30–July 5, Toulouse (to be published).

Xu, R.-L., Zhang, D.-M., and Cao, X.-D.: 1986c, 'The Propagation of a Point Source Plasma in the Neutral Magnetic Field', *Proceedings of the International Symposium of Space Physics*, November 10–14, Beijing, China.

HYDROMAGNETIC WAVES ASSOCIATED WITH POSSIBLE FLUX TRANSFER EVENTS*

L. J. LANZEROTTI and C. G. MACLENNAN
AT&T Bell Laboratories, Murray Hill, New Jersey, U.S.A.

(Received 22 June, 1987)

Abstract. Magnetic fluctuations in the hydromagnetic frequency band ~ 0 to ~ 0.05 Hz are examined at magnetospheric cusp latitudes during two times when ionospheric signatures of possible flux-transfer events were evident in the data. Ultralow frequency power is found to be very broad band in the range ~ 0.02–0.05 Hz and to be more narrowly confined at a frequency ~ 0.0025 Hz. At lower latitudes, the higher frequency (broad-band) power excites narrower-band field line resonances at the fundamental frequency of the respective field line – a standing Alfvén wave. The narrow-band power in the lower frequency band (period around 400 s) is approximately that expected for a field line resonance on a closed field line near the magnetopause; it also corresponds approximately to the width of the convected field-aligned current filament as observed on the ground. The reconnection process at the dayside magnetopause evidently plays an important role in the generation of low-frequency ($\lesssim 0.008$ Hz) hydromagnetic energy in the dayside magnetosphere, energy which can produce Alfvén waves deeper in the magnetosphere.

1. Introduction

The fundamental plasma physical foundation for explaining small amplitude ($\Delta B \ll B_0$) oscillations ('pulsations'; 'micropulsations') of the geomagnetic field with frequencies in the range $\sim 10^{-3}$ Hz to ~ 0.1 Hz is contained in the papers of Hannes Alfvén in the 1940's (Alfvén, 1942a, b, 1947) which discussed the existence of 'electromagnetic-hydromagnetic' waves in an electrically conducting liquid permeated by a background magnetic field. Alfvén was particularly interested in the possible ability of such waves to heat the solar corona. Dungey (1954, 1955) recognized the importance of Alfvén's concepts for explaining geomagnetic pulsations. He pursued detailed examinations of the equations derived by the combination of Maxwell's equations and the hydrodynamic equation, recognized the severe problems that arise in attempts at generalized solutions for a dipole geometry, and derived special wave solutions in a cylindrical geometry.

The work of Sugiura (1961a, b) demonstrated that the characteristics of the magnetic oscillations were indeed wave-like (there were distinct polarizations observed between the vector components of the magnetic field fluctuations as detected on the Earth's surface). Studies of these oscillations at conjugate points on the ground indicated clearly the global nature of the phenomena (e.g., Sugiura and Wilson, 1964; Nagata *et al.*, 1963). The polarization and amplitude patterns suggested that the magnetic field lines were oscillating like guitar strings, as would be expected from an Alfvén wave. However, the reason why a particular magnetic shell would oscillate, uncoupled from a nearby one in the highly non-uniform plasma environment of the magnetosphere, was quite unclear.

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Much statistical morphological work has been carried out in attempts to clarify the occurrence patterns of these pulsations and their associations with geomagnetic activity, local time and season, and interplanetary conditions (e.g., see early reviews by Troitskaya, 1967; Saito, 1969; Jacobs, 1970; Troitskaya and Gul'elmi, 1967; Campbell, 1967).

Observations of highly localized (spatially) pulsation amplitudes in the auroral zones (Samson *et al.*, 1971) and sub-auroral regions (Fukunishi and Lanzerotti, 1974a, b) and of systematic polarization patterns in association with the amplitude peaks, together with theoretical derivations of hydromagnetic wave resonances by Southwood (1974) and by Chen and Hasegawa (1974a, b), provided the experimental and theoretical justifications for understanding how individual magnetic 'field lines' might oscillate 'independently'. The driving energy in the theoretical cases considered was taken to be the solar wind in its interaction with the Earth's magnetic field to form the magnetopause and the upstream bow shock.

A number of mechanisms have been examined observationally and theoretically as energy sources at the dayside magnetopause for producing 'exogenically-driven' hydromagnetic waves. One of these is the excitation of surface waves on the magnetopause produced by a shear flow (Kelvin–Helmholtz-type) instability. The solar wind flow interaction with the Earth's magnetic field which forms the magnetosphere produces the shear between two plasma regions (the theoretical literature is extensive; see, e.g., Southwood (1968) and Yumoto and Saito (1980)). Another important mechanism is the direct transmission through the magnetopause of hydromagnetic waves and turbulence generated in the magnetosheath and at the bow shock (again the literature is extensive; see, e.g., Russell and Hoppe (1983) and Barnes (1970)). The problem of transmission across the magnetopause is a key one for this latter mechanism, and it has received much theoretical, and some observational, attention (e.g., Greenstadt *et al.*, 1983; McKenzie, 1970; Wolfe and Kaufmann, 1975; Hassan, 1978; Kwok and Lee, 1984). Possible observational tests of the relative importance of these pulsation driving mechanisms form an important segment of the 'pulsation literature' at present; reviews of some elements of these tests are contained in, for example, Yumoto (1986), Verö (1986), and Wolfe *et al.* (1987).

Energy also appears capable of being transferred into the dayside magnetosphere via the occurrence of flux transfer events (FTE) – sporadic intervals of magnetic field reconnection at the magnetopause under appropriate interplanetary magnetic field conditions (e.g., Russell and Elphic, 1978; Haerendel *et al.*, 1978; Paschmann *et al.*, 1982). Recent work has provided some evidence for the signature of possible FTE's in the polar ionosphere (e.g., Goertz *et al.*, 1985; Todd *et al.*, 1986; Sandholt *et al.*, 1986).

Lanzerotti *et al.* (1986), in discussing the ground-based magnetic signature of possible FTE's, noted that 'intense magnetic activity in the Pc3–Pc4 ($\sim$ 30–60 s) bands' seems to occur during the intervals when the events are also seen. They also noted that such waves have been attributed by Bolshakova and Troitskaya (1984) to the location of the cusp. Russell and Elphic (1979) and more recently Southwood (1987) proposed that FTE's might generate long-period geomagnetic pulsations in the outer magnetosphere.

Glassmeier *et al.* (1984) provided strong support for the idea that FTE's might also be an energy source for Pc5 pulsations (period 150–600 s). In a statistical study of visual examinations of low resolution chart recordings of magnetic field data gathered at synchronous orbit, Gillis *et al.* (1987) concluded that many pulsations in the period range 60–120 s could be associated with FTE's.

This paper presents the results of an analysis of some hydromagnetic wave activity observed at cusp latitudes during a day when FTE-like field-aligned current signatures were observed in the ground magnetic records at South Pole Station in the Antarctic. The intent is a beginning towards an understanding of the possible relationships between FTE's and hydromagnetic waves at cusp latitudes in order to gain information about hydromagnetic energy sources in the magnetosphere.

2. Experiments

The South Pole magnetic field data used in this paper were acquired in 1982 as one element of the cusp science program at South Pole Station, Antarctica (Lanzerotti *et al.*, 1982). The changes in the local magnetic field were measured by an orthogonal, three-axis flux-gate magnetometer of the Trigg *et al.* (1971) type. The noise level of the instrument is ~ 0.2 nT; the digitization increment is 0.06 nT. All three magnetic field components (geomagnetic north-south, H-component; east-west, D-component; vertical, V-component) are multiplexed and sampled at one second intervals and are written in computer-compatible format on magnetic type. The instrument at Siple Station, Antarctica, (SI; $L \sim 4.1$) is similar, with the analog components sampled at 2 s intervals. Local time at South Pole $\approx$ UT − 3.5 hr; at Siple $\approx$ UT − 5 hr.

Also included are data from a low-latitude ($L \sim 2.9$) magnetic station in Plano, Illinois (PL). These data were acquired at 2 s intervals with a low power, portable magnetometer and data acquisition system (Medford *et al.*, 1981). The noise level of the instrument is ~ 0.1 nT. Local time at Plano is $\approx$ UT − 6 hr. The locations of all of the stations are listed in Table I.

Plotted in local time coordinates in Figure 1 are the locations of South Pole Station (SP) and several other Antarctic research stations. Shown by the shaded region is the nominal auroral oval under average geomagnetic disturbances (Akasofu, 1968). As

TABLE I

Stations used in this study

	Geographic		Approximate L
	Latitude	Longitude	
South Pole	−90.0	–	13.5
Siple	−76.00	276.00	4.1
Plano	41.66	271.57	2.9

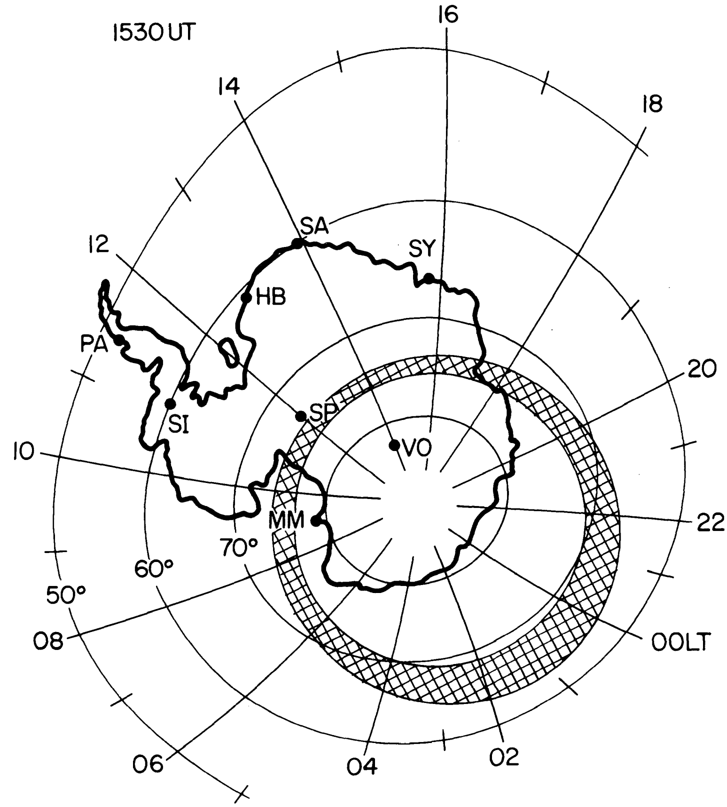


Fig. 1. Location of the nominal auroral zone with respect to the Antarctic continent at local magnetic noon at South Pole Station (SP). Other Antarctic stations are also shown: Halley Bay (HB), Siple (SI), SANAE (SA), Syowa (SY), McMurdo (MM), and Vostok (VO).

indicated by the figure, during local daytime conditions South Pole can be close to the dayside auroral region and, depending upon geomagnetic conditions, can be either on closed or open magnetic field lines; that is, on field lines connected inside the magnetosphere or connected to polar cap field lines extending into the magnetotail.

3. Observations

Magnetic field observations from South Pole for a portion of 3 June, 1982 are shown in Figure 2 (adapted from Lanzerotti *et al.*, 1986). The upper panel plots eight hours of data (08–16 UT) while the lower panel contains plots of an expanded half hour interval, 12 : 00–12 : 30 UT. The upper panel shows evidence for several large, impulsive-like magnetic field changes (particularly in the *D*-component) which may be ionospheric signatures of FTE's. As Lanzerotti *et al.* (1980) show, the signature of a convected field-aligned current in any magnetic field component significantly depends upon the location of the current relative to the observer. The impulsive signatures are super-

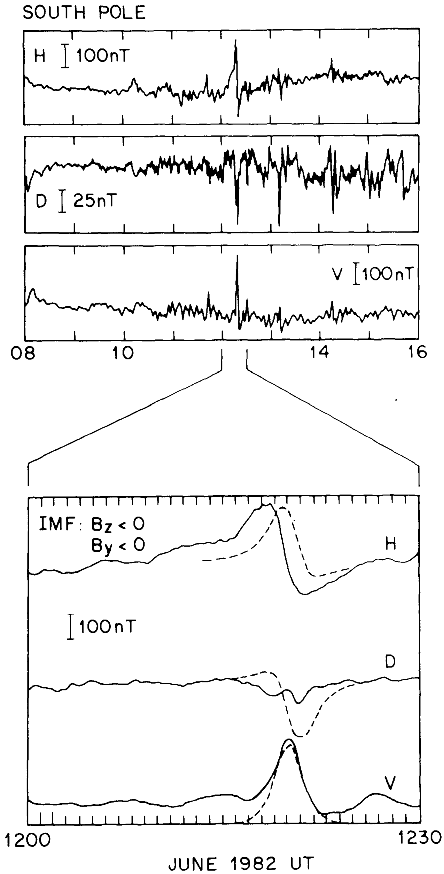


Fig. 2. South Pole magnetometer data acquired on 3 June, 1982. The upper panel plots eight hours of data for the three components in geomagnetic coordinates: H (north-south), D (east-west), and V (vertical). At South Pole upward deflections of the traces correspond to increases in the northward and eastward field components. Upward deflection of the V trace corresponds to an increase in the field looking along the field line (i.e., toward the ionosphere). The lower panel plots higher resolution data for a 30-min interval. The dashed lines correspond to comparisons of the data with a field-aligned current convected over the measuring location. Data adopted from Lanzerotti *et al.* (1986).

imposed on a background of magnetic fluctuations, particularly beginning after $\sim 10:30$ UT. We take the occurrence of the FTE-like signatures to indicate that the South Pole station is located close to the magnetopause, near the last closed magnetic field lines.

The dashed lines in the lower panel are comparisons of the magnetic field variations expected from a field-aligned current being convected in the ionosphere across the station location to the actual measured data (solid lines). The magnitude of the field-aligned current, deduced from the magnetic variations, is $\sim 2 \times 10^5$ A (Lanzerotti *et al.*, 1986). The interplanetary magnetic field (IMF) directions in the north-south (B_Z) and azimuthal (B_Y) directions are indicated.

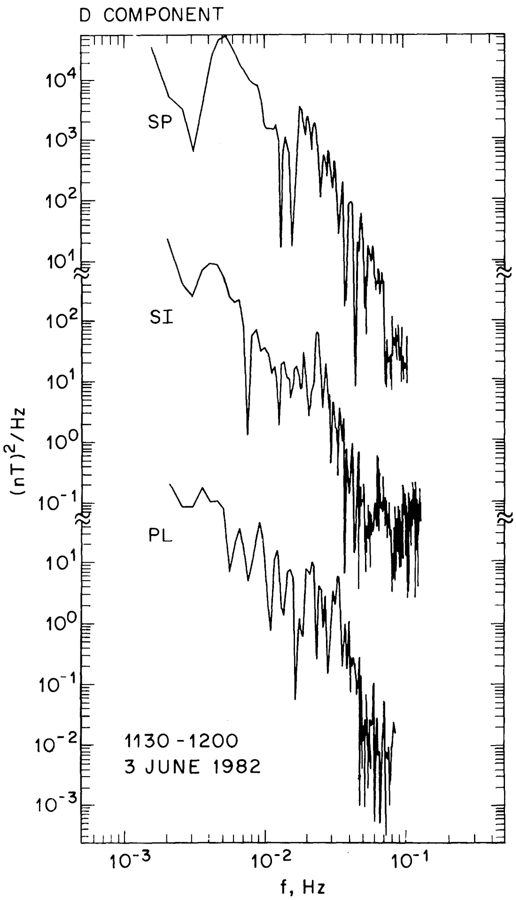


Fig. 3. Power spectra of half-hour of data (11 : 30–12 : 00 UT) from SP, SI, and Plano (Illinois), a northern-hemisphere station at $L \sim 3$ and ~ 1 hour west of the South Pole/Siple meridian.

Representative half-hour power spectra of the east-west (D) component of the magnetic field from the three magnetic stations are shown in Figure 3 for the interval 11 : 30–12 : 00 UT ($\sim$ 08 : 00–08 : 30 LT at South Pole). The spectra were calculated with a fast Fourier transform algorithm after first treating the data with a Thomson window (a single prolate spheroidal data window) (Thomson *et al.*, 1975). The spectral power at the high-latitude, near-cusp station SP is appreciably higher than at the mid-latitude stations; the slope of the spectrum is also steeper. A large, low-frequency peak, centered at $\sim$ 0.005 Hz, is also seen at this station. The higher frequencies are relatively less structured. At the two lower-latitude stations, spectral enhancements are observed at $\sim$ 0.025 Hz at SI and at $\sim$ 0.02 Hz and $\sim$ 0.035 Hz at PL at this time.

The gross spectral features appearing in Figure 3 – the low-frequency peak at the near-cusp station and high-frequency peaks at the lower latitude stations – are clearly evident in digital dynamic spectra of the data for essentially the entire interval contained

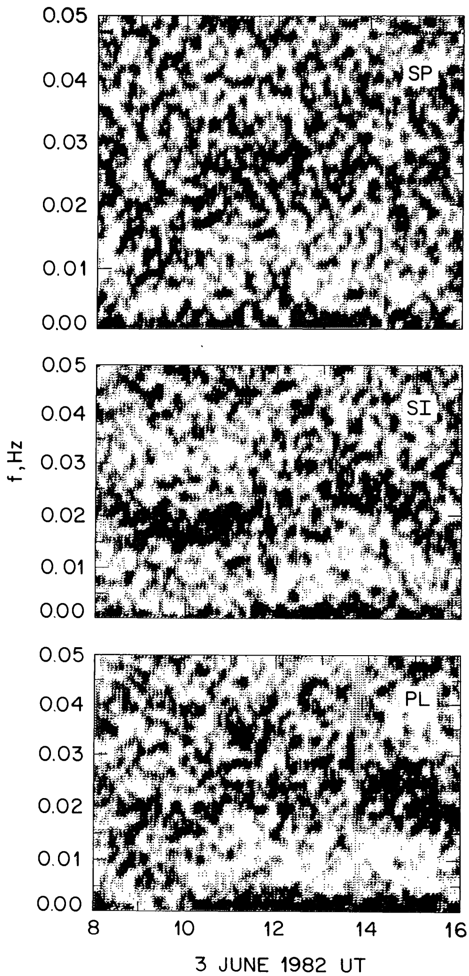


Fig. 4. Dynamic power spectra for eight hours of the D component magnetic variations from the three stations whose individual spectra are shown in Figure 3. The spectral range plotted compresses the lower frequency band at the bottom of the figure. Darker black represents larger amplitude.

in the upper panel of Figure 2. Figure 4 plots the dynamic spectra (0–0.05 Hz) for the interval 08–16 UT on 3 June, 1982 for the east-west (D) magnetic component at the three stations. These representations were produced by first calculating half-hourly power spectra and then subtracting a second-order least-squares 'background' fit. The interval of data was then slid by 5 min and the next spectrum and background subtraction was made. An 8-level gray shade was applied to each residual spectrum. The power at SP in the band $\sim$ 0.02–0.05 Hz is obviously very broad-banded.

In contrast, at SI, there is a strong enhancement in spectral power at $\sim$ 0.018 Hz ($\sim$ 55 s period) from $\sim$ 09 : 00–12 : 00 UT; between $\sim$ 13 : 00–15 : 00 UT the frequency increased to $\sim$ 0.025 Hz ($\sim$ 40 s period). The power is more broad-band at PL than at SI, although from ~11 : 00–16 : 00 UT the band centered between ~0.02–0.025 Hz is most enhanced. There is a significant absence of power at PL in the band ~0.0025–0.015 Hz (~66–400 s period).

Details of the dynamic spectra for the South Pole *H*-component in the band 0–0.02 Hz for 3 June are shown in Figure 5. The enhancements in the spectra over the binomial fit to the data are splotchy and broad. However, there is a rather 'steady' broad band of power centered at ~0.0025 Hz beginning at ~11 : 00 UT and continuing through hour 13. After ~14 : 00 UT, the frequency becomes more variable and decreases slowly with time.

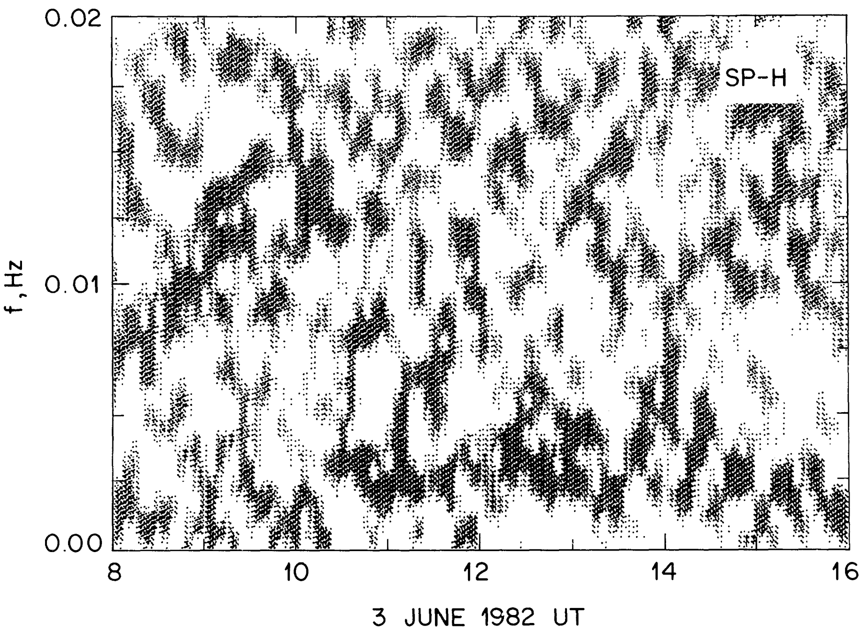


Fig. 5. Dynamic power spectra of *H*-component magnetic variations from South Pole data with an expanded low frequency scale to show the hydromagnetic wave activity at ~0.0025 Hz beginning at ~10 : 30 UT.

The north-south and vertical field components from SP are plotted in the two lower panels of Figure 6 for an eight hour interval on 16 October, 1985. The FTE-like signature at ~12 : 10 UT was discussed in Lanzerotti *et al.* (1987a); another such signature appears at ~14 : 20 UT. The Pc5 (~400 s period) event between 15 : 00–17 : 00 UT as detected by magnetometers and the Sondre Stromfjord incoherent scatter radar is discussed in Lanzerotti *et al.* (1987b). It is clear from the dynamic spectra in the upper panel that considerable power was present in the band ~0.0025–0.0070 Hz from ~12 : 00–14 : 00 UT. After this time, the principal power lies at a frequency of ~0.0025 Hz (period ~400 s) and continues until ~17 : 30 UT.

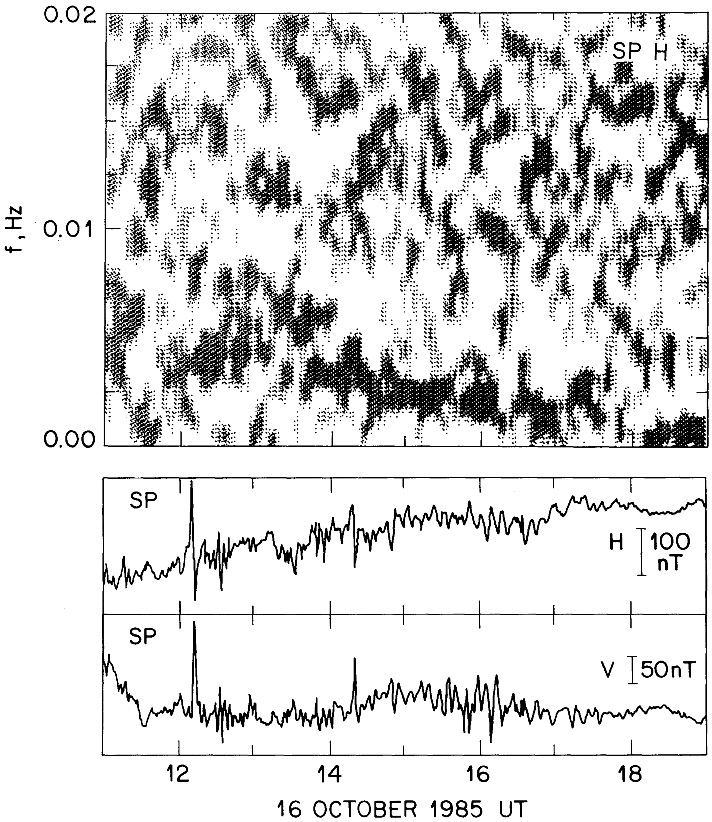


Fig. 6. Dynamic spectra of H-component magnetic data from South Pole (0–0.02 Hz) from 11–19 UT. Plotted at the bottom are the H- and V-component traces. The dynamic spectra show a decrease in the low-frequency band, from ~ 0.005 to ~ 0.002 Hz over this time interval.

4. Discussion

For the examples shown, it is clear that enhancements in hydromagnetic energy occur in a broad band at the magnetosphere cusp region near the time of appearance of intense field-aligned currents detected in the ionosphere. These currents may be evidence of reconnection processes at the dayside magnetopause. In particular, the enhancements in hydromagnetic energy above a background in the frequency range ~ 0.01–0.05 Hz are splotchy and quite broad banded. In contrast, there are significant enhancements in power in a band at a frequency ~ 0.0025 Hz (period ~ 400 s). In fact, this period is of the same order as the width of the field-aligned current events. This period τ is also approximately that which would be expected from calculations of the 'time of flight' of a resonant, transverse hydromagnetic wave at this geomagnetic latitude (e.g., Warner and Orr, 1979; Bamber, 1986)

$$\tau = 2 \int \frac{\mathrm{d}s}{V_{\mathrm{A}}}, \tag{1}$$

where V_A is the Alfvén velocity

$$V_A = \frac{B(s)}{\sqrt{\mu_0 \rho(s)}}\,, \tag{2}$$

$\rho(s)$ is the plasma mass density along the field line and s is the distance along a field line. $\rho(s)$ is usually taken to be expressable as a power law in radial distance r from the center of the Earth, r^n. Radoski (1966) has shown that this time-of-flight approximation corresponds exactly with the calculation of the transverse wave mode if the plasma density varies as r^{-6}. In this case, since $B(s) \propto B_0 r^{-3}$ for a dipole geometry, V_A is independent of r. This is generally not the case in the real magnetosphere. Warner and Orr (1979) note that using (1) results in a value $\sim 16\%$ less than the exact calculation for $\rho \propto r^{-4}$. A more recent calculation in an Olson–Pfitzer model magnetosphere also shows this (Bamber, 1986).

The dominant frequency at ~ 0.0025 Hz is in the Pc5 hydromagnetic band and is the band that Southwood (1987) predicted might be expected to be excited by convected field-aligned currents produced by FTE reconnection processes. This is reasonable since, as noted above, the width of the ionosphere current as deduced from ground magnetic records is of the order of a few minutes. Holzer and Reid (1975) earlier proposed that oscillations of the magnetopause at about these frequencies would occur at the onset of reconnection because of the requirement for an equilibrium field to be established in the ionosphere.

This low frequency band occurring in an interval of possible flux transfer events is consistent with the event of Glassmeier *et al.* (1984) and is inconsistent with the conclusion of Gillis *et al.* (1987) that pulsations with $f \lesssim 0.008$ Hz were not seen in their data at a time when flux transfer events would be expected. These last authors acknowledged that they might have missed some Pc5-band variations because of magnetic activity. However, it is also possible that their analysis scheme is insufficient to draw their conclusions from synchronous altitude spacecraft measurements of the magnetic field. Their paper suggests that they used visual inspection of low resolution paper records of the data, a procedure not designed to allow quantitative and reproducible determinations of the spectral content of time-varying signals.

The relationship observed between the dayside, near-cusp measurements and the low latitude measurements is quite interesting. It is clear that broad-band power in the Pc3–Pc4 band (~ 0.01–~ 0.05 Hz) is available to stimulate the transverse Alfvén wave mode on resonant magnetic field lines at the lower latitudes. The dominant frequency observed at Siple on 3 June, ~ 0.02 Hz shifts to ~ 0.025 Hz, at $\sim 13:00$ UT (just prior to local noon); these are frequencies that might be expected just outside the plasmapause at this station (Fukunishi and Lanzerotti, 1974a). The ratio of the powers between cusp latitude and SI at ~ 0.025 Hz at 12 : 00 LT

$$\left.\frac{P_{SI}}{P_{SP}}\right|_{0.025\ \mathrm{Hz}} \sim 3.5 \times 10^{-2}$$

is far higher than would be expected from surface wave damping from the magnetopause into the magnetosphere (e.g., Yumoto, 1986). Even the ratio of the 'background' powers at this frequency

$$\left.\frac{P_{\mathrm{SI}}}{P_{\mathrm{SP}}}\right|_{\mathrm{BKGND}} \sim 10^{-2}$$

is very high compared to expected damping, $\gtrsim 10^{-4}$, without field-line resonances.

In summary, it is clear that during the time of the possible ionosphere signatures of flux transfer events, hydromagnetic energy is observed near the boundary of the magnetosphere. In the frequency band ~ 0.02–~ 0.05 Hz the power is very broad band at cusp latitudes; this power is seen deeper inside the magnetosphere as narrower-banded hydromagnetic waves at a frequency corresponding to the fundamental resonance of the particular field line. Such resonances are transverse hydromagnetic waves – Alfvén waves. The ratio of the power between $L \sim 4$ and the cusp latitude is $\gtrsim 10^{-2}$. In addition to the higher frequency, broad-band hydromagnetic power, the cusp latitude data also show evidence for power in a narrower frequency band centered in the vicinity of ~ 0.0025 Hz. The period of this band is commensurate with the temporal width of the deduced field-aligned currents convected over the observing station and is the approximate period expected for the fundamental resonance of field lines at these latitudes.

Acknowledgements

We thank A. Wolfe for helpful comments and the U.S. National Science Foundation, Division of Polar Program, for logistical support of measurements at South Pole Station.

References

Akasofu, S.-I.: 1968, *Polar and Magnetospheric Substorms*, D. Reidel Publ. Co., Dordrecht, Holland.
Alfvén, H.: 1942a, *Nature* **150**, 405.
Alfvén, H.: 1942b, *Arkiv. Mat. Astron. Fysik* **29B**, 1.
Alfvén, H.: 1947, *Monthly Notices Roy. Astron. Soc.* **107**, 211.
Bamber, J. F.: 1986, M.Sc. Thesis, Univ. Calgary, Alberta, Canada.
Barnes, A.: 1970, *Cosmic Electrodyn.* **1**, 90.
Bolshakova, O. V. and Troitskaya, V. A.: 1977, *Geomagnetism i Aeronomiya* **17**, 722.
Campbell, W. H.: 1967, in S. Matsushita and W. H. Campbell (eds.), *Physics of Geomagnetic Phenomena*, Academic Press, New York, pp. 821–909.
Chen, L. and Hasegawa, A.: 1974a, *J. Geophys. Res.* **79**, 1024.
Chen, L. and Hasegawa, A.: 1974b, *J. Geophys. Res.* **79**, 1033.
Dungey, J. W.: 1954, Iono. Res. Lab., Pennsylvania State Univ., Sci. Rpt. No. 69, pp. 229–236.
Dungey, J. W.: 1955, in *The Physics of the Ionosphere*, London.
Fukunishi, H. and Lanzerotti, L. J.: 1974a, *J. Geophys. Res.* **79**, 142.
Fukunishi, H. and Lanzerotti, L. J.: 1974b, *J. Geophys. Res.* **79**, 4632.
Gillis, E. J., Rinjbeek, R., Kling, R., Speiser, T. W., and Fritz, T. A.: 1987, *J. Geophys. Res.* **92**, 5820.
Glassmeier, K.-H., Lester, M., Mier-Jedrzejowicz, W. A. C., Green, C. A., Rostoker, G., Orr, D., Wedeken, U., Junginger, H., and Amata, E.: 1984, *J. Geophys.* **55**, 108.

Goertz, C. K., Nielsen, E., Korth, A., Glassmeier, K.-H., Haldoupis, C., Holeg, P., and Hayward, D.: 1985, *J. Geophys. Res.* **90**, 4069.

Greenstadt, E. W., Mellott, M. M., McPherron, R. L. Russell, C. T., Singer, H. J., and Knecht, D. J.: 1983, *Geophys. Res. Letters* **10**, 659.

Hassam, A. B.: 1978, *J. Geophys. Res.* **83**, 643.

Haerendel, G., Paschmann, G., Sckopke, N., Rosenbauer, H., and Hedgecook, P. C.: 1978, *J. Geophys. Res.* **83**, 3195.

Holzer, T. E. and Reid, G. G.: 1975, *J. Geophys. Res.* **80**, 2041.

Jacobs, J. A.: 1970, *Geomagnetic Micropulsations*, Springer-Verlag, Heidelberg.

Kwok, Y. C. and Lee, L. C.: 1984, *J. Geophys. Res.* **89**, 10697.

Lanzerotti, L. J., Hunsucker, R. D., Rice, D., Lee, L. C., Wolfe, A., Maclennan, C. G., and Medford, L. V.: 1987a, *J. Geophys. Res.* **92**, 7739.

Lanzerotti, L. J., Hunsucker, R. D., Rice, D., Wolfe, A., Maclennan, C. G., and Medford, L. V.: 1987b, *Ann. Geophys.* (in press).

Lanzerotti, L. J., Medford, L. V., and Rosenberg, T. J.: 1982, *Antarctic J.U.S.* **17**, 235.

Lanzerotti, L. J., Lee, L. C., Maclennan, C. G., Wolfe, A., and Medford, L. V.: 1986, *Geophys. Res. Letters* **13**, 1089.

McKenzie, J. F.: 1970, *Planetary Space Sci.* **18**, 1.

Nagata, T., Kokubun, S., and Iijima, T.: 1983, *J. Geophys. Res.* **88**, 4621.

Paschmann, G., Haerendel, G., Papamostorakis, I., Sckopke, N., Bame, S. J., Gosling, J. T., and Russell, C. T.: 1982, *J. Geophys. Res.* **87**, 2159.

Russell, C. T. and Elphic, R. C.: 1978, *Geophys. Res. Letters* **6**, 33.

Russell, C. T. and Hoppe, M. M.: 1983, *Space Sci. Rev.* **34**, 155.

Saito, T.: 1969, *Space Sci. Res.* **10**, 319.

Samson, . C., Jacobs, J. A., and Rostoker, G.: 1971, *J. Geophys. Res.* **76**, 3675.

Sandholt, P. E., Deehr, C. S., Egeland, A., Lybekk, B., Viereck, R., and Romick, G. J.: 1986, *J. Geophys. Res.* **91**, 10063.

Southwood, D. J.: 1968, *Planetary Space Sci* **16**, 587.

Southwood, D. J.: 1974, *Planetary Space Sci.* **22**, 483.

Southwood, D. J.: 1987, *J. Geophys. Res.* **92**, 3207.

Sugiura, M.: 1961a, *Phys. Res. Letters* **6**, 255.

Sugiura, M.: 1961b, *J. Geophys. Res.* **66**, 4087.

Sugiura, M. and Wilson, C. T.: 1964, *J. Geophys. Res.* **69**, 1211.

Todd, H., Bromage, J. I., Cowley, S. W. H., Lockwood, M., van Eyken, A. P., and Willis, D. M.: 1986, *Geophys. Res. Letters* **13**, 909.

Trigg, D. F., Serson, P. H., and Camfield, P. A.: 1971, *Publ. 41*, Dept. Energy, Mines, Resources, Ottawa, Ontario, Canada.

Troitskaya, V. A.: 1967, in J. W. King and W. S. Newman (eds.), *Solar-Terrestrial Physics*, Academic Press, New York, pp. 213–274.

Troitskaya, V. A. and Gul'elmi, A. V.: 1967, *Space Sci. Rev.* **7**, 689.

Verö, J.: 1986, *J. Geophys.* **60**, 106.

Warner, M. R. and Orr, D.: 1979, *Planetary Space Sci* **27**, 679.

Wolfe, A. and Kaufmann, R. L.: 1975, *J. Geophys. Res.* **80**, 1764.

Wolfe, A., Kamen, E., Lanzerotti, L. J., Maclennan, C. G., Bamber, J. F., and Venkatesan, D.: 1987, *J. Geophys. Res.* **92**, 168.

Yumoto, K.: 1986, *J. Geophys.* **60**, 79.

Yumoto, K. and Saito, T.: 1980, *Planetary Space Sci.* **28**, 789.

FREE CONVECTION AND MASS TRANSFER EFFECTS ON THE MAGNETOHYDRODYNAMIC FLOWS NEAR A MOVING PLATE IN A ROTATING MEDIUM*

J. N. TOKIS**

Department of Astronomy, University of Manchester, England

(Received 11 December, 1987)

Abstract. An exact solution of the free-convection flow near an infinite vertical plate moving in a rotating medium in the presence of foreign mass and a transverse magnetic field is presented under a constant heating of the plate. It is apparent from this solution that the effects of the motion, the temperature, and the mass transfer are linear and, hence, can be studied independently. Three applications of physical interest are discussed. The non-magnetic case and non-rotating case are are also discussed.

1. Introduction

Magnetohydrodynamic free-convection flows have great significance not only of their own interest but also for the applications in the fields of stellar and planetary magnetospheres, aeronautics, chemical engineering, and electronics (cf. Alfvén, 1950; Gramer and Pai, 1973; Lui, 1987).

Unsteady MHD flows in a rotating medium have been studied by many investigators: see, for example, Debnath (1972, 1975), Puri and Kulshrestha (1976, 1983), and Soundalgekar and Pop (1970). Recently, Singh (1983, 1984) has discussed this problem with free-convection flows on an isothermal plate. Then, considering a sudden heating of the plate, Tokis (1986) has given the exact general solutions of the problem of MHD rotating free-convection flows; he has found that the velocity field consists of two parts, one due to the motion of the plate and the other due to the heating of the plate. Kythe and Puri (1987) have extended this work of Tokis by allowing the boundary condition of the temperature at the plate to be an arbitrary function of the time.

On the other hand, the effects of the phenomenon of mass transfer on a free-convection flow near an infinite vertical plate have been studied under simplifying approximations (cf. Nanousis, 1985; Raptis, 1982, 1984; Soundalgekar *et al.*, 1984). Hence, it appears that the study of the effects of mass transfer on MHD rotating free-convection flows will be of greater interest.

The purpose of the present investigation is the study of the problem of the MHD rotating free-convection flows with the effects of mass transfer. Several flows due to an infinite, vertical, moving plate are considered under the action of a uniform magnetic field and the buoyancy forces, which arise from the combination of concentration and temperature. A general solution for these flows are obtained with the aid of the Laplace

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, on 30 May 1988.
** Permanent address: Technological Education Institute, Koukouli, Patras, Greece.

transform when the plate temperature and the concentration are raised. Furthermore, these results are applied for three important cases of the motion of the plate, which is moving with the arbitrary velocity $U_0 f(t)$ where U_0 is a constant velocity and $f(t')$ a non-dimensional function of the time t'. Finally, we have evaluated the skin friction of the flows on the plate and discussed the results, which were obtained in this work.

2. Formulation of the Problem

Let us consider an unsteady three-dimensional flow of an electrically-conducting, viscous, incompressible fluid near an infinite non-conducting vertical plate (or surface). On this plate an arbitrary point has been chosen as the origin O of a Cartesian coordinate system, with axes Ox' and Oy' fixed on the plate and Oz' normal to it. Both the fluid and the plate are in a state of rigid rotation with uniform angular velocity Ω' about the z'-axis; in order to have Ω' always non-negative the coordinate system is taken at right-handed or left-handed, whichever is convenient. A uniform magnetic field $\mathbf{B}_0'$ is also assumed to be applied normal to the plate (Tokis and Pande, 1981).

Initially, the plate and the fluid are at the same temperature T'_∞ and in stationary condition. Also, the concentration level C'_∞ is the same everywhere. Suddenly, the plate is assumed to be moving with a velocity $U_0 f(t')$ in its own plane along the x'-axis; instantaneously the temperature of the plate and concentration are raised to T'_w $(\neq T'_\infty)$ and C'_w $(\neq C'_\infty)$, respectively, which are hereafter maintained constant.

In physical terms, we also assume that:

(i) All the physical properties of the fluid such as coefficient of viscosity (μ), coefficient of kinematic viscosity (ν), specific heat at constant pressure (c_p), thermal conductivity (k), coefficient of volume expansion (β'), electrical conductivity (σ), chemical molecular diffusivity (D), volumetric coefficient of expansion with concentration (β^*), etc., are constant.

(ii) The influence of variations of density (ρ) (with temperature) and concentration are considered only on the body force term, in accordance with the Boussinesq approximation.

(iii) The magnetic Reynolds number of the flow is taken to be small enough that the induced magnetic field can be neglected in comparison with the applied magnetic field.

(iv) The heat due to friction (or viscous dissipation) can be neglected in comparison with the conduction. This can be allowed because of the small velocities usually encountered in free-convection flows (cf. Holman, 1972). The Joulean dissipation is also neglected because it is of the same order of magnitude as the viscous dissipation.

(v) The species thermal diffusion thermal energy (Soret–Dufour effects) can be negligible, because the level of concentration is assumed as very low.

(vi) All the quantities are functions of the space coordinate z' and time t' only.

After all the above assumptions, it can be shown that the flow is governed by the following equations in the rotating frame:

Momentum equations:

$$\frac{\partial u'}{\partial t'} - 2\Omega' v' = \nu \frac{\partial^2 u'}{\partial z'^2} - \frac{\sigma B_0'^2}{\rho} u' + g\beta'(T' - T'_\infty) + g\beta^*(C' - C'_\infty), \quad (1)$$

$$\frac{\partial v'}{\partial t'} + 2\Omega' u' = \nu \frac{\partial^2 v'}{\partial z'^2} - \frac{\sigma B_0'^2}{\rho} v' . \quad (2)$$

Energy equation:

$$\frac{\partial T'}{\partial t'} = \frac{k}{\rho c_p} \frac{\partial^2 T'}{\partial z'^2} . \quad (3)$$

Diffusion equation:

$$\frac{\partial C'}{\partial t'} = D \frac{\partial^2 C'}{\partial z'^2} , \quad (4)$$

where $\mathbf{v}' = (u', v', 0)$ is the velocity; T' the temperature; and C' the concentration.

The initial and boundary conditions corresponding to the present problem are:

$$\mathbf{v}'(z', t') = \mathbf{0}, \quad T'(z', t') = T'_\infty, \quad C'(z', t') = C'_\infty, \quad \text{for} \quad t' \leq 0 ; \quad (5a)$$

$$\mathbf{v}'(0, t') = \{U_0 f(t'), 0, 0\}, \quad T'(0, t') = T'_w, \quad C'(0, t) = C'_w, \quad \text{for} \quad t' > 0 ; \quad (5b)$$

$$\mathbf{v}'(\infty, t') \to \mathbf{0}, \quad T'(\infty, t') = T'_\infty, \quad C'(\infty, t') = C'_\infty, \quad \text{for} \quad t' > 0 . \quad (5c)$$

The above equations can be reduced to non-dimensional form by the introduction of the following dimensionless quantities:

$$\theta = (T' - T'_\infty)/(T'_w - T'_\infty), \quad C = (C' - C'_\infty)/(C'_w - C'_\infty), \quad (6a)$$

$$z = z' U_0/\nu, \quad t = t' U_0^2/\nu, \quad (u, v) = (u', v')/U_0 , \quad (6b)$$

$$\Omega = \Omega' \nu/U_0^2 , \quad P = \mu c_p/k , \quad m = \nu^2 \sigma B_0'^2/(\mu U_0^2), \quad (6c)$$

$$G_r = \nu g\beta' (T'_w - T'_\infty)/U_0^2 , \quad G_c = \nu g\beta^*(C'_w - C'_\infty)/U_0^2 , \quad (6d)$$

$$S_c = \nu/D , \quad q \equiv u + iv \quad \text{with} \quad i = \sqrt{-1} . \quad (6e)$$

Using these, we obtain

$$\frac{\partial \theta}{\partial t} = \frac{1}{P} \frac{\partial^2 \theta}{\partial z^2} , \quad (7a)$$

$$\frac{\partial C}{\partial t} = \frac{1}{S_c} \frac{\partial^2 C}{\partial z^2} , \quad (7b)$$

$$\frac{\partial q}{\partial t} + 2i\Omega q = \frac{\partial^2 q}{\partial z^2} - mq + G_r \theta + G_c C , \quad (7c)$$

where P is the Prandtl number; S_c the Schmidt number; G_r the Grashof number; G_c the modified Grashof number; and m the magnetic field parameter.

The initial and boundary condition of the system of Equations (7) are

$$q(z, 0^-) = 0\,, \qquad \theta(z, 0^-) = 0\,, \qquad C(z, 0^-) = 0\,; \tag{8a}$$

$$q(0, t) = f(t)\,, \qquad \theta(0, t) = 1\,, \qquad C(0, t) = 1\,; \tag{8b}$$

$$q(\infty, t) \to 0\,, \qquad \theta(\infty, t) = 0\,, \qquad C(\infty, t) = 0\,. \tag{8c}$$

The system (7), subject to the boundary conditions (8), includes the effects of free-convection and mass transfer on the MHD flows near a moving plate in a rotating medium.

3. Solution of the Problem

In order to obtain an exact solution of the system of Equations (7), we shall use the Laplace transform method.

Applying the Laplace transform (with respect to time t) to the system (7) and boundary conditions (8), we find that the solution of the system is

$$\bar{\theta} = \frac{1}{s}\, e^{-z\sqrt{sP}}\,, \tag{9a}$$

$$\bar{C} = \frac{1}{s}\, e^{-z\sqrt{sS_c}}\,, \tag{9b}$$

$$\bar{q} = \bar{f}(s)\, e^{-z\sqrt{h+s}} + \frac{G_r}{s(h + s - sP)}\,(e^{-z\sqrt{sP}} - e^{-z\sqrt{h+s}}) + $$

$$+ \frac{G_c}{s(h + s - sS_c)}\,(e^{-z\sqrt{sS_c}} - e^{-z\sqrt{h+s}})\,, \tag{9c}$$

where

$$h \equiv m + 2i\Omega\,; \tag{10}$$

and a bar over a quantity denotes its Laplace transform with s as the transform variable.

Since the transformed function $\bar{f}(s)$ in expression (9c) has the known inverse arbitrary function $f(t)$, we put $\bar{f}(s) = 1$, which corresponds to the case $f(t) = \delta(t)$, where $\delta(t)$ is the Dirac delta functin; so, the inverse of the first term of the right-hand side of Equation (9c) is obtained in the form

$$\varphi^*(z, t) = \frac{H(t)}{2\pi}\, z t^{-3/2}\, e^{-ht - z^2/4t}\,, \tag{11}$$

where $H(t)$ is the Heaviside unit function.

Then, using the composition rule for Equation (9c), we get the inversion of expressions (9) in the form

$$\theta(z, t) = \operatorname{erfc}\left(\frac{z}{2}\sqrt{\frac{P}{t}}\right), \tag{12a}$$

$$C(z, t) = \operatorname{erfc}\left(\frac{z}{2}\sqrt{\frac{S_c}{t}}\right), \tag{12b}$$

$$q(z, t) = \varphi(z, t) + A_{gj}(z, t) + A_{ck}(z, t), \tag{12c}$$

where $\varphi(z, t)$ is given by

$$\varphi(z, t) = L^{-1}\{\bar{f}(s)\, e^{-z\sqrt{h+s}}\} = \int_0^t \varphi^*(z, \xi) f(t - \xi)\, d\xi\,; \tag{13}$$

and A_{gj}, $j = 1, 2$, and A_{ck}, $k = 1, 2$, denote the sums of inversion forms of the terms, with the factors G_r and G_c at the right-hand side of Equation (9c), respectively: namely,

(i) when $P \neq 1$:

$$A_{g1}(z, t) = \frac{G_r}{2h}\, [e^{bt}\{\Psi_2(z, t; \sqrt{bP}) - \Psi_2(z\sqrt{P}, t, \sqrt{b})\} - \\ - \Psi_2(z, t, \sqrt{h}) + \Psi_2(z\sqrt{P}, t, 0)]\,, \tag{14}$$

(ii) when $P = 1$:

$$A_{g2}(z, t) = \frac{G_r}{2h}\, [\Psi_2(z, t, 0) - \Psi_2(z, t, \sqrt{h})]\,; \tag{15}$$

(iii) when $S_c \neq 1$:

$$A_{c1}(z, t) = \frac{G_c}{2h}\, [e^{at}\{\Psi_2(z, t, \sqrt{aS_c}) - \Psi_2(z\sqrt{S_c}, t, \sqrt{a})\} - \\ - \Psi_2(z, t, \sqrt{h}) + \Psi_2(z\sqrt{S_c}, t, 0)]\,; \tag{16}$$

(iv) when $S_c = 1$:

$$A_{c2}(z, t) = \frac{G_c}{2h}\, [\Psi_2(z, t, 0) - \Psi_2(z, t, \sqrt{h})]\,; \tag{17}$$

with the abbreviations

$$a \equiv \frac{h}{S_c - 1} = \frac{m}{S_c - 1} + i\, \frac{2\Omega}{S_c - 1}\,, \tag{18}$$

$$b \equiv \frac{h}{P - 1} = \frac{m}{P - 1} + i\, \frac{2\Omega}{P - 1}\,, \tag{19}$$

and with the use of the function $\Psi_2(\alpha, t, \beta)$, which is defined in the Appendix.

Equations (12) are the general solution of the present problem. This solution includes the effects of free-convection and mass transfer on the magnetohydrodynamic flows near a moving plate in a rotating medium.

It is apparent from solution (12c) that the effects of the heating, the diffusion and the motion of the plate are linear and, hence, can be discussed independently. Since the expressions (12a, b) for the non-dimensional temperature θ and non-dimensional concentration C are unaffected by $f(t)$, we shall confine our analysis primarily to non-dimensional complex velocity $q(z, t)$ for various types of $f(t)$.

4. Applications of the General Formulation

In this section the previous general results are applied to the three most important cases of flow where the motion of the plate is impulsive, accelerated or decaying in an oscillatory manner. These cases are prescribed by physically acceptable forms of $f(t)$.

Case 1: Impulsive motion

In this case a single impulse of the plate is considered, which corresponds to

$$f(t) = H(t) . \tag{20}$$

On substituting Equation (20) into (13), the expression (12c) gives the complex velocity field

$$q(z, t) = \frac{H(t)}{2} \Psi_2(z, t, \sqrt{h}) + A_{gj}(z, t) + A_{ck}(z, t) , \tag{21}$$

where $A_{gj}(z, t)$ and $A_{ck}(z, t)$ are given by Equations (14)–(17) and the $\Psi_2(z, t, \sqrt{h})$ is defined in the Appendix.

Knowing the velocity field from Equation (21), we can now calculate the axial and transverse components of skin friction of the flow at the plate. In non-dimensional complex form these are given by

$$\tau_x + i\tau_y = \frac{\tau_x' + \tau_y'}{\rho U_0^2} = \left.\frac{\partial q}{\partial z}\right|_{z=0} \tag{22a}$$

$$= -H(t)\sqrt{h}\,\mathrm{erf}\sqrt{ht} - \frac{H(t)}{\sqrt{\pi t}}\, e^{-ht} + \tau_{gj} + \tau_{ck} , \tag{22b}$$

where τ_{gj}, $j = 1, 2$, and τ_{ck}, $k = 1, 2$, denoting by

$$\tau_{gj} \equiv \left.\frac{\partial A_{gj}(z, t)}{\partial z}\right|_{z=0} , \qquad \tau_{ck} \equiv \left.\frac{\partial A_{ck}(z, t)}{\partial z}\right|_{z=0} , \tag{23a, b}$$

are defined by

$$\tau_{g1} \equiv G_r \sqrt{\frac{P}{h(P-1)}}\, e^{bt} (\operatorname{erf}\sqrt{bt} - \operatorname{erf}\sqrt{bPt}) + \frac{G_r}{\sqrt{h}} \operatorname{erf}\sqrt{ht}\,,$$

$$\text{if} \quad P \neq 1\,; \tag{24a}$$

$$\tau_{g2} \equiv \frac{G_r}{h\sqrt{\pi t}} (e^{-ht} - 1) + \frac{G_r}{\sqrt{h}} \operatorname{erf}\sqrt{ht}\,, \quad \text{if} \quad P = 1\,; \tag{24b}$$

$$\tau_{c1} = G_c \sqrt{\frac{S_c}{h(S_c-1)}}\, e^{at} (\operatorname{erf}\sqrt{at} - \operatorname{erf}\sqrt{aS_c t}) + \frac{G_c}{\sqrt{h}} \operatorname{erf}\sqrt{ht}\,,$$

$$\text{if} \quad S_c \neq 1\,; \tag{25a}$$

$$\tau_{c2} \equiv \frac{G_c}{h\sqrt{\pi t}} (e^{-ht} - 1) + \frac{G_c}{h} \operatorname{erf}\sqrt{ht}\,, \quad \text{if} \quad S_c = 1\,. \tag{25b}$$

Case 2: *Accelerated motion*

Consider now a single accelerated motion of the plate. This corresponds to

$$f(t) = \frac{t}{t_0}\, H(t)\,, \tag{26}$$

where t_0 is a constant. In this case Equation (12c) gives the complex velocity field, with the aid of (26) and (13), by

$$q(z,t) = \frac{H(t)}{2t_0}\left[t\Psi_2(z,t,\sqrt{h}) - \frac{z}{2h}\, \Psi_1(z,t,\sqrt{h}) \right] +$$

$$+ A_{gj}(z,t) + A_{ck}(z,t)\,, \tag{27}$$

where the functions $\Psi_{1,2}(z, t, \sqrt{h})$ are given in the Appendix.

The skin friction at the plate due to the velocity $q(z, t)$ is given by

$$\tau_x + i\tau_y = -\frac{H(t)}{t_0}\left(\frac{1}{2\sqrt{h}} + t\sqrt{h}\right) \operatorname{erf}\sqrt{ht} -$$

$$- \frac{H(t)}{t_0} \sqrt{\frac{t}{\pi}}\, e^{-ht} + \tau_{gj} + \tau_{ck}\,, \tag{28}$$

where τ_{gj}, $j = 1, 2$ and τ_{ck}, $k = 1, 2$ are defined by Equations (24)–(25).

Case 3: Decaying oscillatory motion

The case of decaying oscillatory velocity of the plate is finally considered. This corresponds to

$$f(t) = \mathrm{Re}\,[H(t)\, e^{-(\lambda^2 - i\omega)t}]$$

$$= \frac{H(t)}{2}\,[e^{-(\lambda^2 - i\omega)t} + e^{-(\lambda^2 + i\omega)t}]\,, \tag{29a, b}$$

where the real constants λ and $\omega\,(>0)$ are the dimensionless attenuation coefficient and the dimensionless radian frequency, respectively. Substituting (29b) into (13), the expression for the complex velocity is obtained, from (12c) in the form

$$q(z, t) = \frac{H(t)}{4}\, e^{-(\lambda^2 - i\omega)t}\, \Psi_2(z, t, \sqrt{\gamma_1}) + \frac{H(t)}{4}\, e^{-(\lambda^2 + i\omega)t}\, \Psi_2(z, t, \sqrt{\gamma_2}) +$$

$$+ A_{gj}(z, t) + A_{ck}(z, t)\,, \tag{30}$$

with the abbreviation

$$\gamma_1, \gamma_2 = m - \lambda^2 + i(2\Omega \pm \omega)\,. \tag{31}$$

The expression for the complex skin friction is then given by

$$\tau_x + i\tau_y = -\frac{H(t)}{2}\,[\sqrt{\gamma_1}\, e^{-(\lambda^2 - i\omega)t}\, \mathrm{erf}\sqrt{\gamma_1 t} + \sqrt{\gamma_2}\, e^{-(\lambda^2 + i\omega)t}\, \mathrm{erf}\sqrt{\gamma_2 t}] -$$

$$-\frac{H(t)}{\sqrt{\pi t}}\, e^{-ht} + \tau_{gj} + \tau_{ck}\,. \tag{32}$$

5. Special Cases

In this section we shall discuss special cases of the present problem.

5.1. Non-rotating case

In the absence of rotation (i.e., $\Omega = 0$), the expressions (10), (18), (19), and (31) are reduced to

$$h = m\,, \qquad a = \frac{m}{S_c - 1}\,, \qquad b = \frac{m}{P - 1}\,, \qquad \gamma_{1,2} = m - \lambda^2 \pm i\omega\,. \tag{33}$$

All the other forms of the above solutions of the present problem remain unaffected.

The problem of impulsive motion of the non-rotating case has been discussed by Georgantopoulos and Nanousis (1980) with similar results.

5.2. Non-magnetic case

When the magnetic field $\mathbf{B}_0'$ is zero (or the magnetic field parameter $m = 0$), the expressions (9), (11)–(17), (21)–(25), (27), (28), (29), (30), and (32) have the same form with

$$h = 2i\Omega\,, \qquad a = i\,\frac{2\Omega}{S_c - 1}\,, \qquad b = i\,\frac{2\Omega}{P - 1}\,, \quad \gamma_{1,2} = -\lambda^2 + i(2\Omega \pm \omega)\,. \quad (34)$$

This paper describes thus the flow in both the hydrodynamic and the hydromagnetic cases.

It may be noted that the general solution of this problem gives, readily, the results of many cases; i.e., $m = 0$ and $\Omega = 0$; $G_r = 0$ and $\Omega = 0$; and $m = 0$ and $G_c = 0$.

The results for the case $G_c = 0$ are identical with those of Tokis (1986) and for the case $G_c = 0$ and $\Omega = 0$ with those of Tokis (1985).

6. Discussion

Exact solutions for unsteady free-convection problem with mass transfer near a vertical plate immersed in a rotating incompressible viscous medium in the presence of a uniform magnetic field orthogonal to the plate have been determined.

These solutions are given with the Prandtl number (P) different or equal to unity. Indeed, for most gases P is between 0.7 and 0.85, whereas for liquids P is generally greater than one; for $P = 1$ gives a solution for very restricted classes of gas, namely, steam and ammonia.

It is worthwhile pointing out that the velocity field (cf. Equation (12c)) consists of three parts: the first ($\varphi(z, t)$) due to the motion of the plate, the second ($A_{gj}(z, t)$) due to the heating of the plate and the third ($A_{ck}(z, t)$) due to the concentration level near to the plate. The total velocity of the flow can be obtained by superposition of these three parts.

The effects on the flow of the present problem due to the term $\varphi(z, t)$ of the motion have been discussed in Puri and Kulshrestha (1976) with zero suction, in Puri and Kulshrestha (1983) and Tokis (1986). Also, the effects on the same flow due to the term $A_{gj}(z, t)$ of the temperature have been discussed in Tokis (1986) and in Kythe and Puri (1987).

From the expressions (14)–(17) it is obvious that the effects of the term $A_{ck}(z, t)$ of mass transfer are similar to those of the term of temperature. Thus, following the procedures discussed in Tokis (1986) and in Kythe and Puri (1987) we have for mass transfer similar discussions and conclusions to those of the temperature.

It should be pointed out here that the effects of mass transfer are the same for all types of motion of the plate expressed by term $A_{ck}(z, t)$. In the limit, as $t \to \infty$, this term is reduced to

$$A_{ck}(z, \infty) = \frac{G_c}{h}\,(1 - e^{-z\sqrt{h}})\,, \qquad (34)$$

which is similar to the form (cf. Tokis, 1986)

$$A_{gj}(z, \infty) = \frac{G_r}{h} \left(1 - e^{-z\sqrt{h}}\right). \tag{35}$$

These forms reveal that the velocity field depends, finally, on the Grashof number G_r and on the modified Grashof number G_c.

Finally, it can be verified by analysing the expression of the skin friction that for small time the skin friction increases for $G_r > 0$, $G_c > 0$ and all P and S_c, while for large time it becomes constant for $P > 1$ and $S_c \geq 1$, but tends to increase without limit for $P < 1$ and $S_c < 1$.

Appendix

The following functions are introduced. These were obtained from integrals used in the text.

For

$$\beta > 0\,, \qquad t \geq 0\,, \qquad R_e(\alpha) > 0\,, \tag{A1}$$

$$\Psi_1(\alpha, t, \beta) = \frac{2\beta}{\pi^{1/2}} \int_0^t \xi^{-1/2} \exp\left(-\alpha^2 \xi - \frac{\beta^2}{4\xi}\right) \mathrm{d}\xi =$$

$$= e^{-\alpha\beta} \operatorname{erfc}\left(\frac{\beta}{2t^{1/2}} - \alpha t^{1/2}\right) - e^{\alpha\beta} \operatorname{erfc}\left(\frac{\beta}{2t^{1/2}} + \alpha t^{1/2}\right), \tag{A2}$$

$$\Psi_2(\alpha, t, \beta) = \frac{2\alpha}{\pi^{1/2}} \int_0^t \xi^{-3/2} \exp\left(-\alpha^2 \xi - \frac{\beta^2}{4\xi}\right) \mathrm{d}\xi =$$

$$= e^{-\alpha\beta} \operatorname{erfc}\left(\frac{\beta}{2t^{1/2}} - \alpha t^{1/2}\right) + e^{\alpha\beta} \operatorname{erfc}\left(\frac{\beta}{2t^{1/2}} + \alpha t^{1/2}\right). \tag{A3}$$

Particular cases of (A2) and (A3) are

$$\Psi_1(\alpha, \infty, \beta) = 2\, e^{-\alpha\beta} = \Psi_2(\alpha, \infty, \beta)\,. \tag{A4}$$

Furthermore, Equations (A2) and (A3) are valid when $\alpha = 0$ and $\beta = 0$, respectively, so that

$$\Psi_1(0, t, \beta) = 2 \operatorname{erf}(\beta t^{1/2})\,, \tag{A5}$$

$$\Psi_2(\alpha, t, 0) = 2 \operatorname{erfc} \frac{\alpha}{2t^{1/2}}\,. \tag{A6}$$

References

Alfvén, H.: 1950, *Cosmical Electrodynamics*, Oxford Univ. Press, London.

Debnath, L.: 1972, *Z. Appl. Math. Mech.* **52**, 623.

Debnath, L.: 1975, *Z. Appl. Math. Mech.* **55**, 431.

Georgantopoulos, G. A. and Nanousis, N. D.: 1980, *Astrophys. Space Sci.* **67**, 229.

Gramer, K. P. and Pai, Shih-I.: 1973, *Magnetofluid Dynamics for Engineers and Applied Physics*, McGraw-Hill Book Co., New York.

Holman, J. P.: 1972, *Heat Transfer*, McGraw-Hill Book Co., New York.

Kythe, P. K. and Puri, P.: 1987, *Astrophys. Space Sci.* **135**, 219.

Lui, A. T. Y. (ed.): 1987, *Magnetotail Physics*, The Johns Hopkins Univ. Press, Baltimore and London.

Nanousis, N. D.: 1985, *Astrophys. Space Sci.* **113**, 257.

Puri, P. and Kulshrestha, P. K.: 1976, *Trans. Am. Soc. Mech. Eng. J. Appli. Mech.* **43**, 205.

Puri, P. and Kulshrestha, P. K.: 1983, *Z. Appl. Math. Mech.* **63**, 489.

Raptis, A. A.: 1982, *Astrophys. Space Sci.* **86**, 43.

Raptis, A. A.: 1984, *Z. Appl. Math. Mech.* **64**, 127.

Singh, A. K.: 1983, *Astrophys. Space Sci.* **95**, 283.

Singh, A. K.: 1984, *Astrophys. Space Sci.* **103**, 155.

Soundalgekar, V. M. and Pop, I.: 1970, *Bull. Math. Soc. Sci. Math. Rep. Social Roumanie* **14**, 375.

Soundalgekar, V. M., Birajdar, N. S., and Darwhekar, V. K.: 1984, *Astrophys. Space Sci.* **100**, 159.

Tokis, J. N.: 1985, *Astrophys. Space Sci.* **112**, 413.

Tokis, J. N.: 1986, *Astrophys. Space Sci.* **119**, 305.

Tokis, J. N. and Pande, G. C.: 1981, *Trans. Am. Soc. Mech. Eng. J. Appl. Mech.* **48**, 225.

AN ELECTRIC-CURRENT DESCRIPTION OF SOLAR FLARES*

S.-I. AKASOFU

Geophysical Institute, University of Alaska, Fairbanks, Alaska, U.S.A.

(Received 9 July, 1987)

Abstract. The presently prevailing theories of solar flares rely on the hypothetical presence of magnetic flux tubes beneath the photosphere and the two subsequent hypotheses, their emergence above the photosphere and explosive magnetic reconnection, converting magnetic energy carried by the flux tubes to solar flare energy. In this paper, we discuss solar flares from an entirely different point of view, namely in terms of power supply by a dynamo process in the photosphere. By this process, electric currents flowing along the magnetic field lines are generated and the familiar 'force-free' fields or the 'sheared' magnetic fields are produced. Upward field-aligned currents thus generated are carried by downward streaming electrons; these electrons can excite hydrogen atoms in the chromosphere, causing the optical Hα flares or 'low temperature flares'. It is thus argued that as the 'force-free' fields are being built up for the magnetic energy storage, a flare must already be in progress.

1. Introduction

The present guiding concept in searching for basic processes, observationally and theoretically, of solar flares is based on a hypothesis that solar activities are various manifestations of emergence of intense magnetic flux tubes from beneath the photosphere and the subsequent consequences of it. The photosphere is considered to be merely a passive medium through which the magnetic flux tubes penetrate from below. Therefore, the main theoretical efforts have so far been concentrated on understanding the stability of the magnetic flux tubes for sunspots and instability processes for solar flares, leading to explosive annihilation of magnetic energy carried up by the flux tubes (cf. Parker, 1963; Sweet, 1969; Colgate, 1978; Priest and Milne, 1980). For these reasons, all morphological features of solar activities have been described using terms with such theoretical implications, e.g., magnetic flux 'emergence', magnetic energy 'storage', flares 'building-up', 'triggering' instability, and magnetic energy 'release'. Such an elaborate scheme has been conceived primarily because it is widely believed that no photospheric process can supply the energy at the rate of as much as 10^{29} erg s^{-1}. In the next section, it will be shown that a dynamo process in the photosphere can supply the desired energy with the required rate; for details of the discussions contained in this paper, see Akasofu (1984).

Actually, it is desirable that observational studies refrain from usage of such specific model-oriented terms. Indeed, it is unfortunate that most morphological papers on solar flares try to describe solar flares only in such terms. In fact, such a practice has confused some authors. In this respect, it is interesting to note the remark made by Schmahl (1983): "At the present stage of our understanding of flare buildup, we are not sure of the significance of all that has been observed in the way of buildup signatures..."

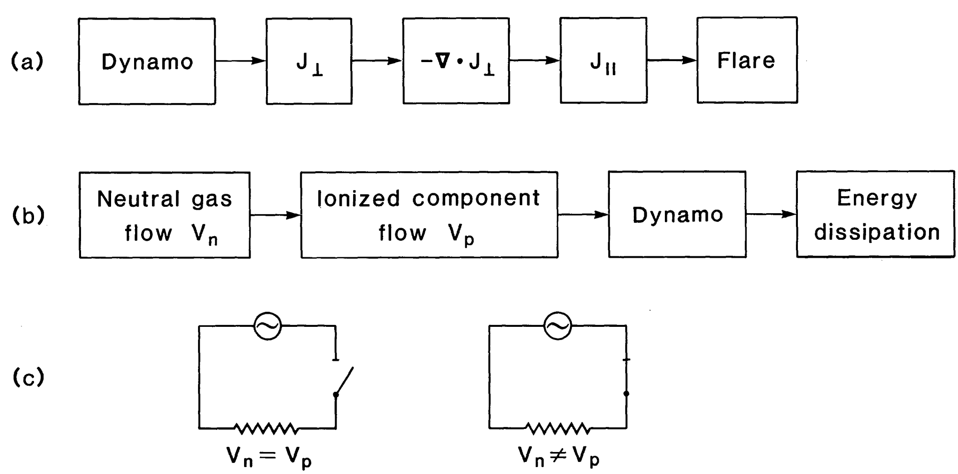
* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 303–309.

2. Need for a Dynamo Process in the Photosphere

It is important to recognize that in discussing solar flares we deal with electromagnetic processes. Thus, we have to consider a dynamo process as the power supply for the relevant electromagnetic processes. It is thus puzzling to find why solar flares have not been discussed in terms of the power supply-dissipation process. Now, since we must deal with the dynamo process as the power supply, it is important to clarify the energy source for our process at the outset. A dynamo is a machine that converts mechanical energy into electrical energy. Thus, we must identify first the mechanical energy for our dynamo. The photosphere is only a very weakly ionized atmosphere, the degree of ionization being 10^{-4}–10^{-5} in the quiet photosphere and perhaps 10^{-6}–10^{-7} in the vicinity of sunspots. Since both the neutral component and the ionized component of the photosphere are expected to move with similar speeds (because of a high collision frequency), the bulk kinetic energy is mostly carried by the neutral component. If we assume an area of $10^5 \times 10^5$ km (a typical radius of an umbra $\sim 10^4$ km) and depth of 10^3 km, the density of $\sim 10^{-7}$ g cm^{-3} and the speed of 1000 m s^{-1}, the total bulk kinetic energy of the photospheric flow is $\sim 10^{31}$ erg. The highest energy dissipation rate in most intense flares is known to be $\sim 10^{29}$ erg s^{-1}.

Now, let us assume that thermal convection, pressure gradient or other mechanisms maintain the flow of the neutral component. The dynamo process associated with a very weakly-ionized gas implies that the bulk kinetic (mechanical) energy of the neutral component must be transferred to the bulk kinetic energy of the ionized component. The bulk kinetic energy of the ionized component thus transferred is converted into electrical power; see the block diagrams in Figure 1.



PHOTOSPHERIC DYNANO AND LOAD

Fig. 1. Block diagram, showing the relationship between the dynamic process and solar flares and the relationship between the neutral component and the ionized component. Note that if $V_n - V_p$, the flare circuit does not generate the power.

The dissipation process of the power thus produced reduces first the bulk kinetic energy of the ionized component and thus its bulk speed. However, in a steady state the differential speed between the ionized component $\mathbf{V}_p$ and the neutral component $\mathbf{V}_n$ (which is assumed to be constant here) ensures the transfer of the bulk kinetic energy from the neutral component to the ionized component. Further, the above estimates of the total kinetic energy of the neutral flow ($\sim 10^{31}$ erg) and of the flare dissipation rate ($\sim 10^{29}$ erg s^{-1}) ensure that the neutral flow is hardly affected by the flare dissipation, so long as thermal convection or pressure gradient can be maintained. Kan *et al.* (1983) showed that the power of the dynamo is given by

$$P = -\Sigma_p \mathbf{E} \cdot (\mathbf{V}_n - \mathbf{V}_p) \times \mathbf{B}\,, \tag{1}$$

where $\mathbf{E} = -\mathbf{V}_p \times \mathbf{B}$ and Σ_p is the Pedersen conductivity which is similar to that of sea water in the lower photosphere (Cowling, 1953). Therefore, however, small it may be, the velocity differential is most crucial for the photospheric dynamo to generate the power for the 'flare circuit'. An important point is that one cannot determine *a priori* the velocity differential on the basis of local conditions alone. One must know the dissipation rate in the whole circuit (including the flare region) in order to estimate the velocity differential, since the dissipation takes place in remote regions connected by the magnetic field lines to the dynamo region. The general practice in solar physics is to assume $\mathbf{V}_n = \mathbf{V}_p$ in the photosphere at all times. This corresponds to a dynamo with an open circuit (or without load); no power is generated by a dynamo if it is connected to an open circuit. The photospheric dynamo was proposed first by Alfvén (1950).

In the past, it has been generally accepted that there is no process to supply the energy at the rate of 10^{29} erg s^{-1} and thus that the energy must be stored prior to flare onset. One can see, however, that the initial assumption of $\mathbf{V}_n = \mathbf{V}_p$ throws away the basic energy supply process at the outset.

3. Force-Free Fields and Sheared Magnetic Fields

One of the most important findings of the past extensive morphological studies of solar flares is that an important condition leading to solar flares is dramatic relative motions of spots, resulting in the so-called 'sheared magnetic fields'. Here, the 'sheared' magnetic field is considered a non-potential field ($\nabla \times \mathbf{B} \neq 0$) which is caused by shear flows in the photosphere. As is well known, a non-potential field ($\nabla \times \mathbf{B} \neq 0$) tends to settle in a particular configuration called 'force-free' field in the chromosphere and the corona, namely $(\nabla \times \mathbf{B}) \times \mathbf{B} = 0$, indicating that the electric current tends to flow along $\mathbf{B}$. The 'force-free' fields tend to have twisted or 'sheared' apearance of magnetic field lines (or of chromospheric fibrils and prenumbral structures) in the vicinity of *active* sunspots. A new sunspot pair is often formed by two spots from different sunspot pairs or groups, but their field configuration is often drastically different from the field of a potential field (see Zirin, 1983; Tanaka, 1980; and Zirin and Tanaka, 1973).

The essense of a force-free field is simply that electric currents ($\mathbf{J} = \nabla \times \mathbf{B}$) flow parallel to magnetic field lines ($(\nabla \times \mathbf{B}) \times \mathbf{B} = 0$). Such currents are generally called 'field-aligned currents'. Thus, if force-free fields are essential for solar flares, field-aligned currents must also be essential for solar flares. However, a force-free field in the corona cannot be isolated from the photosphere and the chromosphere. Obviously, the next question then is how the field-aligned currents are generated. In solar physics, this question has been completely ignored, in spite of the fact that a large number of papers have been written on the force-free fields. Now, since $\mathbf{J}_{\parallel} = -\nabla \cdot \mathbf{J}_{\perp}$, one must find ways to generate the perpendicular current $\mathbf{J}_{\perp}$ which has non-zero divergence ($\nabla \cdot \mathbf{J}_{\perp} \neq 0$). Therefore, if the 'force-free' fields are essential for solar flares, one can conclude that solar flares require a dynamo process in the photosphere to generate non-zero divergence $\mathbf{J}_{\perp}$.

4. Field-Aligned Currents

An important point to repeat here is that when one considers a 'force-free' field, one must consider a dynamo process which can provide the necessary field-aligned currents $J_{\parallel}$. In the past, such an effort was lacking, as pointed out by Kuperus (1983); he commented: 'However, a pure force-free field description bypasses the problem of the source of the currents and thus of the sources of energy input.' Hasegawa and Sato (1980) showed that vortex flows are essential in generating field-aligned currents.

Now, the reason for extensive studies of the 'force-free' fields during the last decade is that it has been suggested that the 'force-free' configuration stores magnetic energy for solar flares. However, the important point to make in this paper is that as the 'force-free' field is being built up, a solar flare must have already been in progress. This is because electrons carrying the upward field-aligned current (needed for the 'force-free' field) must be streaming down into the chromosphere, causing the optical Hα flare. In the case of the terrestrial aurora, the streaming electrons toward the polar region (toward the converging field lines) are accelerated by an interesting potential structure which develops at an altitude of about 10000 km. In our view, as the dynamo power exceeds a critical value and thus the field-aligned currents exceed a critical value, a significant potential drop will develop along the magnetic field lines in the coronal altitude, accelerating current-carrying electrons downward. As these electrons interact with the chromosphere and excite and ionize the atmospheric constituents there, flares will occur. In this respect, the discovery of monoenergetic electrons by Lin and Schwartz (1988) is of great interest; flare electrons and auroral electrons have strikingly similar energy spectra.

The time constant involved in the development of a flare depends on flare-associated phenomena. An Hα flare develops typically in 20–30 min. However, X-ray and microwave observations of flares indicate a much faster rise time, much less than 1 min. A typical auroral substorm develops in about 30 min. However, when the same onset is observed by an X-ray detector, its intensity can increase two to three orders of magnitude in about 1 min. It appears that the auroral potential structure can generate energetic electrons very fast.

As the dissipation rate increases, the bulk kinetic energy of the ionized component is expended and thus the bulk speed of the ionized component tends to be reduced. However, the resulting increase of speed differential will ensure the transfer of the kinetic energy from the neutral gas to the ionized component, so that the flare will continue until the bulk kinetic energy of the neutral gas is expended (see Equation (1) and Figure 1). This description is thus entirely different from the concept of explosive reconnection.

The above statement does not exclude the occurrence of magnetic reconnection, if an anti-parallel magnetic field configuration can arise. (However, an anti-parallel configuration does not seem to be a necessary condition for many flares.) In fact, it may explain many of the fascinating dynamical features associated with solar flares. The point to make is that it is not a spontaneous process which can convert the stored magnetic energy into the flare energy. It has to be *directly* driven by the photospheric dynamo process. Time variations of the output from magnetic reconnection processes may follow fairly well with those of the input (e.g., of the power). Magnetic reconnection at both the dayside magnetopause and the magnetotail is driven by the solar wind in a similar way. In the next section, we examine indications that solar flares are indeed directly driven.

5. Crucial Flares Observations

One of the ways to test the hypothesis of explosive magnetic reconnection is to examine the degree of the shear of the fibrils around a spot *during* (not after) solar flares. Such a test has already been conducted in a study of a flare by Neidig (1979) who showed that the shear actually *increased* during a flare. Furthermore, in a very comprehensive study of flares on 6–7 April, 1980, Krall *et al.* (1982) noted: "... sufficiently time-resolved observations from 14 : 00 to 21 : 00 UT, during which four major flares occurred, showed no evidence of significant change in the transverse field orientations". Similarly, Bruzek (1975) observed: "Before the flare and *even after* its maximum the Hα chromosphere inside the cell showed a pattern of parallel fibrils running roughly in the North–South direction... However, after the flare receded, the pattern had changed considerably; the fibrils were then arranged rather radically inside the cell..." In their observation of two-ribbon flares in a spotless situation, Moore and LaBonte (1980) noted: "... the magnetic field in the chromosphere and in the large filament was strongly sheared across the neutral line before the flare and was still highly sheared in the chromosphere after the filament erupted...".

Therefore, these careful observations indicate clearly that the magnetic field does not necessarily 'relax' from the sheared situation to a potential configuration *during* flares and that the magnetic energy is not necessarily released. There are some indications that the stored energy even increased during one event. It is more likely that the dynamo process and the field-aligned currents are maintained and/or intensified during flares. If the above observations cannot be considered definitive, most of the observations which support the relaxation of the magnetic field cannot be definitive either. Further, if a slight relaxation (if any) would be enough to cause a major flare (as many argue), how can the field relax drastically several hours *after* a flare without causing a flare (as many show)? In our view, the drastic relaxation *after* the flare is only an indication that the dynamic process is decreased as the neutral wind wanes; it is not caused as a result of release of the flare energy.

It is well established that a flare has two components, the low-temperature (normal) component and the high-temperature (relativistic) component (Švestka, 1976). The former is associated with the Hα emission and the latter is associated with impulsive gamma-ray emissions and other impulsive energetic phenomena, as well as enhanced Hα emission. Such a high-energy component may be too energetic to be explained by the potential drop in the field-aligned portion of the circuit, which was discussed earlier. Alfvén and Carlqvist (1967) suggested that current interruption process in the circuit along magnetic field lines is crucial for solar flares. They suggested that a small density depression in the field-aligned current portion of the circuit may be the cause for the current interruption. It is interesting to note in this connection that the plasma density decreases considerably in the auroral potential structure (Calvert, 1981), so that the current interruption may occur as a second stage process (after the potential structure formation during the first stage). If the current interruption can take place, the magnetic energy stored in the circuit may be dissipated impulsively in the density gap, producing a very high voltage there and accelerating the plasma particles to relativistic energy ranges. Alfvén and Carlqvist (1967) estimated that the voltage developed by the current interruption may reach as high as $V = L\,\mathrm{d}I/\mathrm{d}t \sim LI/\tau \sim 10^9$ V where $L = 10H$ and $I = 10^{11}$ A and $\tau = 10^3$ s.

6. Concluding Remarks and Summary

In this paper, we discussed solar flares by introducing the photospheric dynamo process and its consequences. Our emphasis is the generation mechanism for the field-aligned currents which are essential for the 'force-free' field. It is pointed out that as the

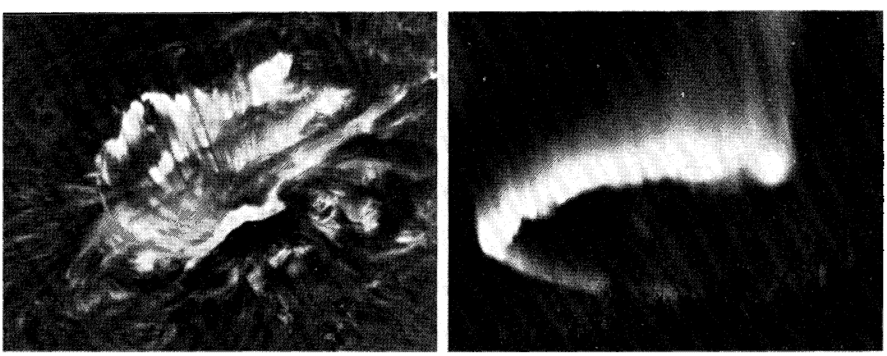


Fig. 2. Photograph of a solar flare (courtesy of the Big Bear Solar Observatory) and of the aurora (courtesy of Kashiwara Astronomy Club).

'force-free' field is being built up for the magnetic energy storage, the electrons carrying the upward field-aligned currents must be streaming down to the chromosphere, causing Hα flares. We have found several statements that the 'stress' or 'shear' of the magnetic field has not changed or even increased during the maximum epoch of some solar flares, indicating that a high dynamo power is required to initiate and maintain the high-energy dissipation during flares.

It is fascinating to infer that an auroral 'curtain' and a flare 'ribbon' are produced by similar processes (Figure 2). Both result from atmospheric emissions, the former mainly from oxygen atoms and nitrogen molecules and the latter mainly from hydrogen and helium atoms. Both phenomena are associated with X-ray emissions. Both are associated with similar electron spectra.

In summary, the magnetic flux tube-based description of solar flares ignores the power generation process for flares. The past studies of the force-free fields ignores the generation of the field-aligned currents. It is unlikely that one can reach a satisfactory understanding of flare processes by neglecting such basic processes. A full, but qualitative, description of sunspots and solar flares in terms of photospheric dynamo process is given in Akasofu (1984).

References

Akasofu, S.-I.: 1984, *Planetary Space Sci.* **32**, 1469.

Alfvén, H.: 1950, *Cosmical Electrodynamics,* Oxford University Press, Oxford.

Alfvén, H. and Carlqvist, P.: 1967, *Solar Phys.* **1**, 220.

Bruzek, A.: 1975, *Solar Phys.* **42**, 215.

Calvert, W.: 1981, *Geophys. Res. Letters* **8**, 919.

Colgate, S. A.: 1978, *Astrophys. J.* **221**, 1068.

Cowling, T. G.: 1953, 'Solar Electrodynamics', in G. P. Kuiper (ed.), *The Sun*, University of Chicago Press, Chicago, p. 532.

Hasegawa, A. and Sato, T.: 1980, 'Generation of Field-Aligned Current During Substorm', D. Reidel, Publ. Co., Dordrecht, Holland, p. 529.

Kan, J. R., Akasofu, S.-I., and Lee, L. C.: 1983, *Solar Phys.* **84**, 153.

Krall, K. R., Smith, J. B., Jr., Hagyard, M. J., West, E. A., and Cummings, N. P.: 1982, *Solar Phys.* **79**,, 59.

Kuperus, M.: 1983, *Space Sci. Rev.* **34**, 47.

Lin, R. P. and Schwartz, R. A.: 1988, *Astrophys. J.* (in press).

Moore, R. L. and LaBonte, B. J.: 1980, 'The Filament Eruption in the 3B Flare of July 29, 1973: Onset and Magnetic Field Configuration', in M. Dryer and E. Tandberg-Hanssen (eds), *Solar and Interplanetary Dynamics*, D. Reidel, Publ. Co., Dordrecht, Holland, p. 207.

Neidig, D. F.: 1979, *Solar Phys.* **61**, 121.

Parker, E. N.: 1963, *Astrophys. J. Suppl.* **8**, 77.

Priest, E. R. and Milne, A. M.: 1980, *Solar Phys.* **65**, 315.

Schmahl, E. J.: 1983, *Adv. Space Res.* **2**, 73.

Švestka, Z.: 1976, *Solar Flares*, D. Reidel, Publ. Co., Dordrecht, Holland.

Sweet, P. A.: 1969, *Ann. Rev. Astron. Astrophys.* **7**, 149.

Tanaka, K.: 1980, 'Prediction of Some Great Flares Based on the Magnetic Field Configuration and Evolution of Sunspot Groups', *Solar-Terrestrial Prediction Proceedings*, Vol. 3, C-1.

Zirin, H.: 1983, *Astrophys. J.* **274**, 900.

Zirin, H. and Tanaka, K.: 1973, *Solar Phys.* **32**, 173.

SOLAR FLARES THROUGH ELECTRIC CURRENT INTERACTION*

CORNELIS DE JAGER

Laboratory for Space Research and Astronomical Institute, Utrecht, The Netherlands

(Received 24 June, 1987)

Abstract. The fundamental hypothesis by Alfvén and Carlqvist (1967) that solar flares are related to electrical currents in the solar chromosphere and low corona is investigated in the light of modern observations. We confirm the important role of currents in solar flares. There must be tens of such current loops ('flux threads') in any flare, and this explains the hierarchy of bursts in flares. We summarize quantitative data on energies, numbers of particles involved and characteristic times. A special case is the high-energy flare: this one may originate in the same way as less energetic ones, but it occurs in regions with higher magnetic field strength. Because of the high particle energies involved their emission seats live only very briefly; hence the area of emission coincides virtually with the seat of the instability. These flares are therefore the best examples for studying the primary instability leading to the flare. Finally, we compare the merits of the original Alfvén–Carlqvist idea (that flares originate by current interruption) with the one that they are due to interaction (reconnection) between two or more fluxthreads. We conclude that a final decision cannot yet be made, although the observed extremely short time constants of flare bursts seem to demand a reconnection-type instability rather than interruption of a circuit.

1. The Original Alfvén–Carlqvist Hypothesis

Some twenty years ago Alfvén and Carlqvist (1967) published a paper in the then brand new journal *Solar Physics* in which they stressed the importance of electrical currents in the solar atmosphere for causing solar flares (Figure 1). These currents had been discovered a few years before by Severny (1964, 1965) who had at that time just developed the technique of vector magnetography, and who had found, by applying the Maxwell's equation

$$i = (c/4\pi)\,\mathrm{curl}\,H\,,$$

that currents of about 10^{11} A were flowing through localized spots in the flare areas.

Alfvén and Carlqvist (1967) thereupon suggested that flares might originate by sudden interruptions of such currents. Quoting from their paper: "An inductive circuit in which a current exists has a general tendency to 'explode'. This means that if we try to interrupt the current at a certain point, the whole magnetic energy of the circuit tends to be dissipated at that point." A well-known laboratory example is of course the spark that originates when a direct current is suddenly interrupted.

This assumption was supported by the subsequent discovery of Moreton and Severny (1968) that the chromospheric seats of these currents coincide with the bright knots seen in the first phase of flares. This discovery was much later fully supported by X-ray imaging observations (Duijveman *et al.*, 1982).

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 311–319.

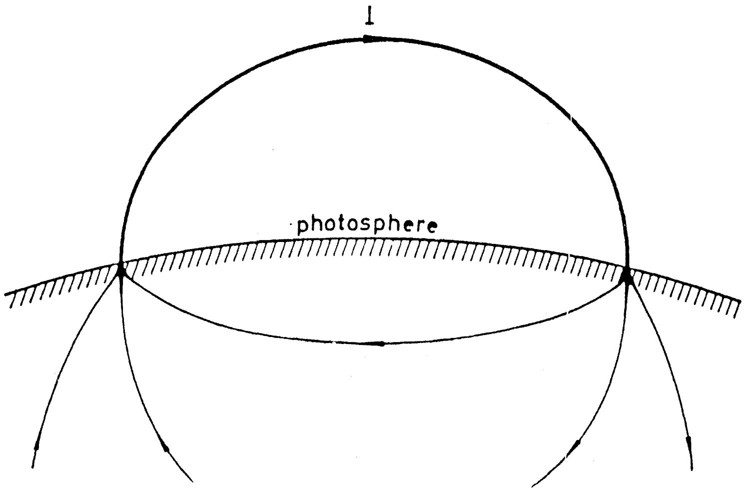


Fig. 1. General pattern of electric currents in the solar atmosphere. The current exists in narrow channels, passing through the solar atmosphere, and is closed in the photosphere or in deeper layers. (Figure and legends from Alfvén and Carlqvist, 1967; reproduced from *Solar Physics*.)

In their paper Alfvén and Carlqvist made an estimate of the magnetic energy of a current of 10^{11} A in a circular loop with a radius of 10^4 to 10^5 km. The inductance of such a loop is about $L = 10H$. Assuming a time constant for the interruption of 10^3 s (at that time believed to be the characteristic time for flare 'ignition') they found a voltage drop over the interruption of $L\,\mathrm{d}I/\mathrm{d}t = 10^9$ V and a magnetic energy of the circuit of $W = 0.5LI^2 = 5 \times 10^{29}$ erg. These numbers were later revised: Carlqvist (1968, 1969) changed the assumed inductance to $20H$ and found that, in order to have a circuit energy of 10^{32} erg, one needs a current of 10^{12} A. An interruption would cause a voltage drop of 8×10^{10} V in a time of 100 s and would accelerate 4×10^{32} electrons and ions to GeV energies. More recently, Alfvén (1981) has stressed the importance of electric double layers: a space charge distribution that gives a potential drop in the layer and a vanishing electric field on each side of the layer. Carlqvist (1986) studied their importance for flare acceleration and found that a current of 3×10^{11} A with an inductance of $100H$ containing a magnetic energy of 5×10^{31} erg would, upon interruption, develop a potential drop of 90 GeV in a characteristic time of 100 s.

A brief comparison with recent observational data: results from solar vector magnetography (Gary *et al.*, 1987) have shown that the total net current in an average flare region is several times 10^{12} A; the average current density is 4×10^{-4} A m^{-2}. We also know (cf. De Jager, 1985b) that the characteristic time for flare 'ignition' (the 'energization' process, during which energy is fed into the flare) must be small, in any case smaller than some 10 s for an average flare burst (the total duration of the impulsive phase, which contains several elementary flare bursts is longer, though), and in many cases shorter than a second, particularly for high-energy events (Kiplinger *et al.*, 1984; De Jager *et al.*, 1987). The energy contained in an average flare is 10^{31} erg.

From this first brief comparison it appears that the interruption assumption is able to explain most of the flare properties, with the exception of the characteristic time which is much shorter than the predicted value. In the next three sections we describe modern flare observations, and in Section 5 we compare the observations with the theoretical models.

2. Flare Ignition: the Impulsive Phase

The Solar Maximum Year (SMY) period, 1979–1981 and subsequent years, was a period of unprecedented increase of our knowledge of solar flares, thanks to the marvellous results obtained with spacecraft such as NASA's Solar Maximum Mission (SMM), the Japanese Hinotori, and NRL's P78-1, a good spacecraft that so dramatically was switched off, being victimized by military experiments. The main findings of the SMM period are summarized in De Jager and Švestka (1987) and in Kundu and Woodgate (1986). The latter book is the condensed synthesis of the results of three 5-day workshops each with some 150 participants from 17 countries. The impulsive phase is dealt with in Chapters 2 and 3 of that publication as well as in some others, particularly Chapter 5. A summary of our present-days knowledge of observations and theories of the impulsive phase of solar flares is given by De Jager (1986). Below we give the essentials.

Images of flares in high energy X-rays (> 15 keV) show the flares to appear first in two or more patches, the 'footpoints', on either side of the magnetic 'inversion line', which is the line where the line-of-sight component of the magnetic vector changes sign (Figure 2). The footpoints brighten up almost simultaneously and are also visible, but less clearly, in soft X-ray images. This observation, as well as those of rapid fluctuations of the hard X-ray flux with time-scales below 1 s, suggest heating by particles or by thermal fronts. Since the footpoints appear to coincide with the places where the electric current reaches maximum but opposite values, it is evident that the heating occurs at the locations where the electric current loop runs through the chromosphere. This supports the 'thick target' model of flare heating, where the footpoint emission results from beam heating, the source of the beam being situated somewhere higher up the loop, while the heating agent propagates along the loops, until it meets the chromosphere.

The heating appears to occur in discrete bursts; these bursts were discovered earlier with ESA's TD1A satellite (Van Beek *et al.*, 1974), and were then called 'Elementary Flare Bursts'. Each of them lasts 5 to 15 s. Observation of the flare's temperature in the footpoints at the very onset of the flare shows it to increase, from the pre-flare values of 10 to 15 MK, to values of > 50 MK. This happens in a time span shorter than a minute. After this phase of initial heating the footpoint temperatures decrease more or less exponentially with an e-folding time of about 90 s. This suggests a very short period of energy injection followed by slower decay. This observation must mean that the burst instability must also last briefly. That several loops must be involved is also obvious since sometimes more than two footpoints brighten up simultaneously (Duijveman *et al.*, 1982).

The model that is, therefore, currently assumed is that of beam heating of the

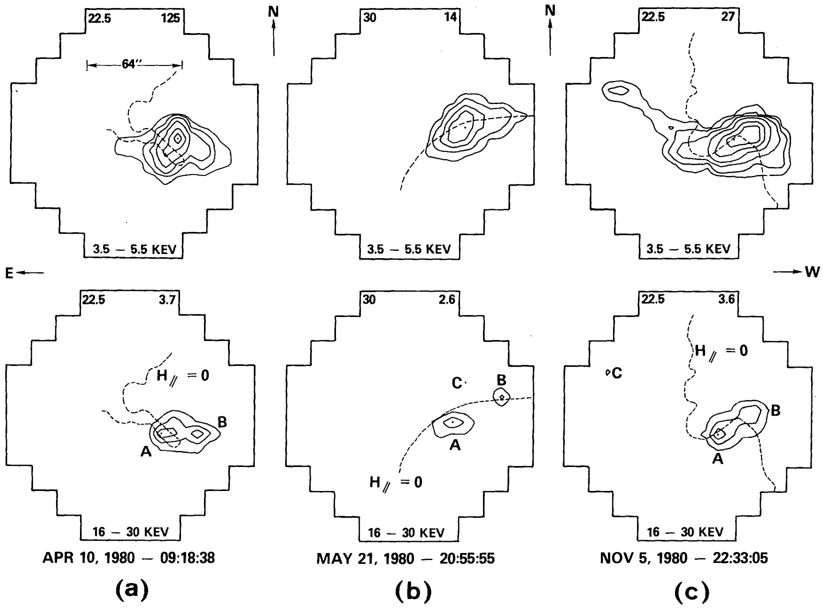


Fig. 2. Hard X-ray images of solar flares show that in the impulsive phase the hard X-ray emission is confined to discrete areas, the footpoints or flare kernals. These are situated on either side of the magnetic inversion line. (From Duijveman *et al.*, 1982; reproduced from *Solar Physics*.)

chromosphere, where the heating agent moves along one or more loops towards the chromosphere and heats it at the spots of impact. A initial picture by Donnelly and Kane (1978) was elaborated for its chromospheric aspects by De Jager (1985) on the basis of various evidences collected by SMM (cf. Figure 3). In that image the heating beam 'evaporates' or 'ablates' the chromosphere. Thus a 'well' originates in the chromosphere, from which heated gas ascends with velocities observed to range between 150 and 450 km s^{-1}. At the bottom of that well a very thin transition layer originates, with a temperature of about 10^4 K. This is the source of the observed flare radiation in Hα and other low-excitation lines observed in the visual part of the flare spectrum.

The above picture is supported by the observed approximate equality of three energies involved:

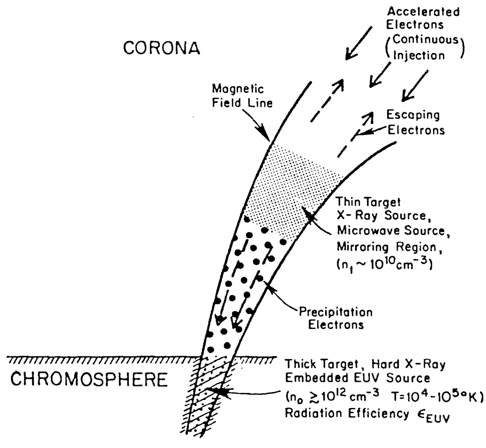
– the energy of the impacting electron beam (if we make the restriction to include only those electrons that are able to penetrate into the chromosphere, in actual practice those with energies above about 15 keV),

– the energy involved in the upward convective motions, and

– the thermal energy content of the plasma that has emanated from the chromospheric 'wells' and has spread over the flare area (the 'coronal explosions').

These energies are about 10^{31} erg for an average flare.

The total number of energetic electrons involved in the chromospheric heating is 10^{38} for an average flare. But there is an hierarchy in the flare-heating processes: a typical



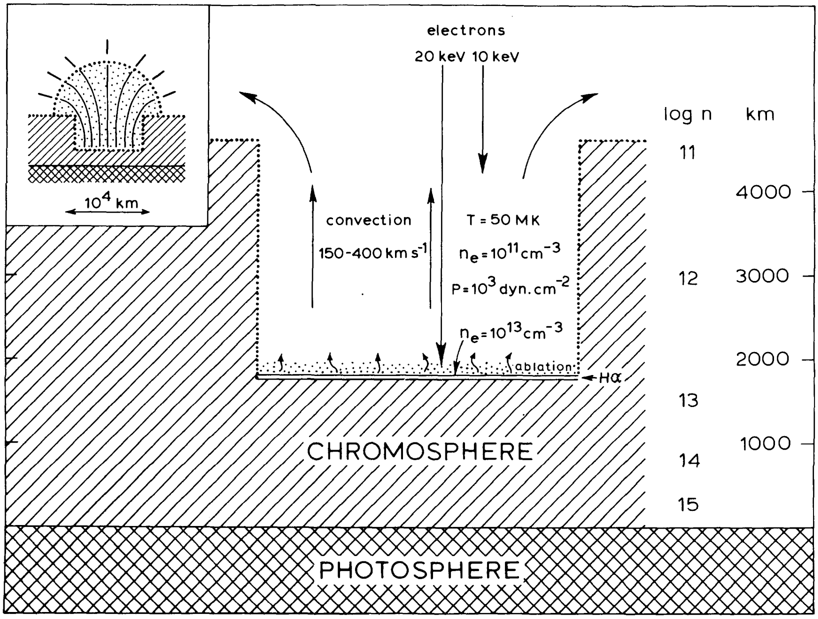


Fig. 3. Heating of the solar chromosphere by a beam of particles. *Above*: the Donnelly–Kane model. *Below*: the De Jager model. (From De Jager, 1986; reproduced from *Space Science Reviews*.)

flare has one or more burst complexes, each with some ten Elementary Flare Bursts, and some 100 to 1000 shorter bursts, down to the 0.1 s level of duration. The shortest bursts have energies of about 10^{28} erg and are due to some 10^{35} electrons of sufficient energy (> 15 keV).

3. Consequences for the Model of the Current Loop System

The description in the preceding section does not address the question *where* primary acceleration takes place and *how* that happens. One important aspect is that flare energization can occur partly (i.e., without fully destroying the flux tube) and repeatedly in the same area. With this latter remark we do not only refer to the homologous flares but also to features such as the Elementary Flare Bursts and the shorter ones, that occur regularly in flares. There are clear indications (see Section 4) that every individual flare burst, whether in X-rays or in microwaves, is the manifestation of a separate acceleration process.

In addition, several authors have drawn attention to the relation between mass motions in active regions and the occurrence in flares. This leads to the suggestion that flares are due to interaction between current threads in the flare area. The interaction between currents is 'driven': the loops are 'frozen' with their feet in the photospheric or chromospheric levels, and are carried along by the strong shearing motions that are observed to occur in the flare area: flux emergence (Martin *et al.*, 1983) or lateral shearing motions (Deszö and Kovacs, 1981; Kundu *et al.*, 1985). This is the way flux tube interaction can be realized. But to that end there must be at least two current loops in the flare.

Observations indicate that there must be many current loops (also called flux tubes) in the flare region. This is shown by various evidences: we mentioned already that three footpoints are sometimes observed; furthermore, the 'filling factor' of the X-ray emitting flare region in the later, gradual, phase of the flare is small, of the order 0.01 to 0.001 (De Jager *et al.*, 1983; Martens *et al.*, 1985; De Jager, 1986, with more references). The X-ray emission in the gradual phase is due to the plasma that streamed out of the flare footpoints after the period of chromospheric impact heating. The two facts: that we see a fairly large area emitting, while the filling factor of the relevant volume is very small, and while the outstreaming plasma must be confined to 'their' fluxtubes, must mean that the flare area outside the footpoints is permeated by a fairly large number of flux tubes. These must all emerge from the flare footpoints: actually it seems that in the footpoints the filling factor is close to unity (De Jager *et al.*, 1983); hence, the flux tubes emerging from the footpoints must diverge outwards and find themselves again close together in the other footpoint. This is our main reason for introducing the 'spaghetti bundle' model for the flux tubes in the pre-flare region. By introducing this idea we suggest that these fluxtubes are more or less randomly intermingled in the region outside the footpoints, which should facilitate their interaction.

We want to call the individual flux tubes *flux threads*, in order to distinguish this concept from the classical fluxtube, of which it was generally thought that there is only one or at most two in the flare. The number of flux threads in the flare is difficult to estimate; Martens *et al.* (1985) find, in one particular case, that there must be some 30 of them.

A few words on the Elementary Flare bursts; these are the individual bursts seen in hard X-ray flux recordings of flares. An interesting feature of these bursts is that they are similar in their time profile for one and the same flare, but that they differ for different flares (De Jager and De Jonge, 1978). For example, their duration may range between 5 and 15 s for different flares. This observation may mean that there must be similarity in the interaction process between fluxthreads in one flare: would that mean succesive interaction between several fluxthreads with only one other fluxthread, as suggested by Kaufmann *et al.* (1980, 1984)? Or should it be an indication for successive interruptions in one double layer, as proposed by Carlqvist (1986)?

4. Inferences from a High-Energy Flare

The flare of 21 May, 1984, studied by Kaufmann *et al.* (1985a, b), Batchelor *et al.* (1985), and De Jager *et al.* (1987) was particular because of its emission of a microwave burst of very high frequency (90 HGz), synchronous with a hard X-ray burst complex (> 100 keV) (cf. Figure 4). The burst complex consisted of at least 13 spikes with life times of about 0.01 s (decay times about 30 ms). The study by De Jager *et al.* (1987) showed that these individual spikes are each associated with the generation of a kernel of extremely hot plasma ($T = 5 \times 10^8$ K), which, because of its high temperature and the consequent intense conductive losses, can only exist very briefly; a conductive loss

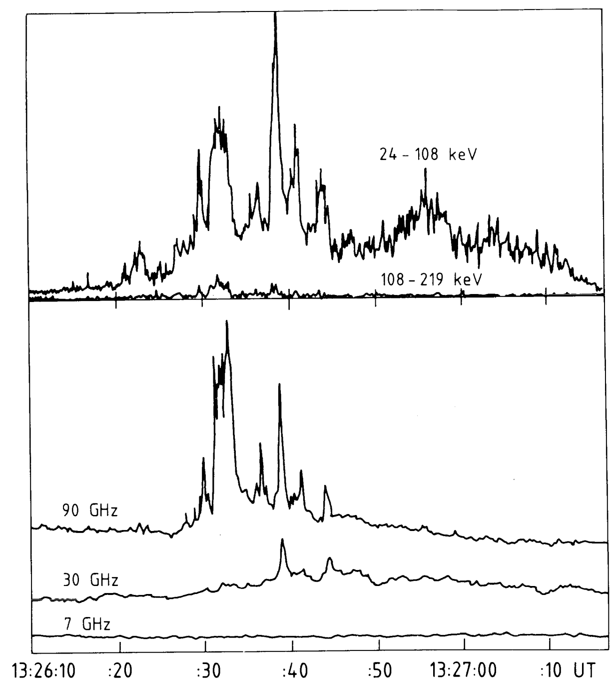


Fig. 4. The solar burst of 21 May, 1984 at 13 : 26 UT, as observed in X-rays by the Hard X-ray Burst Spectrometer aboard the Solar Maximum Mission, and in microwaves at the Itapetinga Radio Observatory. (Kaufmann *et al.*, 1985; reproduced, with some modifications, from *Nature*.)

time of 50 ms was calculated, in rough agreement with the observed decay times of the spikes. During that short time the hot material can only spread over an area not larger than a few hundred km, which is in agreement with the inferred size of the X-ray/microwave emission knots: 350 km. Hence, the location of the emission must be virtually the same as the place where the initial flux threads interact; in this particular case there must have been some thirteen successive flux thread interactions because thirteen spikes had been observed. The reason why this burst complex yielded the observed high temperatures may be related to the exceptionally high magnetic fields at the flare site: between 1400 and 2000 G. To compare: in an average flare particle acceleration occurs in fields around 100 G. In the high-energy case, where the fields are nearly 20 times larger, the induced electric field in a reconnecting sheet $\mathbf{E} = \mathbf{v} \times \mathbf{B}$ (where $\mathbf{v}$ is the Alfvén velocity) is, therefore, nearly 400 times larger.

5. Conclusions

There must be no doubt that electric currents flow through the flare area and that they are seated in many flux-threads. The flare process is related to instabilities in these currents. Are these instabilities double-layer interruptions or field line reconnections?

Our brief comparison of the merits of the two mechanisms and the confrontation with observations shows that the discussion has not yet finished. Both mechanisms have their merits. For assumed values of the currents and of the magnetic energy contained therein – which is equivalent to assuming an 'effective radius' for the circuit – both mechanisms can explain the acceleration to high energies and the recurrence of flare bursts and of flares. That the origin of flares is related to mass motions in the plasma (shear or emergence) can fit into both mechanisms. There is a difference, however, in the calculated characteristic times: this is for suitably chosen energies and currents of the order of 100 s for the interruption model, while observations rather demand times < 10 s for average flares and < 0.1 s for high-energy flares. Reconnection can explain such values because the reconnection time is equal to the speed of reconnection (approximately the Alfvén velocity) times the thickness of the flux-threads involved (assumption: < 1000 km). With a magnetic field $B = 100$ G, and an ambient particle density $n = 10^{10}$ cm^{-3} one finds a time < 0.5 s.

This latter argument is, however, not strong enough for a definite choice of reconnection as the operating mechanism.

Acknowledgement

Thanks are due to Dr P. Carlqvist for critical comments on an earlier version of the manuscript.

References

Alfvén, H.: 1977, *Rev. Geophys. Space Phys.* **15**, 271.
Alfvén, H.: 1981, *Cosmic Plasma*, D. Reidel Publ. Co., Dordrecht, Holland.

Alfvén, H. and Carlqvist, P.: 1967, *Solar Phys.* **1**, 220.
Batchelor, D. A., Crannell, C. J., Wiehl, H. J., and Magun, A.: 1985, *Astrophys. J.* **295**, 258.
Carlqvist, P.: 1968, in Y. Öhman (ed.), 'Mass Motions in Solar Flares and Related Phenomena', *Nobel Symp.* **9**, 193.
Carlqvist, P.: 1969, *Solar Phys.* **7**, 377.
Carlqvist, P.: 1986, *IEEE Transact. in Plasma Sci.* **PS-14**, 754.
De Jager, C.: 1985a, *Solar Phys.* **96**, 143.
De Jager, C.: 1985b, *Solar Phys.* **98**, 267.
De Jager, C.: 1986, *Space Sci. Rev.* **44**, 43.
De Jager, C. and De Jonge, G.: 1978, *Solar Phys.* **58**, 127.
De Jager, C. and Švestka, Z.: 1987, *Adv. Space Res.* **6**, No. 6.
De Jager, C., Kuijpers, J., Correia, E., and Kaufmann, P.: 1987, *Solar Phys.* **110**, 317.
De Jager, C., Machado, M. E., Schadee, A., Strong, K. T., Švestka, Z., Woodgate, B. E., and Van Tend, W.: 1983, *Solar Phys.* **84**, 205.
Deszö, L. and Kovacs, A.: 1981, *Publ. Debrecen Heliophys. Obs.* **4**, 35.
Donnelly, R. F. and Kane, S. R.: 1978, *Astrophys. J.* **222**, 1043.
Duijveman, A., Hoyng, P., and Machado, M. E.: 1982, *Solar Phys.* **81**, 137.
Gary, G. A., Moore, R. L., Hagyard, M. J., and Haisch, B. M.: 1987, *Astrophys. J.* (in press).
Hagyard, M. J. and Rabin, D. M.: 1986, *Adv. Space Res.* **6**, No. 6, 7.
Kaufmann, P., Strauss, F. M., Opher, R., and Laporte, C.: 1980, *Astron. Astrophys.* **87**, 58.
Kaufmann, P., Correia, E., Costa, J. E. R., Zodi Vaz, A. M., and Dennis, B. R.: 1985a, *Nature* **291**.
Kaufmann, P., Correia, E., Costa, J. E. R., and Zodi Vaz, A. M.: 1985b, *Astron. Astrophys.* **157**, 11.
Kiplinger, A. L., Dennis, B. R., Frost, K. J., and Orwig, L. E.: 1984, *Astrophys. J.* **287**, L105.
Kundu, M. R., Gaizauskas, V., Woodgate, B. E., Schmahl, E. J., Shine, R., and Jones, H. P.: 1985, *Astrophys. J. Suppl. Ser.* **57**, 621.
Kundu, M. and Woodgate, B. E.: 1986, *Energetic Phenomena on the Sun*, NASA Conf. Publ. 2439.
Martens, P. H. C., Van den Oord, G. H. J., and Hoyng, P.: 1985, *Solar Phys.* **96**, 235.
Martin, S. F., Deszö, L., Antalová, A., Kučera, A., and Harvey, L.: 1983, *Adv. Space Res.* **2**(11), 39.
Moreton, G. and Severny, A.: 1968, *Solar Phys.* **3**, 282.
Severny, A.: 1964, *Space Sci. Rev.* **3**, 451.
Severny, A.: 1965, *Astron. Zh.* **42**, 217.
Van Beek, H. F., De Feiter, L. D., and De Jager, C.: 1974, *Space Res.* **XIV**, 447.

GLOBAL WAVE PATTERNS IN THE SUN'S MAGNETIC FIELD*

J. O. STENFLO

Institute of Astronomy, ETH-Zentrum, Zürich, Switzerland

(Received 18 September, 1987)

Abstract. When the observed pattern of solar magnetic fields is decomposed in its spherical-harmonic components and a time series analysis is performed, a resonant global wave pattern is revealed. The power spectrum indicates modes with discrete frequencies, obeying a strict parity selection rule in the case of the zonal, rotationally-symmetric modes (with spherical-harmonic order $m = 0$). For instance, the 22 yr resonance that dominates for the anti-symmetric modes (with odd values of the spherical-harmonic degree l) is completely absent for the symmetric modes, which instead exhibit a number of resonances having frequencies increasing with l.

A more traditional way of looking at the evolution of the zonal magnetic pattern is in the form of isocontours in latitude-time space (as in the 'butterfly diagram' of sunspots). We show how this pattern can to a good approximation be represented as a superposition of 14 discrete modes, each with a purely sinusoidal time variation, one mode for each value of l (= 1, 2, ..., 14). This corresponds to the assumption that the true, fully resolved and noise-free power spectrum consists of δ-function peaks, one for each l value.

This approach allows us to analyse the roles of the individual discrete modes in generating the well-known features in the traditional 'butterfly diagrams', e.g., the drift of the sunspot zones towards the equator and the prominence zones towards the poles during the course of the 11 yr cycle. It is shown that these features are accounted for entirely by the odd parity modes with the single, sinusoidal period of 22 yr. The drifts (and thus the arrow of time) are caused by the systematic phase relations between the 22 yr modes. The even modes exhibit an entirely different pattern. Since they have considerably shorter periods, they cause an undulation of the odd-mode contour lines when superposed on the anti-symmetric pattern.

The dispersion, amplitude, and phase relations of the discrete modes are given. It is indicated how they can be used in combination with spectral inversion techniques to determine the depth variation of the parameters in the governing global wave equation.

1. Introduction

Alfvén introduced the concept of magnetohydrodynamic waves in an effort to explain the origin of the sunspots and the 11 yr cycle of solar activity (Alfvén, 1943). These ideas have had a profound influence on the development of astrophysics, and MHD waves have been found to be of importance in all 'corners' of the Universe.

It was Alfvén who in 1965 introduced me to the subject of solar magnetic fields with his ideas that cosmic magnetic fields, in particular the Sun's magnetic field, are extremely inhomogeneous and filamentary in nature (cf. Alfvén, 1967), contrary to the prevailing views. Now it is well established that the magnetic field in the solar photosphere, where it can be measured through use of the Zeeman effect, is in fact extremely intermittent, consisting of fluxtubes with field strengths of 1-2 kG (Stenflo, 1973) in 'quiet' regions anywhere on the Sun, carrying more than 90% of the total magnetic flux recorded in magnetograms (Howard and Stenflo, 1972; Frazier and Stenflo, 1972; cf. also reviews

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 321–336.

by Harvey, 1977; Stenflo, 1984, 1986; and Solanki, 1987). Since the average flux density is only a few G, the strong-field elements cover only a fraction of 1% of the solar surface.

According to the current observational picture there is a coexistence of various spatial and time-scales in the Sun's magnetic field. Although almost all of the observed flux occurs in extremely fragmented form, these strong-field fragments organize themselves into large-scale, global patterns, which evolve over the time-scales of the solar cycle. It is this pattern that is the subject of the present paper.

Since Alfvén's initial MHD theory of the solar cycle, in which the source of the activity was supposed to be located in the core of the Sun, a large number of other MHD theories have been proposed, the most popular ones being the dynamo theories (e.g., Parker, 1955, 1970; Steenbeck and Krause, 1969). These theories aim at explaining the origin of cosmic magnetic fields in general through the interaction between magnetic fields and turbulent plasma motions in a rotating medium. The solar dynamo is of particular importance to these theories, since the Sun serves as a testing ground lending itself to detailed scrutiny. Although initially apparently quite successful (since they had many free, tunable parameters), the dynamo theories for the Sun have in recent years encountered increasing difficulties, and are now further than ever from their goal of being able to explain the solar cycle. This, however, does not rule out that the dynamo theories may eventually be successful. Their general theoretical framework is very broad, and only certain approaches and idealizations have been explored so far (cf. reviews by Schüssler, 1983; and Gilman, 1983).

The cause of dynamo action is the interaction of cyclonic turbulence with magnetic fields. As the turbulence inside the Sun is believed to occur exclusively in the convective envelope, this is where the sources of solar activity are supposed to be located, in contrast to Alfvén's original theory with a much more deeply seated source. The solar core is believed to be in radiative equilibrium with no significant changes occurring over time-scales of the solar cycle. An indication that our knowledge of what goes on below the standard convection zone may need drastic revision has been obtained through the observations of solar neutrinos (Davis *et al.*, 1968). Although the long-standing discrepancy between observed and predicted neutrino fluxes by a factor of 2–3 has recently got a rather plausible explanation in the form of matter-induced neutrino oscillations (cf. Bethe, 1986), recent statistical analyses of the full time series of solar neutrino oobservations by a number of different authors (e.g., Sakurai, 1979, 1981; Haubold and Gerth, 1985a, b; Subramanian and Lal, 1987) have revealed flux variations with periodicities in the range of years, possibly correlated with sunspot activity. Such fluctuations can hardly be explained in terms of neutrino physics. If they should turn out to be statistically significant, as they are claimed to be, the unavoidable conclusion is that there must be large-amplitude temperature variations in the energy-producing solar core over these short time-scales. Such variations are likely to be accompanied by mass motions and other disturbances. If verified, it would have revolutionary consequences for our understanding of stellar interiors.

It seems that time is not yet ripe for a theoretical understanding of solar magnetism without much better empirical guidance than we have had so far. In the past, this guidance has been fragmentary and indirect, since it has largely come from secondary phenomena or proxies, like sunspots and prominences, instead of dealing with the primary parameter, the magnetic field itself. In the present paper we analyse 26 yr of synoptic observations of the Sun's magnetic field carried out daily at the Mount Wilson and Kitt Peak observatories over all solar latitudes and longitudes. We may regard the solar photosphere, where the field is observed, as an outer boundary for the spherical cavity of the solar interior. From the pattern observed at this outer boundary we want to extract the parameters that are vital for the problem of determining the interior structure. This leads to an analysis of the global modes of the Sun.

2. Power Spectrum Analysis of the Global Modes in the Magnetic Field Pattern

For our global mode analysis we have used synoptic observations of the line-of-sight component of the photospheric magnetic field (obtained through recordings of the circular polarization due to the longitudinal Zeeman effect in selected spectral lines). These daily observations were started at the Mount Wilson Observatory on August 8, 1959 (with Carrington rotation No. 1417), and have also been performed (with higher spatial resolution) at the Kitt Peak National Observatory (KPNO) since 1976. As the Sun rotates, all longitudes can be covered during the course of 27.2753 days, which is the defined rotation period of the rigidly rotating Carrington coordinate system (the physical Sun rotates differentially).

Due to the rapid exponential decrease of pressure and density in the outwards direction in the photosphere, strong buoyancy forces act on the magnetic fluxtubes, forcing them to stand up in the vertical direction. Thus, at least in a statistical sense, it is a good approximation to assume that the true direction of the magnetic flux in the photosphere is along the vertical direction (in the largely force-free corona above, the situation is entirely different, but that is not where the field is measured). With this assumption, it is easy to convert through projection the observed line-of-sight component to a vertical or radial magnetic-field component. This and some other corrections that must be applied to the observational data before they can be regarded as maps of the radial magnetic field have been described in detail previously (Stenflo, 1972; Stenflo and Vogel, 1986).

In a global analysis we need to expand the radial magnetic field $B(\vartheta, \varphi, t)$ in spherical harmonics $Y_l^m(\vartheta, \varphi)$, where ϑ is the colatitude, φ the longitude. If the Sun could be observed simultaneously from all directions, we could write

$$B(\vartheta, \varphi, t) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_l^m(t) Y_l^m(\vartheta, \varphi) . \tag{2.1}$$

In reality, due to the sampling of B over the course of one solar rotation, the longitude and time coordinates are not independent of each other, which has to be carefully taken into account when analysing the non-axisymmetric modes (with $m \neq 0$), as done in Stenflo and Güdel (1988). This problem is, however, of no concern to us in the present paper, since we will only be dealing with the zonal, axisymmetric modes (with $m = 0$).

Using the orthonormal properties of the spherical harmonics, we can solve for the time-dependent harmonic coefficients: namely,

$$c_l^m(t) = \int B(\vartheta, \varphi, t) Y_l^{m*}(\vartheta, \varphi)\, \mathrm{d}\Omega\,; \tag{2.2}$$

Fourier transformation

$$\tilde{c}_l^m(\nu) = \int_{-\infty}^{\infty} c_l^m(t)\, e^{-i2\pi\nu t}\, \mathrm{d}t\,, \tag{2.3}$$

and squaring lead to the power spectra $Pc_l^m(\nu)$ of the various modes:

$$Pc_l^m(\nu) = \frac{1}{T_e}\, |\tilde{c}_l^m(\nu)|^2\,, \tag{2.4}$$

where T_e is the effective length of the time series, taking into account apodization and data gaps, as described in Stenflo and Güdel (1988).

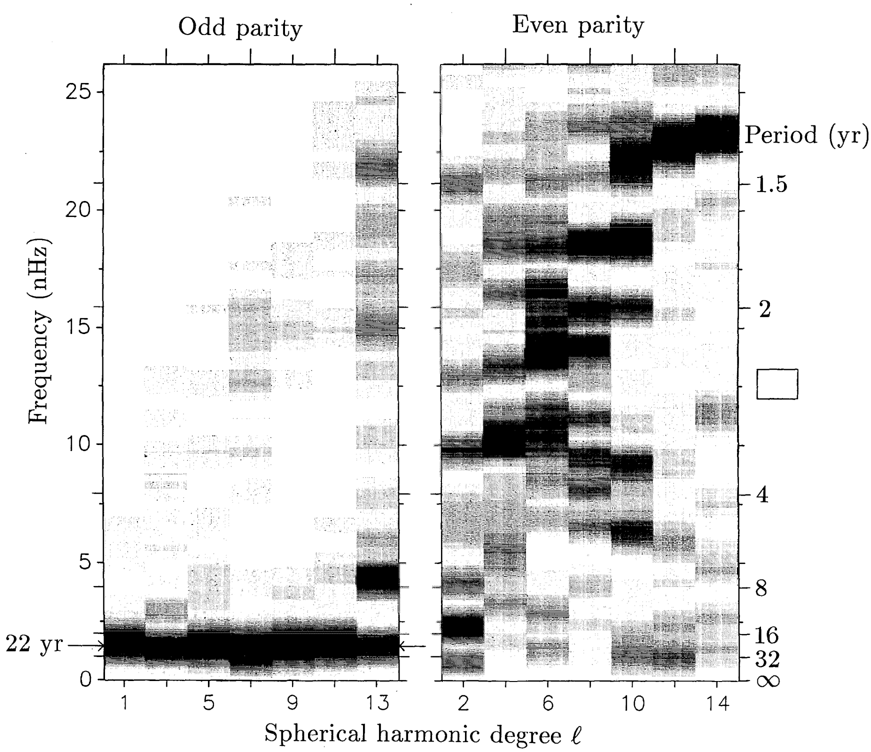


Fig. 1. Power spectra of the spherical-harmonic coefficients for the zonal modes (with $m = 0$) of the radial magnetic field, as a function of frequency and l. Darker areas represent higher power. The left diagram shows the results for the anti-symmetric modes (odd values of l), the right diagram the results for the symmetric modes (even values of l). The resolution is indicated by the small box to the right of the diagram.

The results of such an analysis for the zonal power spectra Pc_l^0, first obtained by Stenflo and Vogel (1986), are illustrated in Figure 1, taken from Stenflo and Weisenhorn (1987). A strict parity-selection rule is governing the behaviour of the modes, which is the reason why we have separate diagrams for odd and even values of l. For each value of l, the gray scale has been normalized such that maximum power corresponds to maximum darkness (for amplitude information, see Stenflo and Weisenhorn, 1987; and Stenflo and Güdel, 1988).

Zonal harmonics with odd values of l are anti-symmetric with respect to reflections in the equatorial plane of the Sun, those with even values of l are symmetric around the equator. We can, therefore, speak of the odd and even *parity* of the patterns.

While the odd modes are dominated by the power spectrum peaks that correspond to a period of 22 yr, independent of l, this period is absent for the even modes, for which the power is concentrated around higher frequencies that increase with the value of l. Since the power amplitudes for the even modes are 5–10 times smaller than those of the odd modes, the background noise in the even-mode diagram has been elevated by the gray-scale normalization procedure.

Figure 1 indicates that we have to do with a pronounced 'emission-line' spectrum, with the power concentrated around a few discrete peaks instead of being smeared out over the diagram. Such a discrete spectrum suggests an underlying origin in terms of global, resonant waves, eigenmodes within the spherical cavity of the solar interior. With an appropriate theory, the observed spectrum could be used to determine the structure of the magnetic field in the solar interior.

Linear plots of power vs frequency shows that harmonics of the 22 yr fundamental mode (for odd l) are almost absent. The power amplitude of the second harmonic at frequency $1/11\ \mathrm{yr}^{-1}$ is less than 10% of the peak of the fundamental mode (Stenflo and Güdel, 1988). The spread and background power in the even-mode diagram may easily be caused by noise due to the limited length, 25 yr, of the time series (which limits the numerical stability and spectral resolution). In the following section we will test a model that assumes the true, noise-free spectrum to be an extreme 'emission-line' spectrum, consisting of one δ-function peak for each value of l, and having frequencies given by the location of the power maxima in Figure 1.

3. The Zonal Pattern as a Superposition of Discrete, Sinusoidal Modes

3.1. Observed Zonal Pattern

By focusing attention on the rotationally-symmetric zonal modes, we avoid all problems connected with the Sun's differential rotation and the non-uniqueness in the choice of longitude system for the Sun. The zonal magnetic-field pattern is obtained by averaging over all longitudes φ, giving $B(t, x)$, where $x = \cos\vartheta$. The resulting matrix can be represented by isocontours in latitude-time space, a representation analog to the 'butterfly diagrams' for sunspots (Stenflo, 1972).

The averaging over all longitudes can be regarded as a smoothing of the data with a rectangular time window of width one Carrington rotation. The results are not sensitive to the width of the time window chosen, except that a wider window (more averaging) leads to reduced noise. Figure 2 shows an isocontour representation of $B(t, x)$ obtained with a smoothing time window having a width of 16 Carrington rotations (1.2 yr).

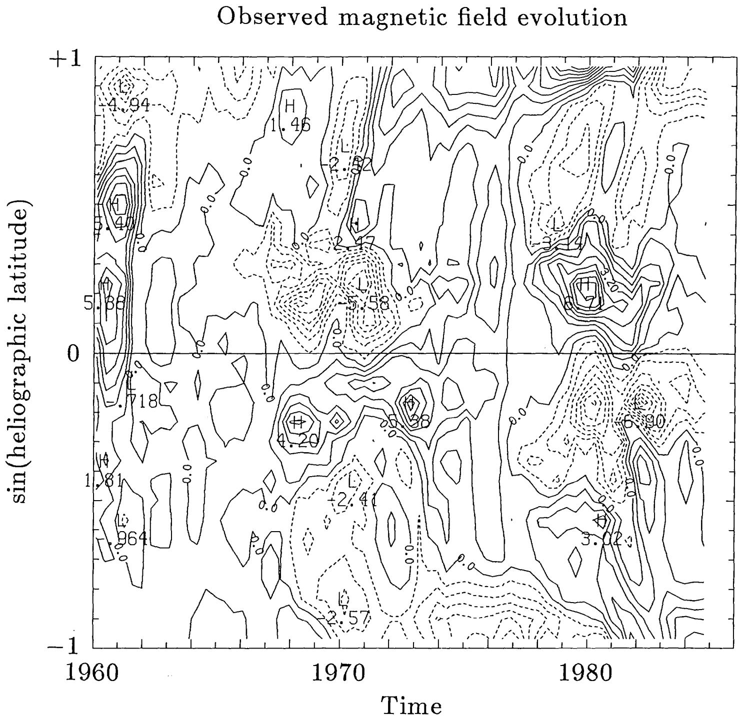


Fig. 2. Evolution of the zonal (longitude-averaged) pattern of the radial component of the observed solar magnetic field, as a function of sin(latitude) and time. A rectangular smoothing window of width 16 Carrington rotations has been applied. The latitude resolution is determined by the grid of 30 equal bins in sin(latitude). Solid (dashed) contours represent positive (negative) magnetic fields, i.e., fields directed out from (into) the Sun. The values for the high (H) and low (L) peak amplitudes are given in units of G.

The observed zonal pattern in Figure 2 is predominantly anti-symmetric, with sign reversal of the polarities every 11 yr, both at low and high latitudes. The lower latitudes are characterized by magnetic patterns drifting towards the equator during the course of the 11 yr cycle, cospatial in latitude-time space with the migrating zones of sunspot activity. The presence of these low-latitude zonal patterns shows that the flux of opposite polarities is not balanced within sunspot groups of active regions, but that a net flux results when we average over the whole activity zone. The polarity of this zonal flux is that of the preceding (westward) spots in sunspot groups. It has been verified that this net flux is insensitive to the sunspot correction used, and should, therefore, be a real property of the activity zones (Stenflo, 1972). The anti-symmetry of the pattern indicates large-scale field line connections (via the corona) between the activity zones in the two hemispheres.

The higher latitudes are characterized by magnetic patterns drifting steeply towards the poles in a way that can hardly be explained in terms of turbulent diffusion (Howard and LaBonte, 1981) as normally assumed in dynamo models. The polewards drifting prominence zones occur along the boundaries between opposite magnetic polarities. During the time around minimum sunspot activity the pattern of a north-south magnetic dipole dominates.

Although the anti-symmetric component dominates, the pattern in Figure 2 cannot be represented without a substantial symmetric component as well. We also notice that there are fluctuations in the shapes of the isocontours indicating variations on time-scales much smaller than 22 or 11 yr. To establish the relationship between these features and the power spectrum peaks of Figure 1, however, requires a fairly elaborate numerical analysis, which will be described next.

3.2. Model and fitting procedure

The power spectrum of Figure 1 has the character of an emission-line spectrum, dominated by a few discrete and rather well-defined peaks. This leads us to introduce a model in which the true power spectrum is approximated by δ-function peaks, one for each value of l. Considering only the rotationally-symmetric zonal modes, and using the proportionality $Y_l^0 \sim P_l$ between the spherical harmonics and Legendre polynomials, we can write Equation (2.1) as

$$B(t, x) = \mathrm{Re} \sum_{l=1}^{N} a_l e^{i2\pi\nu_l t} P_l(x) . \tag{3.1}$$

This is in fact the type of model used for the solution of linearized wave equations in spherical geometry as an eigenvalue problem, e.g., the dynamo equations (cf. Steenbeck and Krause, 1969). The discrete modes are the eigenmodes.

The complex, time-independent coefficients a_l, can be factorized to separate the amplitude and phase factors: i.e.,

$$a_l = |a_l|\, e^{i\phi_l} . \tag{3.2}$$

The phase ϕ_l is related to the relative time lag t_l through

$$\phi_l = -2\pi\nu_l t_l . \tag{3.3}$$

Inserting (3.2) and (3.3) in (3.1), we get

$$B(t, x) = \sum_{l=1}^{N} |a_l| \cos[2\pi\nu_l(t - t_l)] P_l(x) , \tag{3.4}$$

which explicitly shows the zonal field pattern as a superposition of N discrete modes with purely sinusoidal time variations. This model is to be fitted to the observational pattern of Figure 2.

Each mode is characterized by three parameters: amplitude $|a_l|$, frequency ν_l, and time lag t_l. With one mode per value of l, and limiting ourselves to the l range of Figure 1 ($N = 14$), 42 parameters are needed to specify the pattern. In our case we will let the mode frequencies be determined by the center of gravity of the peaks in the power spectrum of Figure 1 (more precisely we have used the center of gravity of the Fourier amplitudes, which represent the square root of the power in Figure 1). This leaves us with two free parameters per mode, thus a total of 28 parameters.

Although such a large number of parameters may at first glance seem excessive for least-squares fitting, this is not so since each mode is fitted separately (the number of free parameters in each fit thus being 2), and the modes represent mathematically orthogonal functions. The nonlinear least-squares fitting procedure of the model to the observed pattern is the following:

In the first step the model uses only the magnetic dipole ($l = 1$) and solves for the two free parameters (amplitude and phase) that minimize the χ^2 deviation between the model and the observed pattern. In the next step (of a loop), the model makes use of the obtained solution for the dipole, but adds the quadrupole ($l = 2$) and solves for its two free parameters. This corresponds to fitting the quadrupole to the residual observed pattern, after the fitted dipole has been subtracted. After subtracting the quadrupole solution, one fits the residual pattern to the $l = 3$ mode, etc., progressively in l until $l = 14$ has been processed.

In principle one could have determined the amplitude and phase of the modes by reading off the values for the integrated peaks in the complex Fourier spectrum (which when squared gives the power spectrum of Figure 1) of the observed harmonic coefficients. However, due to finite spectral resolution the peaks are substantially broadened with extended wings, which redistribute the power, and due to noise there are large phase fluctuations throughout the lines. Therefore, the above fitting procedure gives superior results for the amplitudes and phases under the assumption of discrete modes.

3.3. Results

With the amplitudes and phases determined as explained above, superposition of the 14 discrete modes as in (3.4) results in the zonal pattern of Figure 3, which reproduces the main features of the observed pattern remarkably well, considering the limited number of modes used. Our numerical procedure can be regarded as a 'modal cleaning' of Figure 2, to bring out the resonant components of the observed pattern. As in Figure 2, Figure 3 is dominated by the anti-symmetric pattern, but a weaker symmetric component can also be discerned, as well as undulations of the isocontours with relatively short periods. Our model now allows us to make the individual modal components visible, to clarify their roles in generating this composite pattern.

Let us first separate the symmetric and anti-symmetric components from each other. Figure 4 shows the superposition of all the 7 modes of odd parity, Figure 5 the superposition of the 7 modes of even parity. The sum of the patterns of Figures 4 and 5 of course gives the pattern of Figure 3.

Figure 4 brings out the equatorial and poleward drifts of the magnetic patterns and

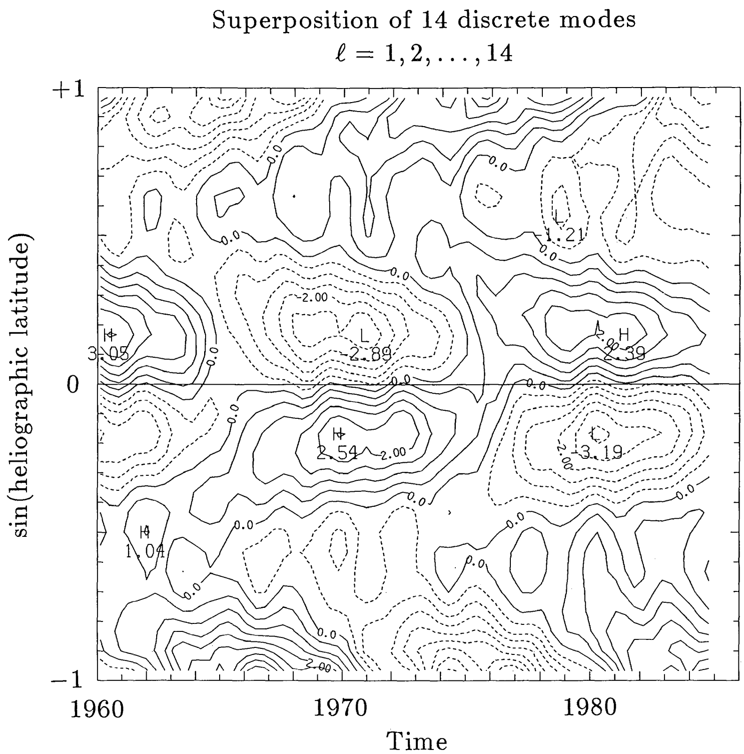


Fig. 3. Synthetic evolutionary diagram, computed as the superposition of 14 discrete (with purely sinusoidal time variation) harmonic modes, one for each value of l. The frequencies of the modes are given by the location of the maxima in Figure 1. The amplitudes and phases of the modes have been determined by a least-squares fitting procedure as described in the text.

their 22 yr periodicity in a very 'clean' way. Note that all the odd modes have periods of about 22 yr (cf. Figure 8 below). The existence of the drifts have nothing to do with the presence of different frequencies in the modal composition, but is exclusively due to the relative phase shifts between the different 22 yr modes. The phase varies systematically with l (cf. Figure 9 below), which establishes the arrow of time here. Each separate mode, having a purely sinusoidal time variation, does not distinguish between the forward or backward time direction.

The symmetric pattern of Figure 5 does not look familiar, since it is normally 'hidden' in the total pattern dominated by the anti-symmetric component. It is composed exclusively of the modes that have frequencies increasing with the value of l, without any 22 yr component. Obvious drifts that characterized the anti-symmetric pattern are absent, although the relative phase relations between the modes show an l dependence similar to that of the odd modes (cf. Figure 9 below). This pattern is responsible for the undulations of the isocontours in Figure 3.

The following diagrams will give us a better feeling for how the pattern is built up by modes of increasing degree l. The upper half of Figure 6 shows the superposition of the

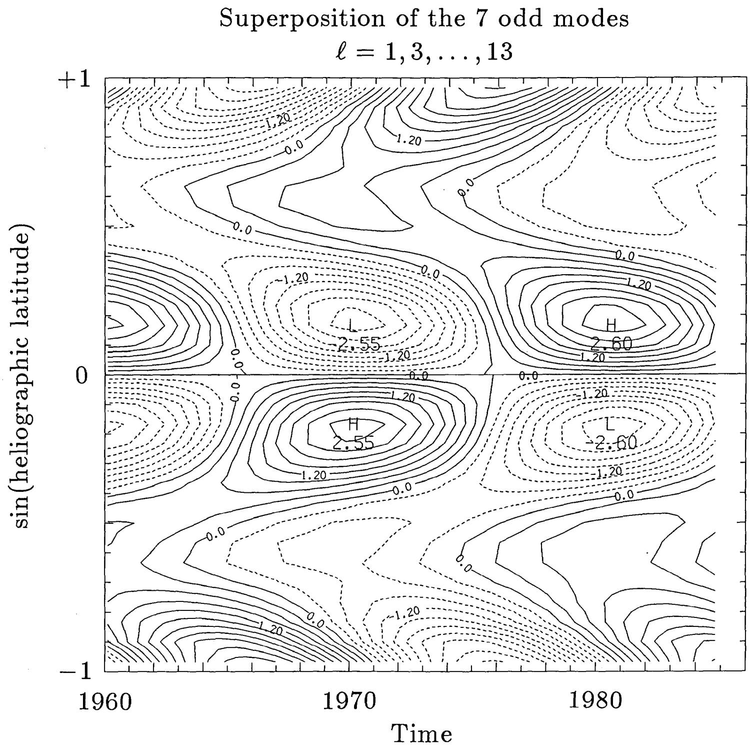


Fig. 4. The anti-symmetric component of Figure 3, obtained if only the 7 modes with odd values of l are superposed, not including any of the even modes.

7 first (odd and even) modes, the lower half the superposition of the remaining modes. With increasing l, smaller and smaller spatial structures can naturally be resolved. In the upper part of Figure 7 the dipole and quadrupole components are shown separately. The quadrupole varies with a shorter period than the dipole. The superposition of the two patterns is shown in the lower left portion of Figure 7. Adding two more modes results in the pattern shown in the lower right portion of Figure 7.

Figure 8 summarizes the 42 parameters of the model. The diagram to the upper left gives the mode frequencies determined by the center of gravity of the power spectrum peaks of Figure 1. The diagram to the upper right shows the fitted mode amplitudes $|a_l|$, the diagram to the lower left the fitted time lags t_l with respect to the epoch 1960.0. Using (3.3), these time lags can be converted to phases ϕ_l, which have been entered in the diagram to the lower right of Figure 8. Due to the phase ambiguity with respect to multiples of 2π, we have in this diagram put all the points inside the interval 0 to 360°.

At first glance the phase diagram may make a rather chaotic impression, but a more careful inspection reveals steep and systematic l-gradients of the phases for both the odd and even modes. Due to the steepness of these gradients, we need to shift the points with $5 \leq l \leq 10$ by 2π, and those with $11 \leq l \leq 14$ by 4π, to bring out the continuous

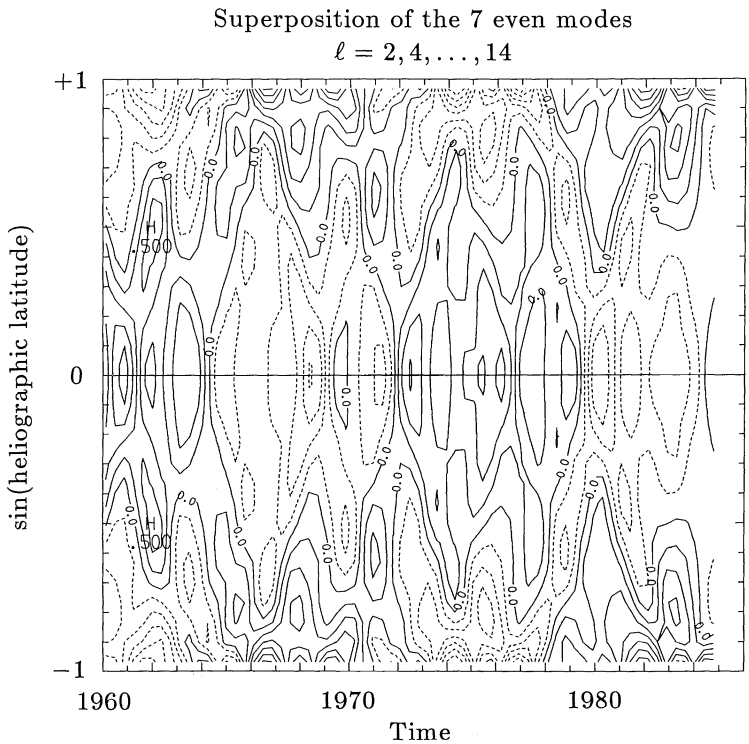
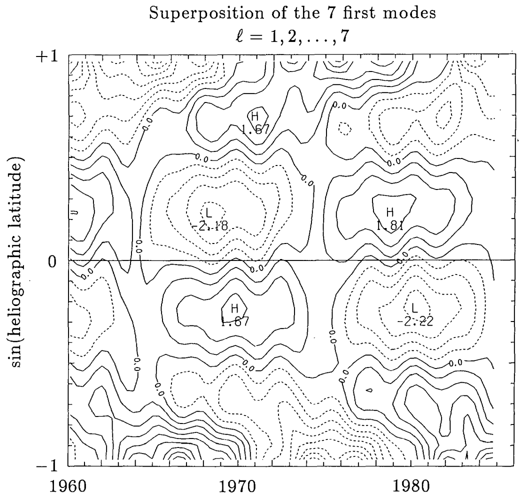


Fig. 5. The symmetric component of Figure 3, obtained by superposing the 7 even modes. Figure 3 is thus the sum of Figures 4 and 5.

phase relations. This has been done in Figure 9. Both the odd and even modes now show a smooth, systematic increase of the phase with l. The circumstance that the gradients of the two phase curves are similar is fortuitous. Since the various even modes have different frequencies (in contrast to the odd modes), the gradient of the phase curve depends on the epoch (1960.0) chosen for the zero point of the phases. For the odd modes there is no such dependence.

4. Concluding Discussion

The circumstance that it is possible to reproduce the observed evolution of the zonal magnetic pattern so well by a single, purely sinusoidal (in time) mode per l value supports the impression from the power spectrum analysis that the spectrum of variability is predominantly a resonant, 'emission-line' spectrum. It appears to be a good approximation to describe the Sun's magnetic field as a linear superposition of discrete global modes. This suggests that a theory representing the magnetic fields as eigenmodes of a linear wave equation should be an appropriate approach to the problem of the magnetic variability of the Sun (and stars).



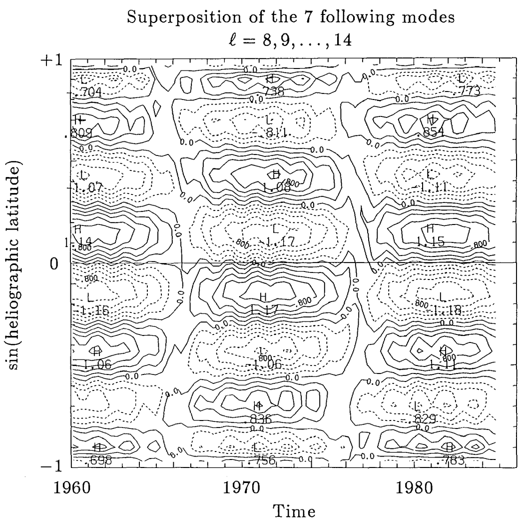


Fig. 6. Contributions of the 7 first (upper diagram) and the 7 last (lower diagram) modes. Figure 3 is the sum of the patterns in the upper and lower diagrams.

Our analysis has revealed a strict parity selection rule for the global zonal modes, and the dispersion, amplitude, and phase relations for the modes have been determined (Figures 8 and 9). In contrast to the l-independent dispersion relation for the odd modes, the even modes are characterized by an increase of the frequency with l. As pointed out by Hoyng (1987), such a dispersion relation may be expected (qualitatively) in linear dynamo theory. If one considers the linear equations of the $\alpha - \omega$ dynamo of Steenbeck and Krause (1969) and disregards the diffusion terms, the dispersion relation for the free dynamo waves is $\omega \sim \sqrt{k}$. As the wave number k is proportional to

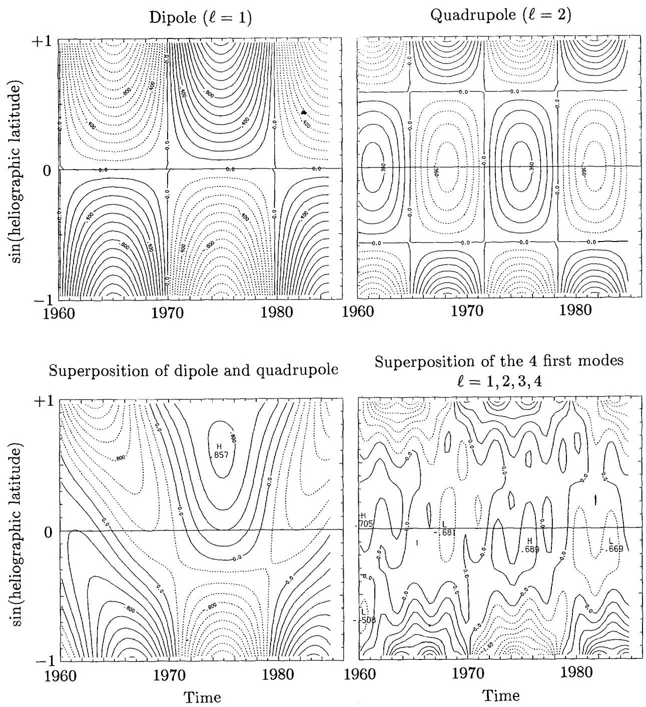


Fig. 7. Illustration of modal contributions to Figure 3. The upper diagrams show the evolutions of the magnetic dipole (*left*) and quadrupole (*right*). The sum of the dipole and quadrupole patterns is shown at lower left. The diagram to the lower right shows the sum of the 4 first modes.

$\sqrt{(l(l+1)}$, this relation can be approximately expressed as $\nu \sim \sqrt{l}$, which is qualitatively similar to the dispersion relation for the even modes in Figure 8.

Another nice feature of the linear $\alpha - \omega$ dynamo is that the equations describing the modes of odd and even parity decouple into two independent sets of equations. Thus a parity selection rule can be directly accomodated by such a theory, but it does not provide an explanation why it actually occurs. The selection rule will be a consequence of the boundary conditions that one postulates, not a consequence of the equations themselves. One possibility is that the symmetry breaking is induced by a weak, fossil seed field that naturally must be of odd-parity, dipole type.

If the correct linear wave equation for the global magnetic field of the Sun could be found, the observed discrete modes may be regarded as the eigenmodes of the wave equation and be used for the solution of the inverse problem, to obtain an empirical determination of the magnetic structure of the solar interior, similar to the spectral inversions of helioseismology (Gough, 1985) or geophysics (Backus and Gilbert, 1967). The mathematical structure of the $\alpha - \omega$ dynamo theory appears promising, since it implies decoupling between the parities and can give the qualitatively correct dispersion

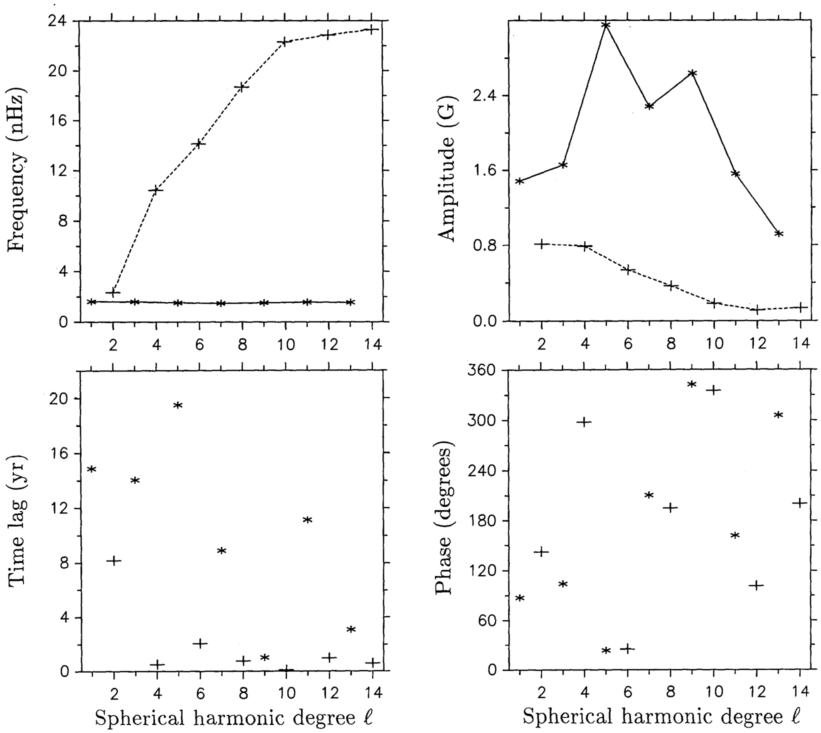


Fig. 8. Variation with l of mode frequency ν_l (*upper left*), amplitude $|a_l|$ (*upper right*), time lag t_l (*lower left*), and phase ϕ_l (*lower right*). Stars and solid lines: odd modes. Pluses and dashed lines: even modes.

relations. However, the three main parameters of the dynamo model, the helicity, turbulent diffusivity, and angular-velocity gradient, must be regarded as unknowns due to our insufficient understanding of convection. It is also not clear whether this parameterization is sufficient or appropriate for a description of the global magnetic-field evolution. Whatever the precise theoretical framework is that has to be used, the depth variations of the parameters of the problem might be derived or at least greatly constrained by the inversion approach. The degrees of freedom of this complex undertaking will be significantly reduced when reliable determinations of the internal angular velocity of the Sun are available from helioseismology inversions.

Acknowledgements

The Mt. Wilson data, which had been recorded under the direction of Robert Howard and subsequently organized in the 30 × 36 array synoptic format at the High Altitude Observatory, had been kindly made available to us by the late Gordon Newkirk. The Kitt Peak synoptic data set had been compiled and made available to us by J. W. Harvey. It is further a pleasure to thank Manuel Güdel for advice on the data reduction.

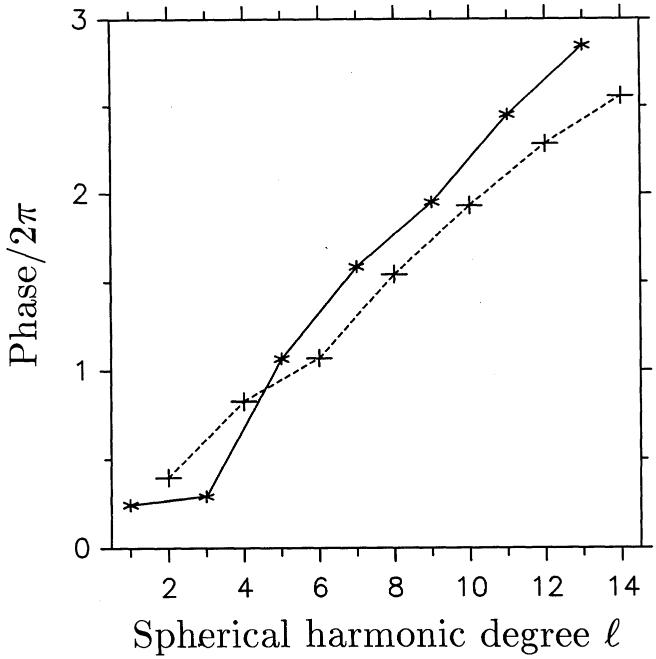


Fig. 9. Mode phase ϕ_l, in units of the period 2π, as a function of l, obtained from Figure 8 by shifting some of the points by multiples of 2π. Stars and solid lines: odd modes. Pluses and dashed lines: even modes.

References

Alfvén, H.: 1943, *Arkiv Mat. Astron. Fys.* **29A**, No. 12.

Alfvén, H.: 1967, in W. R. Hindmarsh, F. J. Lowes, P. H. Roberts, and S. K. Runcorn (eds.), *Magnetism and the Cosmos,* Oliver & Boyd, Edinburgh, London, p. 246.

Backus, G. and Gilbert, F.: 1967, *Geophys. J. Roy. Astron. Soc.* **13**, 247.

Bethe, H.: 1986, *Phys. Rev. Letters* **56**, 1305.

Davis, R., Harmer, D. S., and Hoffman, K. C.: 1968, *Phys. Rev. Letters* **20**, 1205.

Frazier, E. N. and Stenflo, J. O.: 1972, *Solar Phys.* **27**, 330.

Gilman, P. A.: 1983, 'Solar Magnetic Fields: Origins and Coronal Effects', in J. O. Stenflo (ed.), *IAU Symp.* **102**, 247.

Gough, D.: 1985, *Solar Phys.* **100**, 65.

Harvey, J. W.: 1977, in E. A. Müller (ed.), *Highlights of Astronomy* **4**, 223.

Haubold, H. J. and Gerth, E.: 1985a, *Astron. Nachr.* **306**, 203.

Haubold, H. J. and Gerth, E.: 1985b, *Astrophys. Space Sci.* **112**, 397.

Howard, R. and LaBonte, B. J.: 1981, *Solar Phys.* **74**, 131.

Howard, R. and Stenflo, J. O.: 1972, *Solar Phys.* **22**, 402.

Hoyng, P.: 1987, *Astron. Astrophys.* **171**, 357.

Parker, E. N.: 1955, *Astrophys. J.* **122**, 293.

Parker, E. N.: 1970, *Ann. Rev. Astron. Astrophys.* **8**, 1.

Sakurai, K.: 1979, *Nature* **278**, 146.

Sakurai, K.: 1981, *Solar Phys.* **74**, 35.

Schüssler, M.: 1983, 'Solar Magnetic Fields: Origins and Coronal Effects', in J. O. Stenflo (ed.), *IAU Symp.* **102**, 213.

Solanki, S. K.: 1987, in *Proc. Tenth European Regional Astronomy Meeting of the IAU*, Prague, August 24–29, 1987 (in press).
Steenbeck, M. and Krause, F.: 1969, *Astron. Nachr.* **291**, 49.
Stenflo, J. O.: 1972, *Solar Phys.* **23**, 307.
Stenflo, J. O.: 1973, *Solar Phys.* **32**, 41.
Stenflo, J. O.: 1984, *Adv. Space Res.* **4**, 5.
Stenflo, J. O.: 1986, *Mitt. Astron. Ges.* **65**, 25.
Stenflo, J. O. and Vogel, M.: 1986, *Nature* **319**, 285.
Stenflo, J. O. and Güdel, M.: 1988, *Astron. Astrophys.* (in press).
Stenflo, J. O. and Weisenhorn, A. L.: 1987, *Solar Phys.* **108**, 205.
Subramanian, A. and Lal, S.: 1987, *Astron. Nachr.* **308**, 127.

TEMPORAL VARIATIONS OF LOW-ENERGY COSMIC-RAY PROTONS ON DECADAL AND MILLION YEAR TIME-SCALES: IMPLICATIONS TO THEIR ORIGIN*

D. LAL

Scripps Institution of Oceanography, La Jolla, California, U.S.A.

and

Physical Reserch Laboratory, Ahmedabad, India

(Received 21 January, 1988)

Abstract. We discuss the present day information on the prehistoric proton radiation in the solar system at 1 AU based on activation of lunar surface materials, in relation to our present day knowledge of their contemporary fluxes in solar flare events and during quiet time. The bulk of the prehistoric radiation can be attributed to its origin in solar flares. Its energy spectrum is, however, harder than the solar flare radiation observed during solar cycles 19, 20, and 21 (1954–1986). The implications of the data to other sources of particles and/or acceleration mechanisms in the interplanetary region are discussed.

1. Introduction

One of the most significant manifestations of the solar activity is the solar flare in which up to few times 10^{25} J may be released within a few minutes. An appreciable fraction of this energy goes towards acceleration of charged particles to high energies. These particles, designated as solar cosmic radiation (SCR), usually have a soft differential kinetic energy (E) spectrum, $\mathrm{d}N/\mathrm{d}E = \text{const. } E^{-\gamma}$ valid for $1 < E < 100$ MeV, with γ usually lying between 2 and 6 (Biswas and Fichtel, 1965; Lal, 1972; Lanzerotti, 1977; Duggal, 1979). The spectrum is better characterized in terms of particle rigidity R; $\mathrm{d}N/\mathrm{d}R = \text{const. } \exp(-R/R_0)$ over a wider energy interval. The characteristic rigidity R_0 usually lies between 25 and 200 MV (Biswas and Fichtel, 1965; Lal, 1972; Duggal, 1979).

Particle acceleration occurs within the solar system under a variety of conditions; energy can be efficiently transferred to particles from supersonic motions of the background medium via shock waves. Evidence for acceleration of charged particles by collisionless shocks is observed (i) at the Earth's bow shock; (ii) at Jupiter's bow shock; and (iii) at shocks propagating from the Sun through the solar wind (cf. Webb *et al.*, 1985). Shock waves provide a viable accelerating mechanism also for galactic (interstellar) cosmic rays. The first suggestion of acceleration of cosmic rays in stellar winds with terminal shock was made by Jokipii (1968). Axford (1981, 1985) showed that the characteristic supersonic fluid motions can easily provide the cosmic-ray power; often discussed examples of energy sources for the motions are supernova remnants ($\sim 10^{42}$ ergs s^{-1}), and O/B stellar winds ($\sim 5 \times 10^{39}$ ergs s^{-1}).

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 337–346.

Alfvén (1981) has discussed a number of plausible charged particle acceleration mechanisms, e.g., varying magnetic fields, acceleration in double layers and magnetic pumping. He pointed out that the emitted solar wind energy is sufficient to accelerate cosmic rays within the heliosphere to high energies. Reference is made here to his earlier paper (Alfvén, 1949) for the predicted acceleration as well as deceleration of cosmic rays by solar ion streams carrying 'frozen-in' magnetic fields. Although Alfvén (1981) did not present a detailed calculation to determine what cosmic-ray energies and power could be supplied by acceleration processes within the heliosphere, he cautioned against 'blind' usage of the word galactic cosmic radiation (GCR); he suggested instead the use of the term heliospheric cosmic radiation (HCR), for cosmic radiation observed in the solar system. Particle accelerations within the heliosphere are recognized to be important (Barnes and Simpson, 1976; McDonald *et al.*, 1976). In fact, for the July 10–16, 1959, and August 2–10, 1972, high-energy solar events where significant number of particles were accelerated above 1 GeV kinetic energy, it has been suggested that the higher energy particles resulted from the acceleration of ambient low-energy SCR by reflection between two shocks moving with respect to each other in the interplanetary medium (Pomerantz and Duggal, 1974a).

In this paper we examine the time-averaged long-term record of low-energy protons based on observations of radionuclides produced in lunar surface samples in relation to the observed spectra of protons accelerated in solar flares. The lunar record (Lal, 1972; Reedy *et al.*, 1983) provides time-averaged values of flux of protons above 10 MeV kinetic energy, and the characteristic rigidity, averaged over different time periods corresponding to the different radionuclides, e.g., approximately 10^4 yr for ^{14}C, 10^6 yr for ^{26}Al, and 5×10^6 yr for ^{53}Mn. Unfortunately, the record does not go below 10 MeV kinetic energy for protons, the typical threshold energy for producing isotopic changes in matter. However, for charged particles of atomic number ≥ 20, the fossil record of their etchable tracks in lunar materials permits obtaining their energy spectra down to energies of $\lesssim 1$ MeV nucl^{-1} (Lal, 1977; Reedy *et al.*, 1983).

As a result of successful recording of solar flare accelerated particles by the Goddard Space Flight Centre detectors on board IMP-7 and IMP-8 satellites, we now have at hand continuous data for SCR events during the period October 1972–March 1987. Thse data, recently analyzed by Goswami *et al.* (1988), increase the available time series for proton fluxes in solar events by more than factor of two. This data set combined with the earlier proxy and satellite observations summarized by Reedy (1977) provides a continuous time series for 3 solar cycles (1954–1986). We compare these data with the long-term averaged proton spectra based on activation of lunar materials to examine whether the low-energy (10–100 MeV kinetic energies) particle fluxes at 1 AU can be satisfactorily explained as being due solely to their acceleration in solar flares, or else other sources of particles, and/or acceleration mechanisms in the heliosphere are warranted.

2. Solar Flare Cosmic Radiation During Solar Cycles 19, 20, and 21

As discussed above, a long time series is now available for the solar-flare-associated proton fluxes for 1 to > 400 MeV and alpha-particle fluxes for 1–80 MeV nucl^{-1} for the period October 1972–March 1987 (Goswami *et al.*, 1988). For the earlier period, useful proxy and satellite data are available only since 1956. These data have been

TABLE I

Fluences of solar flare protons of $E > 10$ MeV during solar cycles 19, 20, and 21[a]

Solar cycle No. (period)	Peak sunspot No.	No of flare events[b]	Peak event fluence (cm^{-2}) (date of event)		Total proton fluence (cm^{-2}) (11 yr average flux (cm^{-1}))	
			$E > 10$ MeV	$E > 30$ MeV	$E > 10$ MeV	$E > 30$ MeV
19 (1954–1964)	200	32	7.5×10^9 (14 July, 1959)	1.3×10^9 (14 July, 1959)	3.9×10^{10} (110)	8.5×10^9 (25)
20 (1965–1975)	100	39	2×10^{10} (4 Aug., 1972)	8×10^9 (4 Aug., 1972	3.2×10^{10} (92)	9.9×10^9 (28)
21 (1976–1986)	175	45	5×10^9 (23 Sept., 1978)	4×10^8 (23 Sept., 1978)	2.2×10^{10} (64)	1.8×10^9 (5)

[a] For sources of data, see Goswami *et al.* (1988) and Webber *et al.* (1963) as compiled by Reedy (1977).
[b] With event fluence of $> 2 \times 10^6$ for protons of $E > 30$ MeV.

reviewed by Reedy (1977) for protons of $E > 10$ MeV. From these data sets, we can delineate the characteristics of solar flare associated protons of $E > 10$ MeV for the period 1956–1986, covering three complete solar cycles (#19, 20, and #21).

In Table I, we have summarized some of the salient features of SCR during 1954–1986; number of events, peak event fluences (F), and total omni-directional

TABLE II

Spectral characteristics of solar flare protons observed during 1954–1986[†]

Period	Cycle No.	Characteristic rigidity (MV)		
		Minimum	Maximum	Average[a] (standard deviation)
1954–1964	19	39	173	68 (25)
1965–1975	20	24	117	59 (26)
1976–1986	21	22	79	42 (14)

[a] For log-normal distribution the values are 55(27), 52(22), and 48(15), respectively, for cycles #19, 20, and 21, respectively.
[†] For fluxes with event fluences $> 2 \times 10^6$ for protons of $E > 30$ MeV.

fluences and cycle averaged fluxed from all the events for the three cycles. The spectral characteristics are summarized in Table II. As pointed out by Goswami *et al.* (1981, 1988), the SCR events during 1954–1986 have several well-defined features: note that we are considering here only events with fluence F exceeding 10^7 cm^{-2} of protons of energy > 10 MeV:

(i) Both F and R_0 follow log-normal distribution. The distributions of R_0 values can also be represented well by normal distribution (see Table II). However, the F distribution differs markedly from the normal distribution.

The log-normal mean and standard deviation F values for individual cycles lie in the bracket, 6×10^8–1.5×10^9 and 1.7×10^9–3.4×10^9, respectively (Goswami *et al.*, 1981, 1988). In Figure 1 we have plotted the cumulative annual probability for the occurrence of an event with fluence exceeding F. The distribution markedly steepens for $F > 10^9$ cm^{-2}.

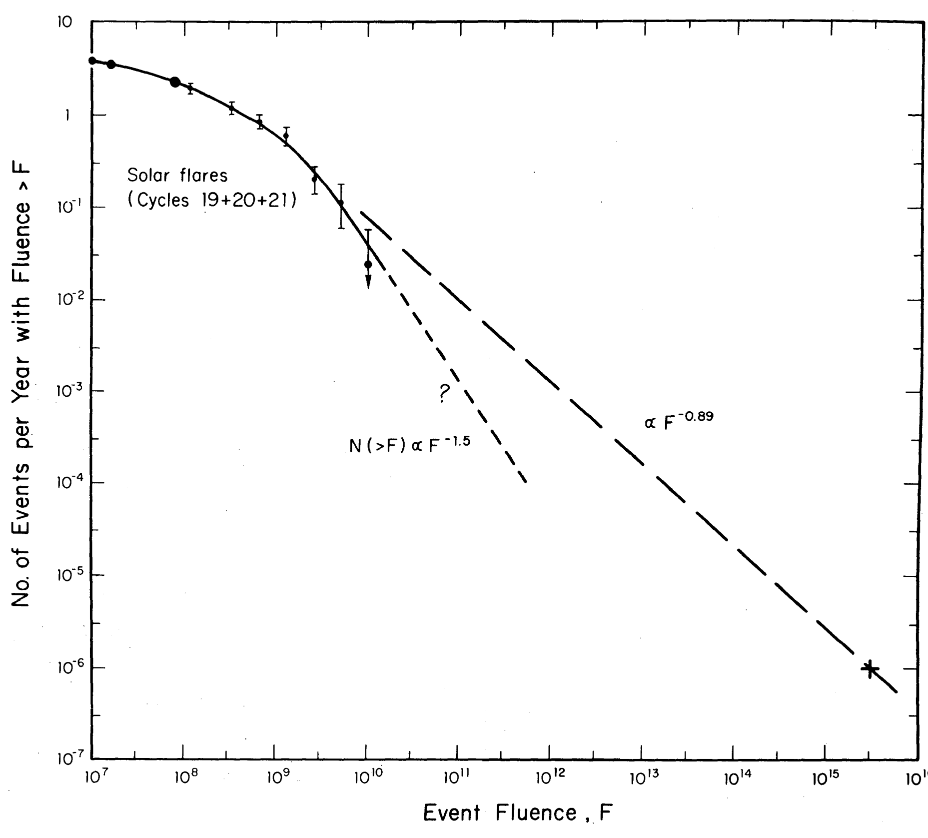


Fig. 1. Cumulative number of events per year with a fluence ($> F$) of protons of $E > 10$ MeV based on solar flare associated proton events during 1956–1986. The cross at 3.5×10^{15} fluence corresponds to the extreme assumption that the long-term averaged flux during the last 1 my was contributed by a single event. See text for the other dotted line.

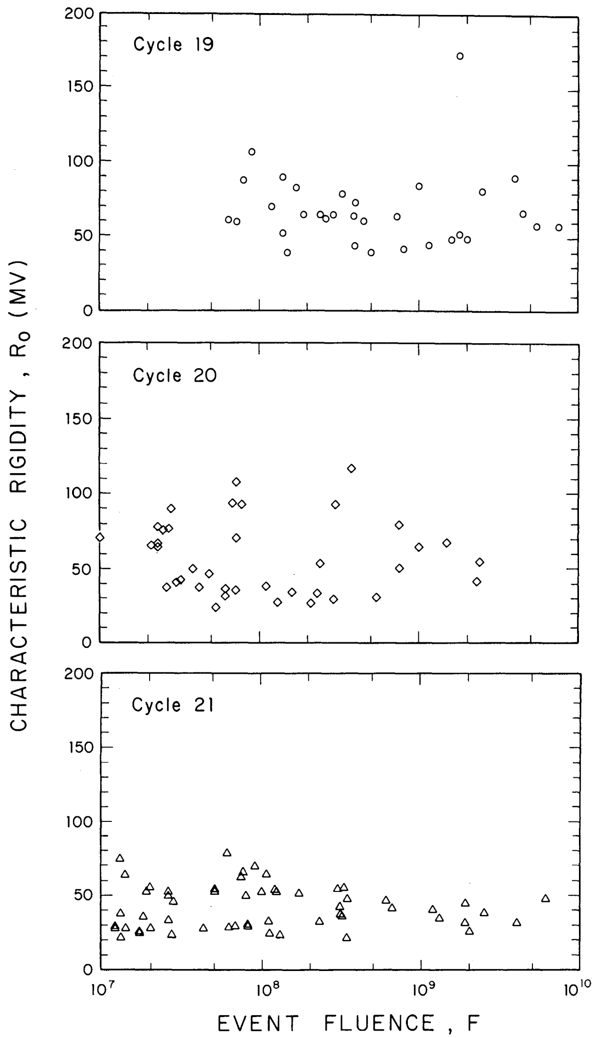


Fig. 2. Observed fluence values are plotted versus characteristic rigidity vlaues for all solar flare events of $F > 10^7$ cm^{-2}, separately for solar cycles 19, 20, and 21.

(ii) Non-parametric test shows that there is no correlation between F and R_0 values in individual events. In any given event of fluence F, all values of R_0 within the range of spread of R_0 values are equally probable. This can be visually seen from Figure 2 where we have plotted F, R_0 values for all the events.

The above features lead to the plausible conclusion that the time series of SCR events is long enough to allow making a reasonable long-term time-averaged prediction for R_0 but not for F.

As regards the time-averaged flux, the log-normal distribution with a large deviation implies that the fluence data observed for the three cycles may differ substantially from the long-term averages. Single events of large fluence can appreciably alter the time-averaged flux for the time period under consideration. This is in fact borne out from the contemporary SCR data. For instance, the flare event of 4 August, 1972 contributed to $\sim 70\%$ of the total fluence in the cycle (Table I).

We will now examine the long-term averaged low-energy proton fluxes and the hardness parameter based on studies of lunar samples. The relevant data are presented in Table III (Reedy, 1980; Goswami *et al.*, 1981, 1983). They represent time averages over periods of the order of 5 my. The spectral form of the time-averaged low-energy proton fluxes sampled by the lunar samples is shown in Figure 3 where we have also shown the quiet time fluxes of protons (GCR), based on published values (Garcia-Munoz *et al.*, 1975).

TABLE III

Contemporary solar flare and ancient low-energy proton fluxes

Period	Time-averaged characteristic rigidity (MV) $(10 < E < 30$ MeV)	Time-averaged flux J (>10 MeV) $(cm^{-2}\,s^{-1})$
Cycle #19	66	110
Cycle #20	85	92
Cycle #21	40	64
Long-term	~ 100	100 ± 25

A general agreement is seen between the contemporary and the million year time-averaged J (flux) and R_0 (hardness parameter) values. During cycle 19, the estimated ground based indirect value of J is probably an underestimate. Another value is 380 $cm^{-2}\,s^{-1}$ (Reedy, 1977) based on lunar sample studies. In this paper we adopt the Earth-based value for proton fluxes in solar cycle 19. Taking this, the range of 64–110 for J $(cm^{-2}\,s^{-1})$ is not inconsistent with the long-term averaged value of 100 ± 25 considering the observed log-normal distribution of F values. As discussed above, a single large event can grossly affect the cycle average. However, the long-term averaged energy spectrum is seen to be considerably harder than that observed during cycles 19, 20, and 21. The mean time-averaged R_0 value for the three cycles is 54 MV with a standard error of 2 MV. This has to be compared with the long-term average value of 100 MV for R_0. The marked difference between contemporary and the long-term average value can be seen from Figure 2 which shows the distribution of R_0 value in solar cycles 19, 20, and 21; there is a paucity of events with $R_0 > 100$ MV. One would expect that unless the solar cycles samples are non-representative, the long-term averaged R_0 value should have been close to 50–55 MV. Also, there seems to be no apparent correlation between peak sunspot number in the cycle and the average R_0 value (Table I). As discussed below, a possible explanation for the observed discrepancy lies in apprecia-ble subsequent acceleration of low-energy solar particles in the heliosphere to higher energies in some of the events.

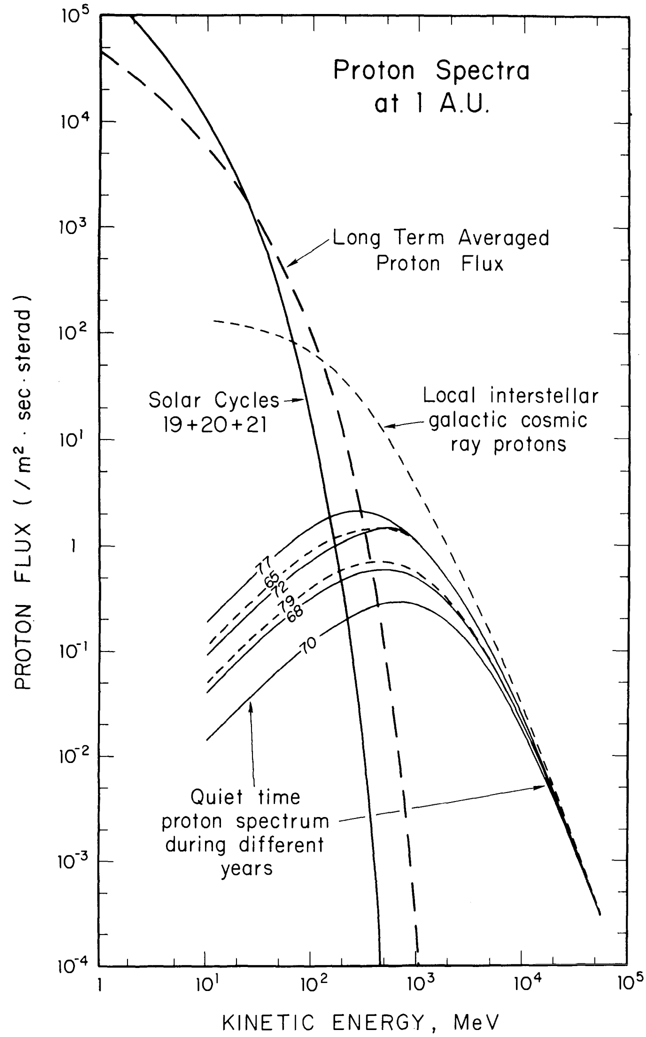


Fig. 3. Observed quiet time differential fluxes of cosmic-ray protons during different years and the theoretically inferred local interstellar proton spectrum for $E > 10$ MeV kinetic energy are shown (Garcia-Munoz *et al.*, 1975; Simpson, private communication). The long-term averaged and contemporary solar flare (cycles 19, 20, and 21) proton spectra are also shown (see text for sources of data). The two curves are normalized to correspond to an integrated omnidirectional flux of 100 protons $cm^{-2}\,s^{-1}$ for protons of kinetic energy $E > 10$ MeV.

3. Discussion and Conclusions

In our discussion of contemporary solar flare events associated with the acceleration of protons to MeV energies (SCR flare events), we have characterized the radiation by two parameters, F (event fluence) and R_0 (hardness parameter). The salient data for SCR events for the three solar cycles, 19, 20, and 21 (Goswami *et al.*, 1988) are presented in Tables I and II. Solar cycle averaged fluxes J and average R_0 values are compared in Table III with the long-term averaged corresponding values deduced from studies of radionuclides observed in lunar samples (^{26}Al, and ^{53}Mn). The generally good agreement indicates that indeed the bulk of the radiation interacting with the lunar samples must be identified with the cosmic radiation accelerated in solar flare events (SCR). The long-term averaged J values ($E > 10$ MeV) are quite consistent with the contemporary SCR events. In fact, the frequency distribution of F values is such that if the long-term averaged values were higher or lower than the contemporary SCR average by an order of magnitude or so, we would consider that within the range of statistical fluctuations. However, the time-averaged hardness parameter R_0 seems to be appreciably different, being softer for the SCR events during the past 35 yrs compared to the million year average. The deviation is appreciable at energies of tens of MeV, as can be seen from Figure 3 where the contemporary SCR spectra and long-term averaged proton fluxes are plotted. Considering the frequency distribution of R_0 in individual SCR events, the deviation is indeed large, outside statistical fluctuations.

Based on the above, we may conclude that either the cycles 19, 20, and 21 are atypical or that there are other sources of protons and/or that some acceleration mechanism is operative on the protons accelerated by the Sun. Fortunately direct observations have been made of proton radiation in the heliosphere (Banes and Simpson, 1976; McDonald *et al.*, 1976; Simpson, 1978; Tsurutani *et al.*, 1982). One observes a persistent flux of low-energy proton radiation (1–10 MeV primarily) corotating with the Sun. Intensity increases of thousand-fold occur at the well-defined concentrated magnetic field regions which recur with the rotation period of the Sun. These nuclei, energized to about 10 MeV kinetic energies bear evidence for acceleration of protons to such energies far beyond the Sun, and confined within the magnetic field spiral. The mechanism of acceleration is now well established to be a shock wave corotating with the Sun far beyond the orbit of Jupiter, also having the contour of the spiral magnetic fields extending out from the Sun (cf. Simpson, 1978).

The corotating flux of (1–10) MeV protons does not, however, constitute an important flux, being several orders of magnitude lower than the time-averaged SCR proton flux. The proton radiation has been continuously monitored in the interplanetary space for more than a decade and no other transient radiation has yet been found other than the corotating particles and the solar flare accelerated protons. The observed deviation between contemporary SCR and the long-term averaged flux of protons could be explained away by the hypotheses that either the solar cycles 19, 20, and 21 are not representative or that we do not yet know enough about particle acceleration mechanisms in the heliosphere. However, it seems to us that if an explanation has to be found in terms of our present data set on low-energy proton radiation and the observed particle acceleration mechanisms in the heliosphere, we can identify the discrepancy as being due to occurrence of the rare events such as those during July 10–16, 1959 and August 2–10, 1972. In these epochs several flares occurred within a week and ambient lower energy solar flare accelerated protons were further accelerated to > 1 GeV kinetic energies by reflection between two shocks moving with respect to each other in the interplanetary medium. These events caused marked ground level enhancements in cosmic-ray flux (Pomerantz and Duggal, 1974b). The proton spectrum as recorded by satellite data for 10–30 MeV interval was also hard; the value of R_0 for 16 July, 1959 and 10 August, 1972 events was 80 and 100 MV, respectively. We have on record only two such events during solar cycles 19 and 20. On a long-term basis, these events could well lead to appreciable modification in time-averaged R_0 values because these are also to high fluence events; these events are responsible for a greater part of the total proton fluence during the cycles in which they occurred.

Thus we propose that solar flare events such as those during July 1959 and August 1972 make a significant impact not only on the time-averaged proton fluxes but also on the spectrum hardness. In these events ambient solar flare accelerated protons get accelerated to high energies in the interplanetary medium. Implicit here is the assumption that we know well enough the characteristic spectra of SCR events (based on observation of contemporary solar flare events), and that an appreciable part of the flux sampled by lunar materials is due to further acceleration of ambient solar flare accelerated protons to high energies in the interplanetary space in rare multiple flare events. Thus the radiation we call as the SCR is in part accelerated in the heliosphere, underscoring the caution voiced by Alfvén (1981), as stated in the introduction.

The analysis presented here suggests a search for higher energy proton record in the lunar samples. There exist possibilities of making such measurements.

Before closing, we would like to allude to the quesiton of probabilities of occurrence of SCR events of different fluences. This question which also has relevance to radiation risks in space applications has been considered by several investigators (Lingenfelter and Hudson, 1980; Wdowczyk and Wolfendale, 1977; van Hollebeke *et al.*, 1975; Webber *et al.*, 1963). We would like to point out that based on the long-term fluxes we can fairly well fix the probability of observing an event of fluence F. Taking the flux of $E > 10$ MeV protons, 100 cm^{-2} s^{-1}, to be contributed by a single event in 1 my, we obtain the point marked by a cross in Figure 1. This gives the following power-law expression for the cumulative annual probability of an event of fluence F:

$$N(>F) = 6.35 \times 10^{7} F^{-0.89} . \tag{1}$$

This expression is clearly an upper limit since we assumed the long-term averaged J to be contributed from one event. A more realistic and plausible expression would be that given by a power law which satisfies the condition that the time-averaged flux of protons of $E > 10$ MeV is 100 cm^{-2} s^{-1}. This expression, valid for $F > 10^{-9}$ cm^{-2} (Figure 1) is deduced to be

$$N(>F) = 4 \times 10^{13} F^{-1.5} , \tag{2}$$

not inconsistent with the experimental data (Figure 1) for $10^{9} < F < 10^{10}$ events.

Acknowledgements

I am grateful to G. Arrhenius, J. N. Goswami, and R. E. Lingenfelter for a critical review of the paper. I would like to take this opportunity to thank my colleagues

R. E. McGuire, J. N. Goswami, R. C. Reedy, and R. Jha for exciting collaboration on studies of contemporary solar flare associated charged particle data.

I gratefully acknowledge a travel grant received for the Smithsonian U.S.–India Foreign Currency Program to visit S.I.O.

References

Alfvén, H.: 1949, *Phys. Rev.* **75**, 1732.

Alfvén, H.: 1981, *Cosmic Plasma,* D. Reidel Publ. Co., Dordrecht, Holland.

Axford, W. I.: 1981, 'Origin of Cosmic Rays', in G. Setti, G. Spada, and A. W. Wolfendale (eds.), *IAU Symp.* **94**, 339.

Axford, W. I.: 1985, n F. D. Kahn (ed.), *Cosmical Gas Dynamics*, VNU Press, p. 119.

Barnes, G. W. and Simpson, J. A.: 1976, *Astrophys. J.* **210**, L91.

Beer, H. F., Käppeler, F., Reffo, G., and Venturini, G.: 1983, *Astrophys. Space Sci.* **97**, 95.

Biswas, S. and Fichtel, C. E.: 1965, *Space Sci. Rev.* **4**, 709.

Duggal, S. P.: 1979, *Rev. Geophys. Space Phys.* **17**, 1021.

Garcia-Munoz, M., Mason, G. M., and Simpson, J. A.: 1975, *Astrophys. J.* **202**, 265.

Goswami, J. N., Jha, R., and Lal, D.: 1981, *J. Astrophys. Astron.* **2**, 201.

Goswami, J. N., McGuire, R. E., Reedy, R. C., Lal, D., and Jha, R.: 1988, *J. Geophys. Res.* (in press).

Jokipii, J. R.: 1968, *Astrophys. J.* **152**, 799.

Lal, D.: 1972, *Space Sci. Rev.* **14**, 3.

Lal, D.: 1977, *Phil. Trans. Roy. Soc. London* **A285**, 69.

Lanzerotti, L. J.: 1977, in O. R. White (ed.), *The Solar Output and Its Variation*, Colorado Assoc. Univ. Press, Boulder, pp. 383–403.

Lingenfelter, R. E. and Hudson, H. S.: 1980, in R. O. Pepin, J. A. Eddy, and R. B. Merrill (eds.), *Proc. Conf. on Ancient Sun*, Pergamon, New York, pp. 69–79.

McDonald, F. B., Teegarden, B. J., Trainor, J. H., von Rosenvinge, T. T., and Webber, W. R.: 1976, *Astrophys. J.* **203**, L149.

Morral, H. and Axford, W. I.: 1983, *Astron. Astrophys.* **125**, 204.

Pomerantz, M. A. and Duggal, S. P.: 1974a, *Rev. Geophys. Space Phys.* **12**, 343.

Pomerantz, M. A. and Duggal, S. P.: 1974b, *J. Geophys. Res.* **79**, 913.

Reedy, R. C.: 1977, *Proc. Lunar Sci. Conf. 8th*, Pergamon, New York, pp. 825–839.

Reedy, R. C.: 1980, in R. O. Pepin, J. A. Eddy, and R. B. Merrill (eds.), *Proc. Conf. on Ancient Sun,* Pergamon, New York, pp. 365–386.

Reedy, R. C., Arnold, J. R., and Lal, D.: 1983, *Science* **219**, 127.

Simpson, J. A.: 1978, *Astron. Aeronautics* **16**, 96.

Tsurutani, B. T., Smith, E. J., Pyle, K. R., and Simpson, J. A.: 1982, *J. Geophys. Res.* **87** (A9), 7389.

Van Hollebeke, M. A. I., Ma Sung, L., and McDonald, F. B.: 1975, *Solar Phys.* **41**, 189.

Wdowczyk, J. and Wolfendale, A. W.: 1977, *Int. Cosmic Ray Conf. 15th Conf. Papers* **5**, 270.

Webb, G. M., Forman, M. A., and Axford, W. I.: 1985, *Astrophys. J.* **298**, 684.

Webber, W. R., Benbrook, J. R., Thomas, J. R., Hunting, A., and Duncan, R.: 1963, *An Evaluation of the Radiation Hazard Due to Solar-Particle Events*, Boeing Co. Report D2–90406 (NTIS Document No AD-805442), 109 pp.

ELECTRODYNAMICS OF THE EUV/X-RAY BRIGHT POINT AND FILAMENTARY FLUX LOOP COMPLEXES*

W.-H. IP

Max-Planck-Institut für Aeronomie, Katlenburg-Lindau, F.R.G.

(Received 24 June, 1987)

Abstract. The electrodynamic model of generation of electric currents in the solar atmosphere, by means of twisting of emerging magnetic flux loops, is investigated with emphasis on the small-scaled EUV/X-ray bright points. It is found that the corresponding power input from such conversion of kinetic energy of the turbulent photospheric plasma into magnetic energy could amount to about 25% of the total energy flux of the solar wind and solar radiation. However, if similar filamentary structures containing colder material are formed in abundance, the total energy budget would be correspondingly larger and the resulting mass injection phenomena may be related to the so-called coronal bullets observed in UV. These energetic features suggest that the coronal dynamics and heating could be dictated by plasma structures with angular sizes $<0.1''$–$1''$. The Solar and Heliospheric Observatory (SOHO) mission will be essential in addressing these issues basic to solar corona and solar wind acceleration.

One interesting development in solar physics is that whenever better spatial resolutions can be achieved, observations always showed that the solar surface is permeated with complex structures with size scales of the order of the angular resolutions reached. As a result, the emerging picture is that the coronal structures and heating process are very much dictated by narrow filamentary magnetic fields in the turbulent solar atmosphere – a view which was first advocated by Alfvén and Carlqvist (1967). One good example is the X-ray bright points investigated in detail by the Skylab soft X-ray telescope (Vaiana *et al.*, 1973; Golub *et al.*, 1974) (see Figure 1). Statistical study by Golub *et al.* (1974) showed that these bright points occurred near-uniformly at all latitudes of the Sun and that flare-like activities were frequently observed with a time scale of a few minutes. These small-scaled X-ray emitting features with a typical dimension of 2×10^8 km^2 are associated with bipolar magnetic field, and when resolved at spatial resolution of $2''$ in the Fe XV 284 Å spectroheliograms, are fould to consist of several miniature closed loop structures with lengths ~ 12000 km and diameters ~ 2500 km (Shelley and Golub, 1979). Since the lifetime of these small loops was estimated to be comparable to the radiative cooling time or conductive cooling time, heating mechanism must exist to maintain the X-ray emitting temperature of 1.3–1.7×10^6 K (Golub *et al.*, 1974, 1980; Shelley and Golub, 1979).

There is perhaps a general agreement that the energy dissipated must be derived from the turbulent twisting of the magnetic flux loops via the photospheric convection motion. With a relative azimuthal velocity V_ϕ turning the both foot points of a magnetic loop (see Figure 2), the time variation of the azimuthal component of the magnetic field B_ϕ

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Fig. 1. A soft X-ray image of the Sun taken from a Sounding Rocket on January 31, 1978. Photography Courtesy from the Solar Physics Group, American Science and Engineering, Inc., Cambridge, Mass., U.S.A.

is related to the longitudinal component through the relation

$$\dot{B}_\phi = B_z \frac{\partial v_\phi}{\partial z} \, . \tag{1}$$

For a projected length l and radius r, the magnetic energy available for heating the loop can be written as

$$E = \frac{\pi^2 l B_\phi^2 r^2}{4\mu_0} \, ; \tag{2}$$

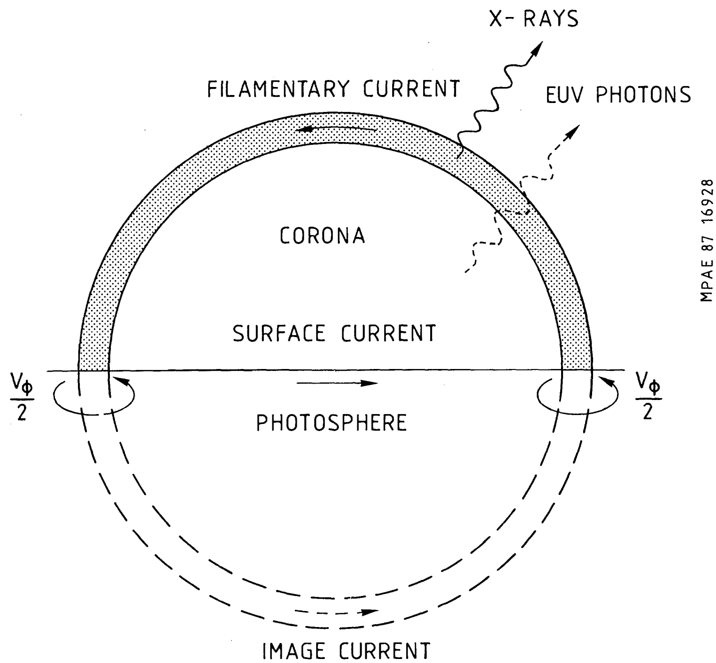


Fig. 2. A model of the flux loop current system associated with the EUV/X-ray bright points. The filament circuit is closed by the photospheric surface current which may be approximated by an image current as shown.

and since $B_\phi/B_z \approx 2V_\phi t/\pi l$,

$$\dot{E} = \frac{2B_z^2 V_\phi^2 r^2 t}{\mu_0 l} . \tag{3}$$

The above equation then approximates the energy storage rate in the twisted loop system. Also, with $I = 2\pi r B_\phi/\mu_0$, we have

$$I = \frac{4B_z v_\phi r_t}{\mu_0 l} . \tag{4}$$

For $V_\phi \approx 1$ km s^{-1}, $B_z \approx 10 - 100$ G, $l \approx 1.2 \times 10^4$ km, $r \approx 1.3 \times 10^3$ km, $I \approx (3.3 \times 10^5 – 3.3 \times 10^6) \times t(s)$ A.

In a model for the loop prominences, Kuperus and Raadu (1974) suggested that the associated current system may be represented by a circuit (Carlqvist, 1972) with the coronal filament current along the prominence connected to the photospheric surface current. The magnetic field generated by the surface current may be further approximated by invoking an image current of the filament current with a upward pointing $J \times B$ Lorentz force (see Figure 2).

In this circuit analog, the self-inductance of the magnetic loop with uniform current density may be approximated as (see also Carlqvist, 1969)

$$L = \frac{\mu_0 l}{2}\left[\ln\left(\frac{4l}{r}\right) - \frac{7}{4}\right]. \tag{5}$$

Note that the time taken for one turn of the flux loop foot point is $t = 2\pi r/V_\phi = 1.5 \times 10^4$ s (4.1 hr) at which point $B_\phi \approx B_z$ and $I \approx 2.3 \times 10^9$–$2.3 \times 10^{10}$ Å. Since from Equation (5), $L \approx 9H$, the stored energy amounts to $E \approx 2.4 \times 10^{19}$–$2.4 \times 10^{21}$ J. According to Golub *et al.* (1974) the production rate of the X-ray bright points with a lifetime ~ 8 hr is on the order of 1500 per day, the corresponding energy supply rate – if each flux loop carries a current of $I \approx 10^{10}$ A on the average – would be $P \approx 3 \times 10^{26} \times N$ ergs s^{-1} where N (≈ 2–3) is the number of miniature loops in each X-ray bright point. In comparison, the total radiative energy loss from the Sun has been estimated to be about 32×10^{27} ergs s^{-1} and similar value for the solar wind energy flux (Kopp, 1972). The X-ray bright points thus account for about 17–25% of the energy budget of the solar corona.

Rosner *et al.* (1978) have investigated the heating of coronal plasma by anomalous DC current dissipation in closed magnetic loops. This requires the occurrence of some plasma instability to generate the turbulence. Consider the case of ion-acoustic instability in which the current carrying electron drift velocity has to be larger than the ion sound speed, i.e., $V_d > V_s$ with $V_s = (3KT_e/m_i)^{1/2}$, then the critical current density will be $j_c \approx 25$–250 A m^{-2} for a plasma density of $n_e = 10^9$–10^{10} cm^{-3}. The average value of current density $\langle j \rangle \approx \langle I \rangle/\pi r^2 \approx 10^{-3}$ A m^{-2} in the twisted flux loops is, therefore, far too small to satisfy this condition. If the continuous heating of the X-ray bright points is taken as evidence of anomalous Joule heating, one possible solution – as first pointed out by Carlqvist (1972) – is to form narrow sheets or filaments with characteristic width given as $r_j \approx (\langle j \rangle/j_c)^{1/2}\, r \approx 10$ km. In this event, the miniature loops would be expected to be resolved into a system of ultrafine-scaled helical structures when observed in EUV or X-ray emission with an angular resolution of less than 0.1″.

Another possible effect of plasma instability in the flux loop is current disruption with energy dissipation concentrated in discrete regions (Carlqvist, 1972, 1979). The potential drop across these areas can be estimated to be $\Phi = L\, \mathrm{d}I/\mathrm{d}t \approx 10$ MV producing a population of suprathermal electrons and ions with energies reaching 10 keV or more (Carlqvist, 1972). Such process may be associated with the flare activities of the X-ray bright points.

The existence of the X-ray emitting flux tubes as well as the explosive events could be accompanied by the formation of twisted flux loops containing cold chromospheric material with temperature $\sim 10^5$ K. As mentioned before, the Fe XV 284 Å spectroheliographic observations have provided evidence for such a combination (Shelley and Golub, 1979). If the ratio of the numbers of cold loops to the hot ones is 10 to 1, the stored magnetic energy will be correspondingly larger making the EUV/X-ray bright point complexes an important source of coronal energy input.

Some relevant information may have also come from the high spatial resolution UV observations by Brueckner and Bartoe (1983) concerning the detection of hight velocity jets (~ 400 km s^{-1}) emitted from the quiet corona. According to these authors, the production rate of such high-energy jets is on the order of 24 events per second over the whole Sun and the power involved is $P \sim 6 \times 10^{17}$ ergs s^{-1} and the mass flux $\dot{M}_r \sim 2 \times 10^{12}$ g s^{-1}. Interestingly, the dimension of these jet-like structures (or coronal bullets) is only about 4000 km, hence, setting a strong constraint on the dimension of their source regions at the chromosphere/photosphere. The continuous acceleration of the coronal bullets has been interpreted in terms of the action of Lorentz force (Karpen *et al.*, 1984) and a model of magnetic propulsion of plasmoids embedded in closed field configuration has been proposed by Cargill and Pneuman (1984). This could come about from magnetic reconnection of the twisted flux loops due to chaotic photospheric motion (see Figure 3(a)). One alternative, following the idea of Mouchovias and Poland (1978) and Carlqvist and Alfvén (1980) is to allow continuous electrodynamic coupling of the ejected material via a system of field-aligned currents as shown in Figure 3(b). In addition to the initial conversion of the stored magnetic energy into kinetic energy of the coronal jets, the continuous twisting of the connected flux loops permits further acceleration.

Making the assumption that there could be a hierarchy of magnetic configurations from prominences to twisted flux loops associated with the X-ray bright points, we

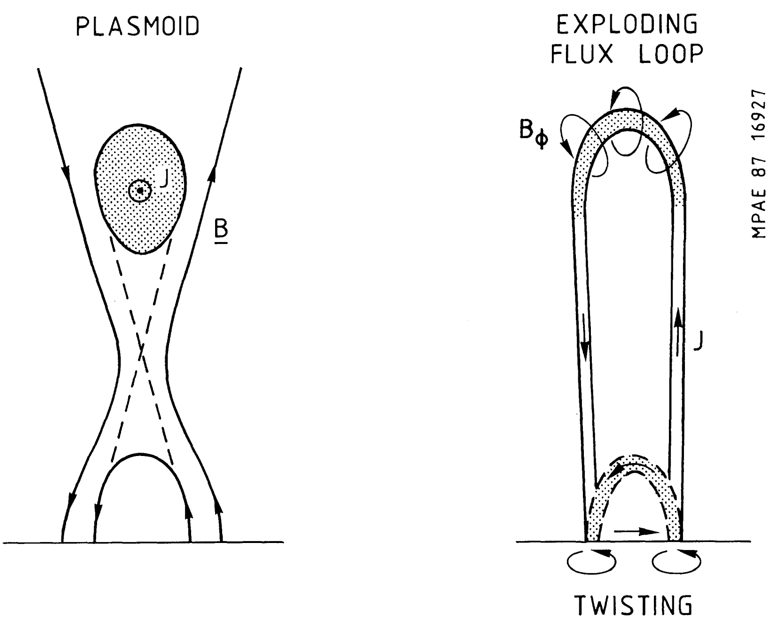


Fig. 3. Two schematic models of the high-speed coronal jets discovered by Brueckner and Bartoe (1983): (a) magnetic propulsion of plasmoid from reconnected field line; (b) exploding current-carrying flux loop. In the latter case, the connection of the ejected plasma to the photospheric foot points provides the continuous acceleration of the coronal jet.

explore, under what circumstances, these small-scaled structures might be related to the energy and mass budgets as well as other dynamical features in the solar corona. The simple approximation invoking an electric circuit analog provides useful insights to the complicated electrodynamic phenomena involved in the conversion of kinetic energy from the photospheric turbulence to coronal heating via magnetic energy stored in twisted flux loops. To maintain continuous heating in closed flux loops, ultra-fine current structures with widths ~ 10 km appear to be necessary. Even though the upcoming Solar and Heliospheric Observatory (SOHO) mission would be able to observe the solar X-ray and UV emissions with angular resolutions not much better than $1''$, the acquired data would be most valuable in addressing the general behaviours and inter-relation between the X-ray bright points and other UV emission features. Such information would be essential in planning future observations capable of resolving the ultra-fine current structures as in a Solar Probe Mission reaching ≤ 20 solar radii from the Sun.

Acknowledgement

I thank Professor Hannes Alfvén for directing my attention to many interesting topics in Solar System Studies including coronal electrodynamics, a field he pioneered. Interactions with him during my study at San Diego have been most educational in many ways for me; and I look forward to learning about his many more exciting new ideas in the future.

References

Alfvén, H. and Carlqvist, P.: *Solar Phys.* **1**, 220.
Brueckner, G. E. and Bartoe, J.-D.: 1983, *Astrophys. J.* **272**, 329.
Cargill, P. and Pneuman, G.: 1984, *Astrophys. J.* **276**, 369.
Carlqvist, P.: 1969, *Solar Phys.* **7**, 377.
Carlqvist, P.: 1972, *Cosmic Electrodynamics* **3**, 377.
Carlqvist, P.: 1979, *Solar Phys.* **63**, 353–367.
Carlqvist, P. and Alfvén, H.: 1980, *Astrophys. Space Sci.* **71**, 203.
Golub, L., Krieger, A. S., Silk, J. K., Timothy, A. F., and Vaiana, G. S.: 1974, *Astrophys. J.* **189**, L93.
Golub, L., Maxson, C., Rosner, R., Serio, S., and Vaiana, G. S.: 1980, *Astrophys. J.* **238**, 343.
Karpen, J. T., Oran, E. S., and Boris, J. P., 1984, *Astrophys. J.* **287**, 396.
Kopp, R. A.: 1972, *Solar Phys.* **27**, 373.
Kuperus, M. and Raadu, M. A.: 1974, *Astron. Astrophys.* **31**, 189.
Mouschovias, T. Ch. and Poland, A. I.: 1978, *Astrophys. J.* **220**, 675.
Rosner, R., Tucker, W. H., and Vaiana, G. S.: 1978, *Astrophys. J.* **220**, 643.
Shelley, N. R. and Golub, L.: 1979, *Solar Phys.* **63**, 119.
Vaiana, G. S., Krieger, A. S., and Timothy, A. F.: 1973, *Solar Phys.* **32**, 81.

PECULIAR SOLAR FLARES AND PLAGES OF POSSIBLE INTEREST FOR THE STUDY OF ELECTRIC QUADRUPOLE LINES OF NaI AND MgII*

Y. ÖHMAN

Stockholm Observatory, Saltsjöbaden, Sweden

(Received 20 March, 1987)

Abstract. Attention is drawn to flare-like plages showing the resonance lines of Na I and Mg II as strong emission lines. Spectra of such objects may prove of interest for the study of electric quadrupole lines of the type $3\,^2S - n\,^2D$.

On July 28, 1966, Gurtovenko (1967) observed a very remarkable flare-like phenomenon on the Sun in the light of the sodium *D*-lines. No satisfactory explanation has been given of this interesting phenomenon, but no doubt the object appeared on the Sun (Gurtovenko, 1971). When trying to find an explanation the author would like to draw attention to the possibility, that grazing-incidence selective reflection may appear in the sodium *D*-lines, if surface structures due to shock fronts appear in the solar atmosphere. Figure 1 indicates how such reflection effects in a shell with an opening in its centre may produce a selective concentration of light in the central area – that is to say, a flare-like area in the sodium *D*-lines. Compare similar dark structures discussed before by the author, particularly near the solar limb (Öhman, 1985).

About 30 years ago the author draw attention to the fact (Öhman, 1958) that solar flares may appear fairly strong in the H and K lines of Ca II, but be less conspicuous in the Hα line. Even in such cases selective reflection may contribute to the intensity, but other mechanisms may be at work as well. Anyhow, if a flare appears strong in Ca II, it would be expected to appear with high intensity in the resonance lines of Mg II as well. The reason why the author wants to draw attention to the possibility of finding electric quadrupole lines in the spectra of Na I and Mg II plages and flares is the following.

Recently the study of electric quadrupole lines in the sodium sequence has attracted great interest among theoretical physicists as well as laboratory spectroscopists. A valuable study of transition probabilities of such lines has been published by Ali (1971), and a number of emission lines of this type have been observed and analysed at the Physics Department of the Lund University by Godefroid *et al.* (1985). In fact, Risberg (1956) already observed at the same institution the $3\,^2S - 3\,^2D$ transition of Na I in emission and found the wavelength 3426.86 Å. The present writer has tried to find this line as a non-identified Fraunhofer line or a chromospheric line, but without success.

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 353–356.

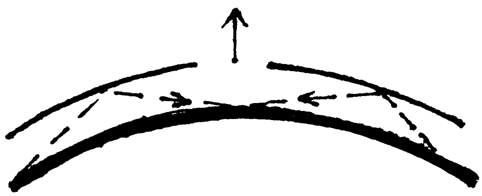


Fig. 1. Schematic drawing showing how selective grazing incidence reflections in a gaseous shell may produce a flare-like object in the centre, where the shell is assumed to be more or less transparent. Even tube-shaped radially-orientated fibrils may transport light in a similar way.

It should be noted here, that there is a misprint in Ali's Table 3, where two wavelengths are given for this transition and without correction to air as well.

The corresponding $3\,^2S - 3\,^2D$ transition of Mg II does not seem to have been observed in the laboratory, probably depending on experimental difficulties, because the line should be expected to be stronger than the Na I line. It is very interesting, therefore, that there are some rather good indications that the Mg II line is present in the solar spectrum in emission. The line should be expected to appear as a close doublet with the wavelengths (vacuum) 1398.77 Å and 1398.79 Å. In the very valuable list of solar lines in this region, recently published by Sandlin *et al.* (1986) we find two not identified emission lines on each side, namely 1398.76 with the intensity 17 in plages and 4 at the limb, and 1398.82 with the corresponding intensities 4 and 7. According to a private communication from Dr R. Tousey, the two wavelengths seem to be good to $\pm$ 0.01 Å and $\pm$ 0.03 Å, respectively. The presence in plages make both these lines very interesting. When trying to make a choice between the two 'candidates' spectrographic studies of flares, where the Mg II resonance lines appear with high intensity, might be of great value.

In a way the E2-lines of Mg II mentioned above resemble the classical He I transition $2\,^3P - 4\,^3F$ studied by Struve many years ago in early B-stars (Struve, 1929). Several investigations have been made in the laboratory of this forbidden line, for instance by Burrell and Kunze (1972), who were able to find plasma satellites as well. This agrees well with the fact that the upper level of the forbidden line is close to the upper level of the allowed transition $2\,^3P - 4\,^3D$. Compare the possible plasma satellites among chromospheric lines discussed recently by the writer (Öhman, 1983; and references given in this publication).

But also for the forbidden lines corresponding in Na I to higher n-values of transitions $3\,^2S - n\,^2D$ close upper levels of allowed transitions have been found to stimulate the appearance of the forbidden lines, and particularly in the presence of electric fields. Here the pioneering work of Segrè (1934) should be particularly mentioned. Already in 1931 a joint work together with Bakker (Sègre and Bakker, 1931) had shown that the $n = 3$ transition showed the Zeeman-splitting of a quadrupole line. The observations were made in absorption. In 1934 he observed himself higher members, and in fact with and without electric fields. Work of similar kind was made by Ny Tsi-Ze and Weng Wen-Po (1936) and particularly by Thackeray (1949).

By using the extension of the Rowland table published by Moore *et al.* (1982) the writer has found an unidentified and very wide Fraunhofer line with the wavelength 2477.85 and intensity 3 which agrees fairly well with Thackeray's line 2477.62, corresponding to $n = 10$ in the series $3\,^2S - n\,^2D$. The great width agrees well with the fact that the allowed transition $3\,^2S - 11\,^2P$ is very close, in fact its wavelength should be 2475.53. Moreover, we have the inter-atomic electric fields. In a similar way it is found that a wide Fraunhofer line with the wavelength 2493.69, also of intensity 3 and not identified, agrees still better in wavelength with Thackeray's line 2493.60 Å, corresponding to $n = 9$. Here a nearby upper level of an allowed transition is also present, in fact corresponding to $n = 10$ and with the wavelength 2490.70 Å. No doubt it would be of great value to confirm these tentative identifications with spectrographic observations of bright sodium plages and flares.

If a a real proof can be found of the presence of E2-lines of Na I and Mg II in the solar spectrum, we have reason to look for them even in stars with peculiar spectra. Stellar spectra showing the resonance lines of Na I and Mg II as strong emission lines may be of special interest. Compare the writer's early search of the forbidden Ca II line $4\,^2S - 3\,^2D$, corresponding to the wavelengths 7291.47 Å and 7323.89, lines which were later on identified in the spectrum of ν Sagittarii by Merrill (Öhman, 1934; Merrill, 1943).

Acknowledgements

The author is much obliged to Professor E. Segrè and Dr R. Tousey for valuable correspondence. He is grateful also for many valuable discussions with Professor I. Martinson, Dr S. Johansson, and Dr C. E. Magnusson at the Department of Physics of the Lund University.

Addendum. In connection with Figure 1 it has been suggested that even tube-shaped narrow filaments may transport light along the tube because of selective reflection at grazing incidence. As shown in *Solar Phys.* **23** (1972) p. 134, observations made by the writer of ball-shaped spectrum features when 'pinch effects' appear in a rotating filament, and when the slit of the spectrograph is parallel to the filament, give strong reason to assume reflection effects of similar kind. In fact, the definition of the ball-shaped structures is so high that the Doppler features must have been produced by nearly monochromatic light. The explanation may be, that selective reflection has not allowed much light of neighbouring wavelengths to escape from the walls of the rotating filament in the direction of the observer. Compare Öhman (1985).

References

Ali, M. A.: 1971, *J. Quant. Spectr. Rad. Trans.* **11**, 1611 (see also Biémont, E. and Godefroid, M.: 1978, *Phys. Scripta* **18**, 323).

Burrell, C. F. and Kunze, H.-J.: 1972, *Phys. Rev. Letters* **29**, 1445.

Godefroid, M., Magnusson, C. E., Zetterberg, P. O., and Joelsson, I.: 1985, *Phys. Scripta* **32**, 125.

Gurtovenko, E. A.: 1967, *Solar Phys.* **1**, 389.
Gurtovenko, E. A.: 1971, *Solar Phys.* **16**, 413.
Merrill, P. W.: 1943, *Publ. Astron. Soc. Pacific* **55**, 242.
Moore, Charlotte, Tousey, R., and Brown, C.: 1982, *The Solar Spectrum 3069 Å–2095 Å,* Washington, D.C.
Ny Tsi-Ze and Weng Wen-Po: 1936, *J. Phys. Radium* **7**, 193.
Öhman, Y.: 1934, *Astrophys. J.* **80**, 171.
Öhman, Y.: 1958, *Arkiv Astron.* **2**, 165, Table 1.
Öhman, Y.: 1983, *Solar Phys.* **85**, 53.
Öhman, Y.: 1985, *Solar Phys.* **96**, 209.
Risberg, P.: 1956, *Arkiv Fysik* **10**, 583 (see also Edlén, B.: 1976, *Mém. Soc. Roy. Sci. Liège*, **IX**(6), 235).
Sandlin, G. D., Bartoe, J.-D., Brueckner, G. E., Tousey, R., and Vanhoosier, M. E.: 1986, *Astrophys. J. Suppl.* **61**, No. 4.
Segrè, E.: 1934, *Rendiconti Lincei* **19**, 595.
Segrè, E. and Bakker, C. J.: 1931, *Z. Phys.* **72**, 724.
Struve, O.: 1929, *Astrophys. J.* **69**, 173.
Thackeray, E. R.: 1949, *Phys. Rev.* **75**, 1840, Table II.

BAND STRUCTURE OF THE SOLAR SYSTEM: AN OBJECTIVE TEST OF THE GROUPING OF PLANETS AND SATELLITES*

GUSTAF ARRHENIUS

Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California, U.S.A.

and

SUSANNE ARRHENIUS

Department of Biology, University of California, Santa Cruz, California, U.S.A.

(Received 24 July, 1987)

Abstract. Alfvén in his early work on the origin of the solar system (1942–1946) noted a pronounced band structure in the gravitational potential distribution of secondary bodies, and suggested this feature to be directly related to the formation process. When the critical velocity phenomenon was later discovered, a close agreement was found between the planet-satellite bands on one hand, and the critical velocity limits of the major compound elements in the interstellar medium on the other, suggesting a specific emplacement mechanism for the dusty plasma which presumably constituted the solar nebula.

Since the originally perceived band structure was outlined in a qualitative fashion, an attempt is made here to analyze the distribution by a statistical technique, testing the significance of clustering of the observational data in the bands. The results show that, with proper scaling of the parameters, such a band structure indeed appears, with features closely similar to those originally conceived. Some deviations are indicated by the cluster analysis, however; their significance is discussed in terms of process involved in the formation of the solar system.

1. Introduction

Alfvén's approach to the early evolution of the solar system is based on two major premises. One is the requirement that modern concepts of the properties of the space medium, established by *in situ* measurements, be used exclusively. Even with this limitation imposed, the variety of possible scenarios remains high. The reason for this is the complexity and nonlinear nature of phenomena involving magnetized plasma in the real world in contrast to homogeneous models commonly used as a first approach in theoretical plasma physics. This complexity is the basis for the second requirement, that only such evolutionary model elements should be accepted, which are derived from observed features, considered to be preserved from the formative era, or representing the formative process today.

Among such observed features there are, in our solar system, particularly two that are invoked as direct evidence for the nature of the process by which matter was or is emplaced in the circumsolar region. The first is what has been referred to as the band structure of the solar system. The second is the cosmogonic shadow effect, manifest in the $\frac{2}{3}$ relationship between the orbital radius of satellite rings and gaps in the Saturnian

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

system; analogous effects are also found in the asteroid belt (Alfvén and Arrhenius, 1976; Alfvén, 1981, 1983, 1984).

Because of the significance ascribed to these two phenomena as observational evidence for the origin and evolution of the solar system, they attracted special attention. The present study is an attempt to test quantitatively the proposed features underlying the band structure concept.

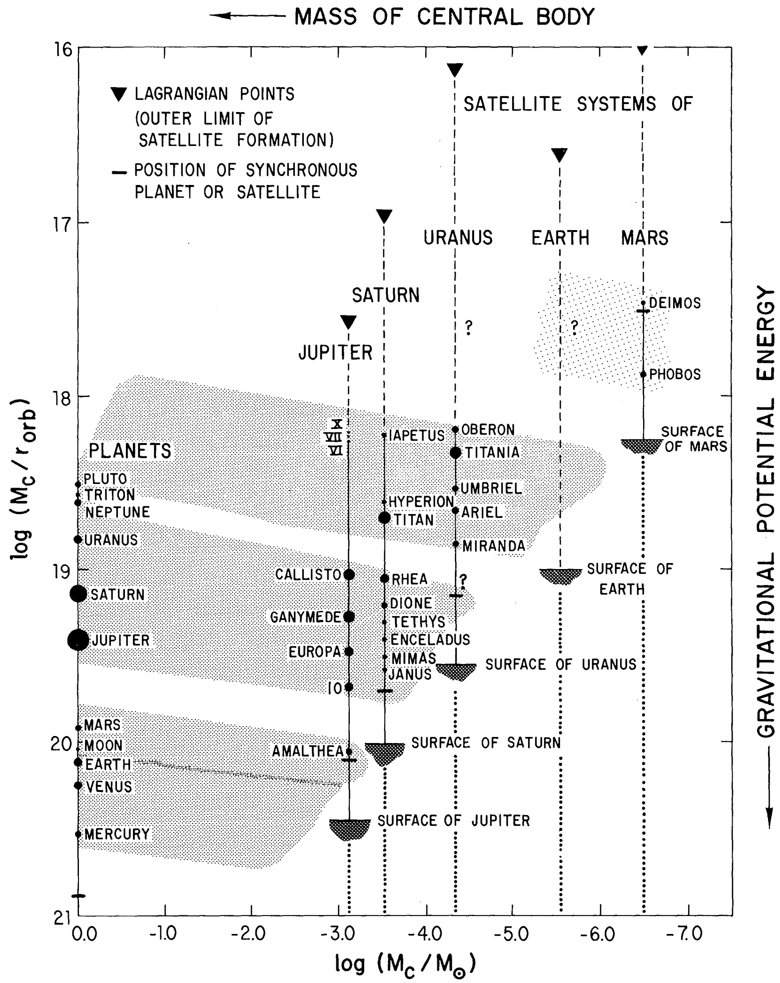


Fig. 1. The band structure of the solar system as outlined in Alfvén and Arrhenius (1975). Among other features, the extension of the C-band through the inner Uranian region suggested the presence of matter there (De, 1972), subsequently discovered as rings and satellies (Elliot *et al.*, 1977; Bhattacharyya and Kuppuswamy, 1977). The formation of a normal satellite system around the Earth and its destruction during the evolution of the orbit of the Moon is discussed in Alfvén and Arrhenius (1972). The modifications of this original concept suggested by cluster analysis are shown in Figures 4–7.

2. The Hetegonic Band Structure and the Critical Velocity Effect

It was noted by Alfvén (1942, 1943, 1946, 1954) that the secondary bodies in the solar system form discrete groups with regard to distance from the central body and that these groups, when they are scaled by gravitational potential and by relative mass of the central body, form bands, separated by gaps where matter is depleted by a factor exceeding 10^{15} (Figure 1). This feature pointed to a common mode of origin of the planets in the solar magnetic field and of the satellites around the magnetized planets.

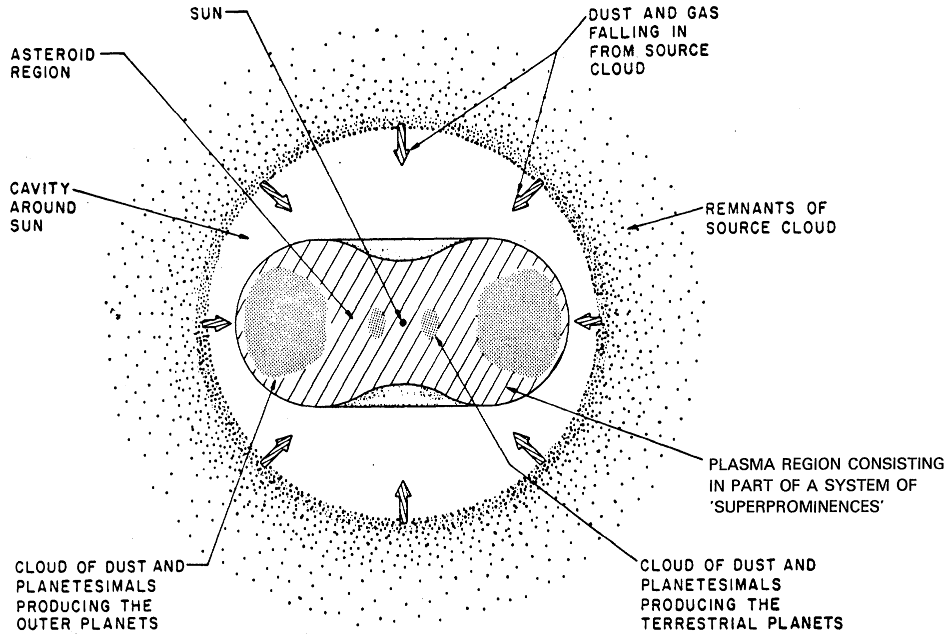
At the time when this zoning of the secondary bodies was first pointed out a physical explanation was lacking. The critical velocity effect, discovered by Angerth *et al.* (1962), provided a suggestive explanation for the band structure as a result of the interaction between a thin magnetized plasma in the circumsolar cavity, the magnetic field of the central bodies, and neutral gas and dust falling in toward them (Danielsson, 1970; Danielsson and Brenning, 1975). Current theories of the critical velocity phenomenon are discussed and referenced by Piel and Möbius (1978), Raadu (1978), Brenning (1982), and Haerendel (1984).

3. Conceptual Origin of the Band Structure

Briefly, the concept of the mechanism generating the band structure can be described as follows. The partially ionized protostellar source cloud (Figure 2) gives rise to the Sun by gravitational settling of grains and gas (Alfvén and Carlqvist, 1978). Neutral gas would continue to fall in toward the proto-Sun from the remaining protostellar cloud, which is assumed to be partially ionized and magnetized in accordance with interstellar cloud observations. While the ionized component of the cloud is suspended by the magnetic field, the neutral component is free to settle gravitationally toward the proto-Sun. As the neutral gas is accelerated in free fall, it eventually reaches a critical velocity v_c defined by:

$$v_c = (2e_i/m)^{1/2},$$

where e_i is the ionization energy and m the mass of the atom being ionized. Ionization at the critical velocity limit consequently occurs at widely different velocities for different atomic species. Since the ratios of ionization potential to atomic mass divides the elements in three discrete groups (Figure 3), separated by wide gaps, the infalling neutral gas would be expected to be chemically segregated, ionized, and stopped in the solar magnetic field in three discrete regions or bands in the circumsolar cavity (Figure 2). Hydrogen and helium would thus penetrate closest to the Sun, forming an ionized cloud of H^+–He^+ plasma with charged dust grains. It should be noted that ionization arises not because of photoexcitation (which in the case of a luminous star may also occur), but due to the critical velocity effect which is unrelated to the photochemical effects. Due to the familiarity of astrophysicists with photoionization, and the less widely recognized hydromagnetic effects in space, it is not uncommon in reviews and discussions of the



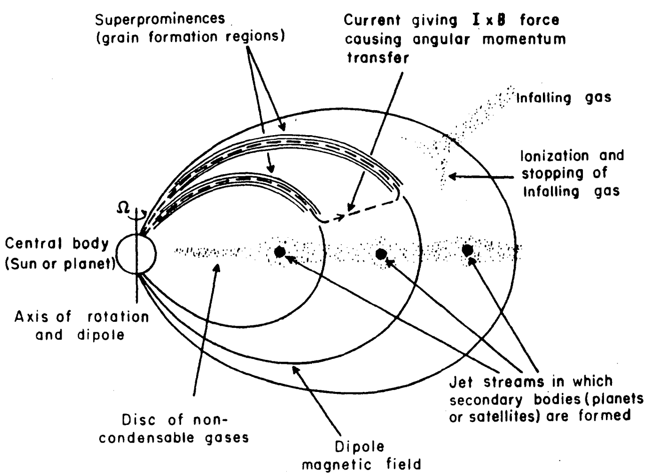


Fig. 2. Schematic representations of the emplacement and angular momentum transfer mechanisms assumed to give rise to the solar nebula in the formative era, and to the stellar disk systems seen in process of formation today. Lower graph from De *et al.* (1977).

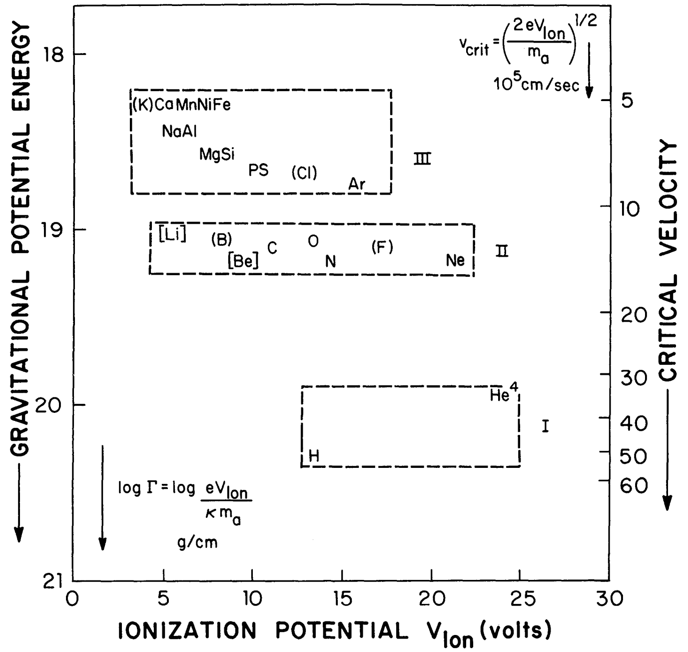


Fig. 3. Gravitational potential energy, critical velocity, and ionization potential of the 23 most abundant elements. Clouds of neutral atoms, molecules and dust falling through a collisionless plasma toward the magnetized central bodies (the Sun in the planetary system; the planets in the regular satellite systems) will beocme ionized at the critical velocity limit and stopped by the magnetic field (cf. Figure 1). The discrete grouping of critical velocities is held responsible for the corresponding band structure (Figure 2) of the distribution of mass around the Sun and the magnetized planets. Much of this mass, and in the case of the H–H_2–He band (A and B clouds in Figure 2) practically all of the small (terrestrial planet + Amalthea) mass now accumulated in solid bodies, must be derived from the dust associated with the gas. The dust immersed in the chemically fractionated plasma, in the critical velocity region, couples dynamically to the plasma, which results in momentum transfer to the dust (Mendis *et al.*, 1982a, b). The strongly reducing atomic and ionized hydrogen in the B cloud may have reacted with the dust to form volatile species such as SiO and MgO, leaving refractory metals, primarily iron, exceptionally concentrated in the solid condensates in this region, accreting to form planet Mercury.

theory to find it misinterpreted as a proposed photoionization process (see, for example, Reeves, 1978).

Another important consideration which has been overlooked and led to criticism in the literature is the important role that the dust component associated with the infalling gas would play in determining the chemical composition of the material accumulating in the band regions. The dust component of the infalling plasma would provide the silicate and metal component, particularly in the inner clouds. This is an important consideration in the case of the hydrogen-helium plasma cloud (A–B cloud) penetrating closest to the Sun, and responsible for the formation of the terrestrial planets (and for Amalthea in the Jovian system) which obviously have retained little of the hydrogen and helium in the emplacement medium. Lack of recognition of the proposed role of dust in the plasma clouds emplaced around the Sun has led (for example, Delsemme, 1980) to criticism of the concept based on the fact that the terrestrial planets do not consist of hydrogen and helium. A fundamental role of the partially corotating circumsolar (and circumplanetary) plasma shells corresponding to the bands would be to orbitalize interstellar dust associated with the plasma at its emplacement. The electrodynamic interaction between interstellar and interplanetary dust with space plasmas in a central gravitational field has been developed by Mendis and his collaborators (e.g., Mendis *et al.*, 1982a, b; see also Mendis, 1981) and has provided a means for a physical understanding of the Saturnian and Jovian ring structures, and for the motion of charged interstellar dust in the heliosphere.

4. Orbitalization of Circumsolar and Circumplanetary Plasma

Alfvén's theory thus provides a solution to the central problem of angular momentum distribution in the solar system, achieved by transfer of momentum from the central bodies, particularly from the Sun, which in this way has lost a major part of its original angular momentum to the planets. The transfer is achieved by the current system connecting the central bodies to the surrounding dusty plasma, which is thereby brought into partial corotation. The secondary bodies forming from this material inherit its orbital angular momentum. The same process is observed today in the corotation of the planetary magnetospheres with the spin of the magnetized planets, and the motion of dust particles influenced by the corotating plasma.

Another important consequence of this process is that solid particles, condensing from or captured in the corotating plasma, will have larger orbital radii ($\frac{3}{2}$) than they will eventually assume in Kepler orbits when they become decoupled from the plasma by growing to such a large size ($\sim 1\ \mu m$) that gravitational forces will dominate over electrodynamic effects. The bodies eventually accreted in the dusty plasma shells will thus eventually move in orbits, which after circularization by accretional inelastic collision will be near circular, with radii or semi-major axes which are $\frac{2}{3}$ of the original emplacement radius (Baxter and Thompson, 1971, 1973). Suggestive evidence of the correctness of this deduction is found in the form of the $\frac{2}{3}$ relationships observed with an accuracy of 1% in 20 documented cases in the Saturnian ring system (Alfvén, 1986).

5. Formation of Planets

As indicated above, a close parallelism is visualized between the formation of prograde satellites around the magnetized planets and the formation of planets around the Sun. The suggestive evidence for operation of the same mechanism in both cases comes from the close similarity in structure of the regular satellite systems and the planetary system, manifest particularly in the band structure, and furthermore by the observation of the $\frac{2}{3}$ relationship both in the planetary system (preserved in the asteroidal belt) and in the Saturnian ring system, where the emplacement process has been suggested to be continuing through recent time (Morfill *et al.*, 1983; Northrop and Connerney, 1987). The development of the regular satellite systems would begin at a time when the accretion of the planets had proceeded to such a stage where their magnetic dipole fields had been developed, and while gravitational infall from the interstellar source cloud was still taking place. For planets with relatively rapid accretion times, this time differential is estimated in the range of 10^7 to 10^8 years, a formation time interval which is in agreement with the radiochronologically based estimates (see, for example, Kirsten, 1978). In the case of the outermost planets with their satellite systems, the much longer time-scale required, comparable to the age of the solar system, poses a problem in all accretion theories which depend solely on classical collision dynamics for accretional evolution (e.g., Safronov, 1954, 1969). It is likely that the solution to this general problem in earlier solar system formation theories lies in the accelerating effects of accretion in dusty plasma which follows from gravitoelectrodynamic considerations (Mendis *et al.*, 1982a, b). The acceleration of planetesimal growth follows from the differential orbital velocity of the faster corotating dusty plasma and the slower planetesimals moving with Kepler velocities. This differential motion is observed today in the Jovian and Saturnian magnetospheres. Applied to the planetesimal accretion process, it has the possible effect of favoring a single major planetesimal growing in each accretion region, rather than a large number of Moon-size objects which appear as a result of purely mechanical-dynamic interactions (Wetherill, 1980). In the latter case the gravitational interaction between these large objects degenerates the system into highly eccentric orbits, differing from the low eccentricity and inclination observed for the regular planets and satellites.

6. Statistical Coherence of Band Structure

Some of the features in the band structure, as shown in Figure 1, have been questioned as ambiguous or arbitrary. Hence, it appears useful to subject the distribution to an objective test. Such a test procedure is provided by cluster analysis (Johnson, 1967), originally designed and extensively used to investigate the significance of discrete grouping of experimental data, particularly in biological data sets.

Cluster analysis is generally used to establish association of data when there is no *a priori* basis for the grouping on a theoretical basis. In contrast, in the present case, theory and observation exist which evoke such grouping, and we wish to use cluster analysis as an objective means for measurement of the cohesion of these groups.

The iterative procedure in cluster analysis starts out by defining each single observation as one cluster. The set of observational data is then used to compute a distance matrix in which the distance between two observations is

$$\mathrm{d}(x_i, x_j) = (x_i - x_j)' \, (x_i - x_j) \, ,$$

where x_i is the ith observation vector.

The two data points (initial clusters) separated by the smallest distance are combined into a new cluster, and the two new clusters closest to each other form the next generation cluster; this procedure is continued to form a hierarchy of increasingly larger clusters, with the ultimate (and trivial) grouping of all data into one single cluster.

The largest distance between an observation in one cluster and an observation in another cluster is used as a measure of the distance between these two clusters. Parameters such as the minimum average and maximum distances between clusters and within individual clusters can be generated and intercompared to give a measure of the statistical meaningfulness of the grouping.

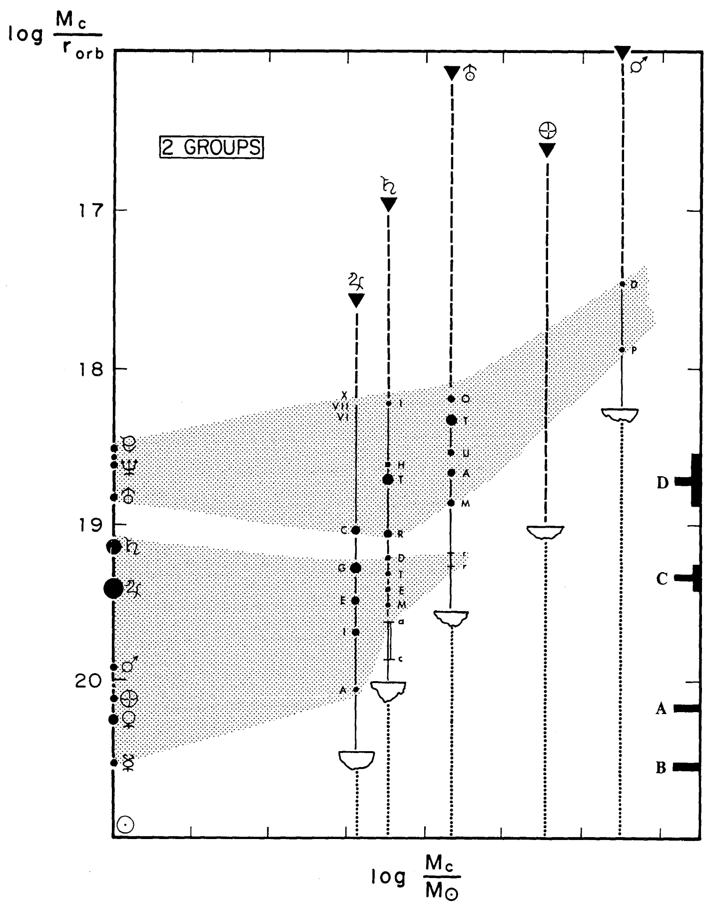


Fig. 4. Cluster analysis of the mass distribution in the solar system; two clusters specified. Under these conditions a first gap forms, separating the outer planets and satellites from the rest (inner part) of the planet-satellite systems. The arguments used in the analysis are $\log M_c$ and $4 \ln G$ in this graph, and in Figures 5–7. The location of the bands A–D on the right-hand side are determined by the critical velocity limits (Figures 2 and 3) multiplied by the factor $\frac{2}{3}$. This orbital reduction arises at the transition of solid grains from plasma-controlled orbits at the critical velocity limit to initially eccentric Kepler orbits, circularized by momentum exchange due to inclastic collisions and thus eventually accreting at $\frac{2}{3}$ of the original orbital radius. The planets are designated by their Christian-astrological symbols; the satellite names, corresponding to the letter markings, are given in Table I. The body inserted between Neptune and Pluto is Triton, now an eccentric satellite of Neptune, but thought to be a captured planet (McCord, 1966). Callisto (C) in the Jovian system and Rhea (R) in the Uranian system were in the original outline (Figure 2) placed in the C-band, but are in the cluster analysis assigned to the lower edge of the D-band.

7. Hierarchy of Planet-Satellite Clusters

Figures 4–6 illustrate the emergence of connectivities as the number of clusters is increased. When only two clusters are specified (Figure 4), the solar system is split at the gravitational potential level of Uranus–Rhea–Miranda. This then appears as the most significant gap in the system.

With three clusters (Figure 5), the Martian satellites Phobos and Deimos are split off as a separate group. At the request of four clusters (Figure 6), a new gap appears between the terrestrial planets and Amalthea on the inside, and Jupiter, the Gallilean satellites, and the inner Saturnian and Uranian satellites on the outside.

This is essentially the subdivision visualized in the original concept (Figure 1), except for the fact that the gap falls between Saturn and Jupiter rather than as previously

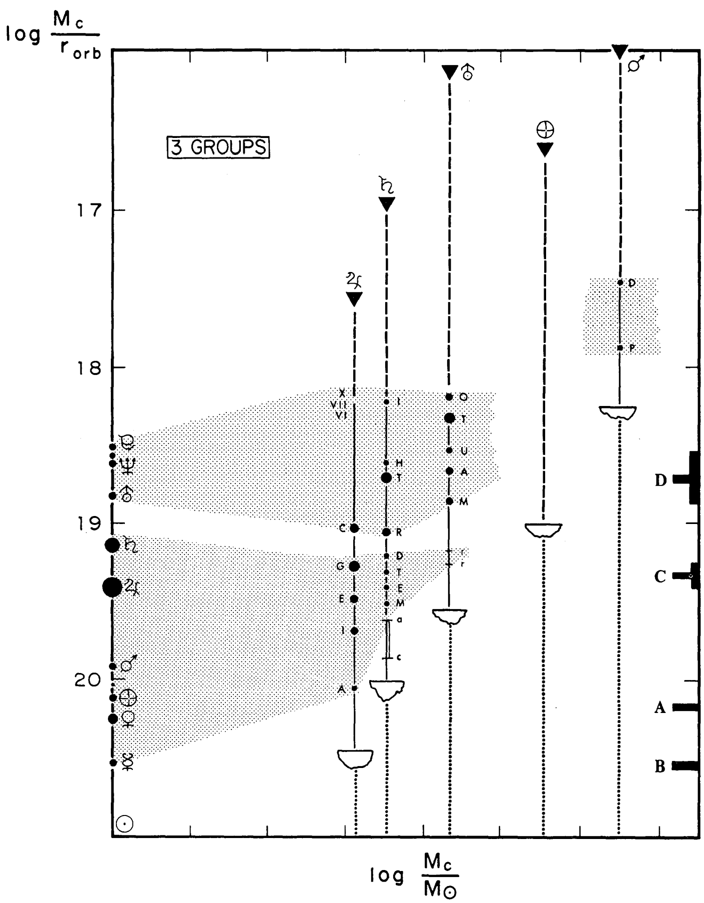


Fig. 5. At specification of three clusters, the (anomalous) Martian moons separate from the two main bands. If original moons of Earth existed and had not been destroyed by the Moon (Alfvén and Arrhenius, 1972), the cohesion of the original outer band (Figure 4) would probably be sufficiently reinforced to cause the new split in Figure 6 to occur already here, rather than at the four-cluster level.

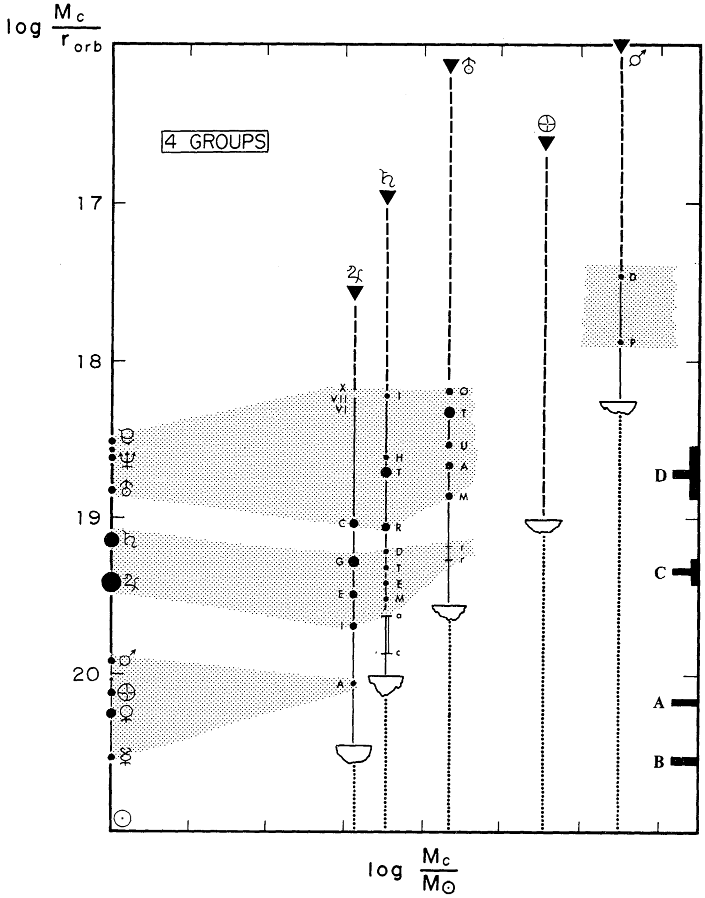


Fig. 6. Clustering in four groups gives rise to the pronounced gap between the terrestrial and giant planets, occupied by the negligible mass of the asteroids and extending between Amalthea and the inner Gallilean satellites Io–Europa–Ganymede (I–E–G) in the Jovian system. This four-band distribution is closely similar to the original one, visualized in Figure 2.

assumed, between the outermost planets. Also, Callisto in the Jovian system and Rhea in the Saturnian system are transferred to the outer ban (D) rather than belonging to the C-band as originally thought. The discovery of the innermost Uranian satellites and rings (Bhattacharyyia and Kuppuswamy, 1977; Elliot *et al.*, 1977) confirms the extension of the C-band into this region, as predicted by De (1972) on the basis of the band structure. The slope of the bands, originally ascribed physical meaning, becomes more indeterminate as a result of the cluster analysis. Furthermore, if the Saturnian ring system is included in the C-band, the A–C gap is somewhat decreased from its originally perceived width, and infalling matter is found to penetrate substantially beyond the critical velocity limit for C, O, N, and Ne in this region.

Further subdivision into five clusters (Figure 7) leads to split-off of the outer satellites

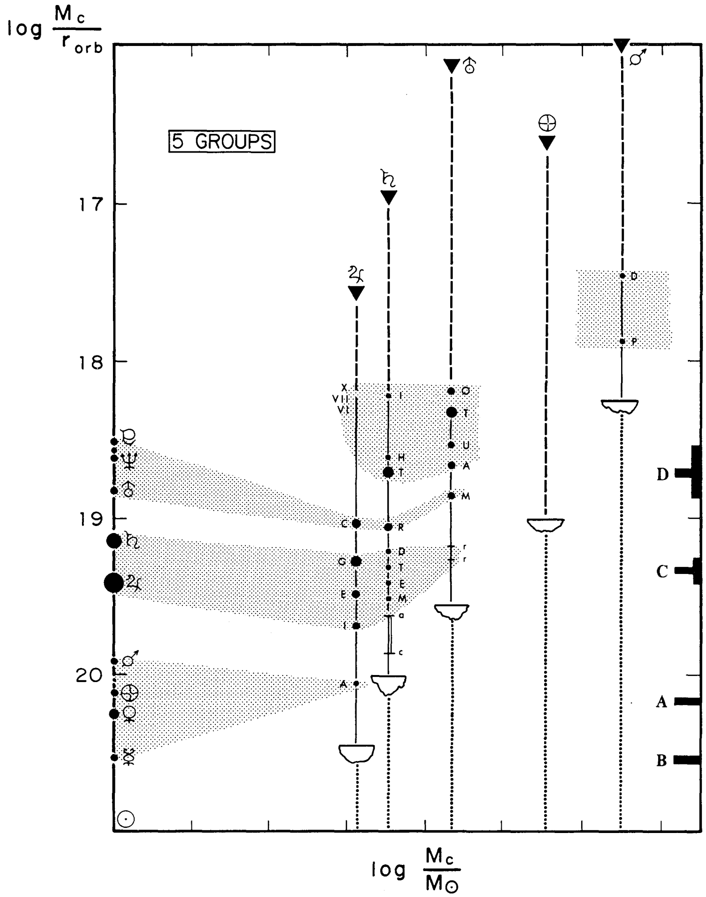


Fig. 7. Introducing a request for yet another cluster results in dissociation of the outer satellites to form a separate group, appearing also in Figure 8. The strong band cohesion between the outer planets, Callisto, Rhea, and Miranda (C–R–M_u) causes them to remain as a separate band which is broken only at cluster 9 by Miranda, at 14 by Saturn and at 28 between Callisto and Rhea, one of the most persistent couples.

from the planetary system, leaving Callisto, Rhea, and Miranda connected to the outermost planet group. At the introduction of six clusters the Saturnian–Jovian complex divides in two, but only at the level of seven clusters does Mercury, representing the innermost (B) cloud, controlled by hydrogen penetration, separate itself from the outer terrestrial planets – Amalthea complex, perceived as formed from the A-plasma-dust cloud, controlled by the critical velocity of helium.

Around 10–11 clusters the band structure coherence between the planetary system and the satellite systems is largely lost at the scaling used here, and the lateral coherence is giving way to local groups, which obviously decrease in size as the cluster number is increased further.

8. Effect of Scaling

For clusters (bands) to develop, such as in gravitational potential – mass space in Figures 4–7, it would appear necessary that the arguments are numerically of the same order of magnitude. If the families of secondary bodies, represented by the vertical lines in the diagrams, were numerically far apart from each other, a cluster in one family would not seek its counterparts in neighbouring families. At the specification of decreasing numbers of clusters, these would thus tend to remain as isolated groups within the individual families, and bands connecting the families would not appear.

In Table I, mass is given in g × 10^{27} and orbital distance in cm × 10^{10}. If, arbitrarily, $\log M$ and $\ln G$ (from Table I) are used as input parameters in the cluster analysis, the distribution shown in Figure 8 results. The outer satellites in this case form a group separate from the rest of the system at the level of two clusters, and the rest of the satellites split off from the planetary system when three clusters are specified.

If, however, the values for $\ln G$ are multiplied by increasingly larger coefficients, a

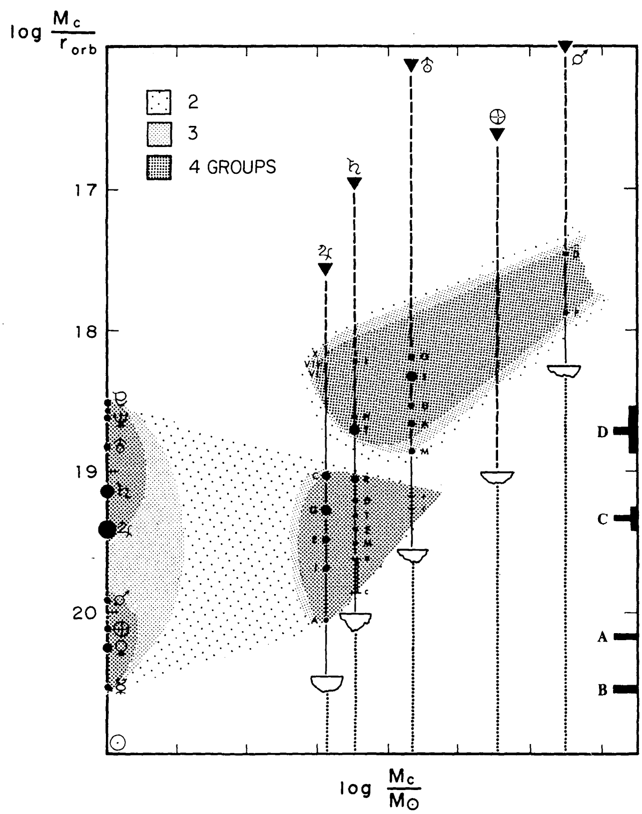


Fig. 8. Cluster analysis using as arguments $\log M_c$ and $\ln G$. In this scaling the prominent features include a persistent split of the inner and outer satellites from each other and from the planetary system, which at the level of four clusters divides in two groups.

TABLE I

Mass of central bodies, and orbital radius and gravitational potential of secondaries

No.	Central body	Mass, M_c ($g \times 10^{27}$)	$\log M_c$	No.	Secondary body	r_{orb} ($cm \times 10^{10}$)	Gravitational potential ($\ln G = \ln(M_c/r_{orb})$)
1	Sun	2 000 000	6.30	1.1	Mercury	579	8.15
				.2	Venus	1 080	7.52
				.3	Earth	1 500	7.20
				.4	Mars	2 280	6.78
				.5	Jupiter	7 780	5.55
				.6	Saturn	14 300	4.94
				.7	Uranus	28 700	4.24
				.8	Neptune	45 000	3.79
				.9	Pluto	59 000	3.52
5	Jupiter	1 899	3.28	5.13	inner ring	1.282	7.30
				.1	Amalthea	1.81	6.96
				.2	Io	4.22	6.11
				.3	Europa	6.71	5.65
				.4	Ganymede	10.7	5.18
				.5	Callisto	18.8	4.62
				.6	6	115	2.80
				.7	10	117	2.79
				.8	7	119	2.77
6	Saturn	568	2.75	6.00	D-ring (inner edge)	0.11	6.74
				.01	C-ring (inner edge)	0.73	6.03
				.02	A-ring (outer edge)	1.37	5.93
				6.1	Mimas	1.86	5.72
				.2	Enceladus	2.38	5.48
				.3	Tethys	2.94	5.26
				.4	Dione	3.77	5.02
				.5	Rhea	5.27	4.68
				.6	Titan	12.2	3.84
				.7	Hyperion	14.8	3.65
				.8	Iapetus	35.6	2.77
7	Uranus	87.2	1.94	7.01	ring ξ	0.42	5.34
				.02	ring ε	0.51	5.14
				.5	Miranda	1.29	4.21
				.1	Ariel	1.91	3.82
				.2	Umbriel	2.66	3.49
				.3	Titania	4.36	3.00
				.4	Oberon	5.83	2.71
4	Mars	0.64	−0.19	4.1	Phobos	0.09	1.97
				.2	Deimos	0.23	1.03

lateral grouping becomes increasingly pronounced; with a coefficient of 4 the insular grouping has spread into the pronounced band pattern seen in Figures 4–7, where this factor is used. Further increase of the coefficient, e.g., to 10, does not change the clustering pattern.

9. Discussion

By proper numerical scaling of gravitational potential vs mass of the central body, the Cluster Program can be brought to recognize a cohesive band structure of the secondary bodies in the solar system, persisting through a total number of 8 or 9 clusters. The shape of the band clusters, and particularly the hierarchial ranking in their cohesion, differs on minor points from the earlier more subjectively assigned structure, but retains its general features. Among these stands out the proposed association with the critical velocity limits for the four major element groups. Given the fact that the critical velocity phenomenon appears to operate in the space medium today (Haerendel, 1982; Galeev *et al.*, 1986), the close coincidence between the critical velocity limits and the major band features must be taken as strongly suggestive evidence for the relationship between these phenomena and thus for the emplacement mechanism and the hydromagnetic structuring of the solar nebula, proposed by Alfvén.

Notwithstanding this general agreement between observation and theory, there are some anomalies that raise further questions. One concerns the Martian satellites that appear to violate the critical velocity-band structure systematics. They could be written off as captured bodies from the asteroid belt were it not for the fact that their orbits are almost perfectly circular and ecliptic (Burns, 1977; Pollack, 1977). To give them an appropriately large mass, permitting the evolution of originally eccentric and inclined capture orbits, it has been suggested that they are the small, rocky residues of originally large icy bodies that subsequently have lost their volatile components (Singer, 1971). On the other hand, their surfaces are now known to show features that are characteristic of planetesimal accretion rather than resulting from a thawing process. To account for their original accretion in closely similar and near-stationary orbits, Gold (1975) has proposed a mechanism based on Lorentz force acting on dusty plasma in a planetary magnetic field.

Another question concerns the origin of the, albeit small, mass that constitutes the Saturnian rings and associated small satellites; this mass broadens the C-band to close vicinity with the A-region and diminishes the impressiveness of this band gap, at least in the satellite systems. In general the C-band is broad and complex and breaks in two even before the A and B bands separate from each other at the seven-cluster level. The fixation of the discussion on three bands is thus somewhat arbitrary – it could from a statistical point of view involve two, four, or five. It is possible that the extended structure of the C-band is due to the second-order modifications of the critical ionization distance caused by the partial corotation of the interacting plasma. These effects are outlined in Section 21.13 (Alfvén and Arrhenius, 1976). Broadening may also be caused by the differential critical velocities of abundant stable molecular ions such as CH_4^+ and

CO^+ forming in the C-cloud. Further experimental exploration of the critical velocity phenomenon may cast light on these phenomena.

Finally, the statistical technique used here for objective verification of the band structure needs further analysis to clarify the importance of the scaling parameter discussed above.

Acknowledgements

The authors wish to dedicate this work to Hannes Alfvén on his 80th birthday, with gratitude for many decades of cosmic inspiration and insight. Helpful criticism and comments from Asoka Mendis are warmly acknowledged.

References

Alfvén, H.: 1942, *Stockholm Observatorium Ann.* **14**(2), 3.

Alfvén, H.: 1943, *Stockholm Observatorium Ann.* **14**(5), 3.

Alfvén, H.: 1946, *Stockholm Observatorium Ann.* **14**(9), 3.

Alfvén, H.: 1954, *On the Origin of the Solar System*, Oxford Univ. Press, London.

Alfvén, H.: 1981, *Cosmic Plasma*, Astrophysics and Space Science Library, Vol. 82, D. Reidel Publ. Co., Dordrecht, Holland, 164 pp.

Alfvén, H. and Arrhenius, G.: 1972, *Moon* **5**, 210.

Alfvén, H. and Arrhenius, G.: 1975, *Structure and Evolutionary History of the Solar System*, D. Reidel Publ. Co., Dordrecht, Holland.

Alfvén, H. and Arrhenius, G.: 1976, *Evolution of the Solar System*, NASA SP-345, U.S. Govt. Printing Office, Washington, D.C., 599 pp.

Alfvén, H. and Carlqvist, P.: 1978, *Astrophys. Space Sci.* **55**, 487.

Alfvén, H., Axnäs, J., Brenning, N., and Lindqvist, P.-A.: 1985, *Further Explorations of Cosmogonic Shadow Effects in the Saturnian Rings*, TRITA-EPP-85–09, Roy. Inst. Tech., Stockholm.

Angerth, B., Block, L., Fahleson, U., and Soop, K.: 1962, *Nucl. Fusion Suppl.* **I**, p. 39.

Baxter, D. and Thompson, W. B.: 1971, 'Jet Stream Formation through Inelastic Collisions', in T. Gehrels (ed.), *Physical Studies of Minor Planets*, Proc. 12th Colloq. IAU, NASA SP-267, p. 319.

Baxter, D. and Thompson, W. B.: 1973, *Astrophys. J.* **183**, 323.

Bhattacharyya, J. C. and Kuppuswamy, K.: 1977, *Nature* **267**, 331.

Brenning, N.: 1982, *A Necessary Condition for the Critical Ionization Velocity Interaction*, TRITA-EPP-82–01, Roy. Inst. of Techn., Stockholm.

Burns, J. A.: 1977, 'Orbital Evolution', in J. A. Burns (ed.), *Planetary Satellites*, Univ. of Arizona Press, Tucson, p. 113.

Danielsson, L.: 1970, *Phys. Fluids* **13**, 2288.

Danielsson, L. and Brenning, N.: 1975, *Phys. Fluids* **18**, 661.

De, B. R.: 1972, Verbatim original of De (1978), submitted to *Icarus*, December 1972, entitled 'On the Possibility of Existence of a Ring of Uranus' (unpublished).

De, B. R.: 1978, *Moon and Planets* **18**, 339.

De, B. R., Alfvén, H., and Arrhenius, G.: 1977, 'The Critical Velocity Phenomenon and the Origin of the Regular Satellites', in J. A. Burns (ed.), *Planetary Satellites*, Univ. of Arizona Press, Tucson, p. 472.

Delsemme, A. H.: 1980, *Icarus* **42**, 448.

Elliot, J. L., Dunham, E., and Mink, D.: 1977, *Nature* **267**, 328.

Galeev, A. A., Gringauz, K. I., Klimov, S. I., Remisov, A. P., Sagdeev, R. Z., Savin, S. P., Sokolov, A. Yu., and Verigin, M. I.: 1986, *Geophys. Res. Letters* **13**, 845.

Gold, T.: 1975, *Icarus* **25**, 489.

Haerendel, G.: 1982, 'Critical Velocity Experiments in Space', in G. Haerendel and E. Möbius (eds.), *Workshop on Alfvén's Critical Velocity*, Max-Planck Institute for Extraterrestrial Physics, Garching, p. 216.

Johnson, S. C.: 1967, *Psychometrica* **32**, 241.

Kirsten, T.: 1978, 'Time and the Solar System', in S. F. Dermott (ed.), *The Origin of the Solar System*, J. Wiley and Sons, New York, p. 267.

Mendis, D. A.: 1981, 'The Role of Electrostatic Charging of Small and Intermediate-Sized Bodies in the Solar System', in F. D. Kahn (ed.), *Investigating the Universe*, D. Reidel Publ. Co., Dordrecht, Holland, p. 353.

Mendis, D. A., Houpis, H. L., and Hill, J. R.: 1982a, *J. Geophys. Res.* **87**, 3449.

Mendis, D. A., Hill, J. R., and Houpis, H. L.: 1982b, *J. Geophys. Res.* **88**, 4929.

Morfill, G. E., Fechtig, H., Grün, E., and Goertz, . K.: 1983, *Icarus* **55**, 439.

Northrop, T. G. and Connerney, J. E.: 1987, *Icarus* **70**, 124.

Öpik, E. J.: 1978, *Observatory* **98**, 32.

Piel, A. and Möbius, E.: 1978, *Measurement of Structure and Properties of a Critical Velocity Rotating Plasma in Xenon with Respect to Space Application*, SFB Plasmaphysik Bochum, Jülich, Rept. 78-M2-040.

Pollack, J. B.: 1977, 'Phobos and Deimos', in J. A. Burns (ed.), *Planetary Satellites*, Univ. of Arizona Press, Tucson, p. 319.

Raadu, M. A.: 1978, *Astrophys. Space Sci.* **55**, 125.

Reeves, H.: 1972, 'Seven Questions Related to the Origin of the Solar System and the Answers Given in Different Models', in H. Reeves (ed.), *Symposium on the Origin of the Solar System*, CNRS, Paris.

Reeves, H.: 1978, 'The Origin of the Solar System', in S. F. Dermott (ed.), *The Origin of the Solar System*, John Wiley and Sons, New York, p. 1.

Safronov, V. S.: 1954, *Astron Zh.* **31**, 499.

Safronov, V. S.: 1969, *Evolution of the Protoplanetary Cloud and Formation of Earth and the Planets*, Moscow (English translation, Israel Program for Scientific Translation, Jerusalem, 1972).

Singer, S. F.: 1971, 'The Martian Satellites', in T. Gehrels (ed.), *Physical Studies of Minor Planets*, NASA SP-267, U.S. Govt. Printing Office, Washington, D.C., pp. 399–405.

Sokal, A.: 1981, *Biometry*, W. H. Freeman Co., San Francisco.

Varma, R. K.: 1978, *Astrophys. Space Sci.* **55**, 113.

Wetherill, G. W.: 1980, *Ann. Rev. Astron. Astrophys.* **18**, 77.

THE COSMOGONIC SHADOW EFFECT*

MICHEL AZAR

Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA, U.S.A.

and

WILLIAM B. THOMPSON

Department of Physics, University of California, San Diego, La Jolla, CA, U.S.A.

(Received 4 September, 1987)

Abstract. We consider the Alfvén–Arrhenius fall-down mechanism and describe an approximate model for the infall, capture and distribution of dust particles on a given magnetic field line and their possible neutralization at the '$\frac{2}{3}$' points, the points at which the field aligned components of the gravitational and centrifugal forces are equal and opposite. We find that a small fraction ($<10\%$) of an incoming particle distribution will actually contribute to the above '$\frac{2}{3}$' fall-down process. We also show that if at the $\frac{2}{3}$ points, the ratio of dust to plasma density is

$$\frac{n_D(\frac{2}{3})}{n_p(\frac{2}{3})} > \frac{10^{-3}}{r_{g_\mu} T_{\mathrm{eV}}}$$

(r_{g_μ} = radius of a grain in microns, T = plasma temperature in eV), then the dust particles will lose their charge, decouple from the field line and follow Keplerian orbits in accordance with the Alfvén–Arrhenius mechanism.

We then determine the limits on the plasma parameters in order that rotation of a quasi-neutral plasma in thermal equilibrium be possible in the gravitational and dipole field of a rotating central body. The constraints imposed by the above conditions are rather weak, and the plasma parameters can have a wide range of values. For a plasma corotating with an angular velocity $\Omega \sim 10^{-4}\ \mathrm{s}^{-1}$, we show that the plasma temperature and density must satisfy

$$10^{-1} \ll T_{(\mathrm{eV})} \ll 10^2, \qquad 10 T_{\mathrm{eV}}^2 \ll n^p(\mathrm{cm}^3) \ll 10^6 .$$

1. Introduction

Various gaps and density minima have been observed in the Saturnian ring system. Attempts have been made to attribute these observations to gravitational resonances with the inner satellites, thus causing the removal of particles from the dark regions of the ring system (Alexander 1953, 1962).

According to an alternative theory (Alfvén and Arrhenius, 1976) the observed dark regions are the result of a 'cosmogonic shadow effect' produced by the 'two-thirds fall-down mechanism'. The basis of this mechanism is the following: plasma contained in the magnetic dipole field of a central (celestial) body is brought up to a state of partial corotation. In the corotating frame of reference, the plasma experiences an outward centrifugal force which drives a current of density **J**. In addition to the gravitational and

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

centrifugal forces. a plasma element is therefore subject to a magnetic force $\mathbf{J} \times \mathbf{B}$. These forces confine a plasma particle to a field line, but if the gyroradius is small, allow it to move rather easily along the field line. If the magnetic mirroring effect in the field gradient is neglected, such a particle will find equilibrium along the field line at the '$\frac{2}{3}$' points, i.e., the points at which $V_\phi^2 = \frac{2}{3}V_k^2$, $\mathbf{V}_\phi$ being the circular velocity and $\mathbf{V}_k$ the Kepler velocity at the same radial distance (Figure 1).

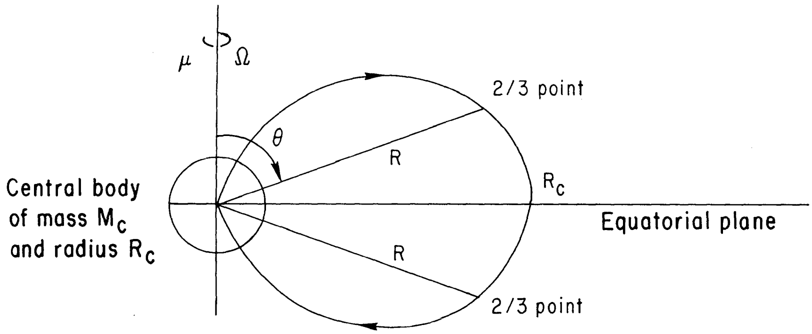


Fig. 1. A magnetic dipole field line with $\mathbf{\Omega}_c \parallel \boldsymbol{\mu}$ and $R = R_e \sin^2\theta$. As shown by Alfvén and Arrhenius, the field aligned components of the gravitational and centrifugal force balance at the '$\frac{2}{3}$' points, where the rotational speed is $(2/3)^{1/2}$ of the Kepler speed at the same radial distance. If neutralized, the particle leave the field line, follows an inclined elliptical orbit of eccentricity $e = \frac{1}{3}$ which will cross the equatorial plane at $R_\Omega = \frac{2}{3}R = \frac{2}{3}R_e \sin^2\theta$.

If the plasma components recombine at the '$\frac{2}{3}$' points, the effect of the magnetic force will disappear and the neutralized particles will move under the influence of the gravitational force alone. Because of the initial condition, i.e., $V_\phi^2 = \frac{2}{3}V_k^2$, particles at the same central distance R will follow inclined elliptical orbits with the same ecentricity $e = \frac{1}{3}$, with the same major axis $\frac{3}{4}R$ which will cross the equatorial plane at the circumference of a circle of radius $\frac{2}{3}R$. Mutual collisions between a large number of such particles will eventually result in a final circular orbit (of radius $\frac{2}{3}R$) of solid particles in the equatorial plane.

The bulk structure of the Saturnian ring can be explained as being caused by this fall-down process. If before the fall down, plasma in a certain region $R_2 - R_1$ was absorbed by a planet or satellite, then after the fall down there will be no matter, i.e., a gap or a 'shadow' in the region $\frac{2}{3}(R_2 - R_1)$.

This Alfvén–Arrhenius mechanism raises two questions. First, the equilibrium at the '$\frac{2}{3}$' points is unstable. This is because the effective potential (gravitational plus centrifugal) has a local maximum at the '$\frac{2}{3}$' points (Figure 2). For plasma in thermal equilibrium, the density along a given field line is given by the Boltzmann factor

$$\rho \sim e^{-m_i\psi/2kT},$$

where ψ is the effective potential, m_i the mass of an ion, T and ρ the plasma temperature and density. It is immediately clear that, on a given field line (for which $R_e > R_1$), the

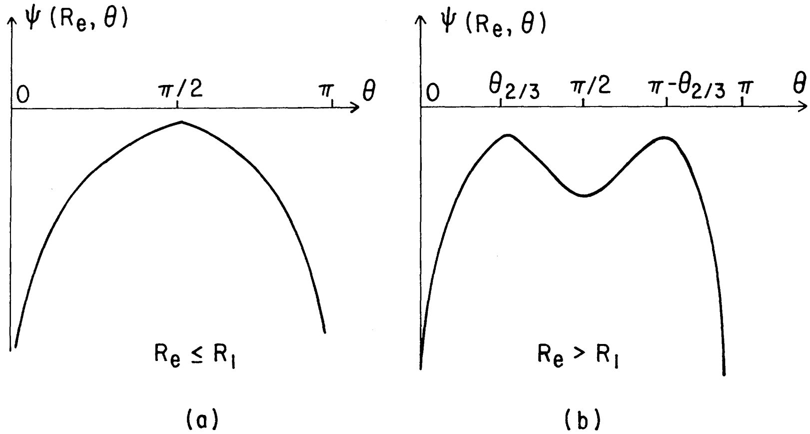


Fig. 2. For $R_e \leq R_1$, the effective potential (gravitational plus centrifugal)

$$\psi(R_e, \theta) = -\frac{GM_c}{R_e \sin^2\theta} - \frac{1}{2}\Omega^2 R_e^2 \sin^6\theta$$

has a maximum at $\theta = \frac{1}{2}\pi$ (the equator). For $R_e > R_1$, $\psi(R_e, \theta)$ has a minimum at $\theta = \frac{1}{2}\pi$ and two maxima at $\theta_{2/3}$ and $\theta_2 = \pi - \theta_{2/3}$ ($\frac{2}{3}$ points). $R_1 \equiv (\frac{2}{3}(GM_c/\Omega^2))^{1/3}$, $\sin\theta_{2/3} = (R_1/R_e)^{3/8}$, this is equivalent to $V_\phi^2 = \frac{2}{3}V_K^2$.

plasma density is maximum at the equatorial plane (bottom of the potential well) and minimum at the '$\frac{2}{3}$' points (hilltops of the potential). We, therefore, conclude that recombination is very unlikely to occur at the '$\frac{2}{3}$' points if plasma were present alone.

Second, the factor '$\frac{2}{3}$' is a result of the geometry of the magnetic field, i.e., a dipole aligned with the rotation vector of the central body. Plasma rotating in the magnetic field of the central body will cause a distortion of this field. In the second part of this paper we will consider this problem and determine the conditions under which the distortion of the dipole field of the central body, is negligible. This requirement, along with that for quasi-neutrality and thermal equilibrium, impose restrictions on the plasma parameters.

For clarity, we first outline the arguments, summarize the results, and then present the details of the calculation.

2. Summary

2.1. Role of Dust

To resolve the first question raised in the previous section (i.e., the effective potential has a maximum at the '$\frac{2}{3}$' points), we consider the role of dust in the '$\frac{2}{3}$' fall-down mechanism, and suggest a simple model for the distribution of the dust particles, their interaction with the ambient plasma, and their possible neutralization at the '$\frac{2}{3}$' points. As mentioned before, the plasma is assumed to be in thermal equilibrium in the gravitational and magnetic field of the rotating central body.

We envisage micron size dust grains falling in from infinity, and acquiring a charge as they pass through the ionized region. In our present oversimplified model, we imagine the charge as rising rather suddenly to such a large value that grains are effectively trapped on the field line.

It is unlikely that the grains will fall directly toward the central body, instead they will have some distribution in angular momentum, and of this, part will appear as a corotation with the plasma, and part as a gyromotion about the field lines.

Only those particles which are close to corotation will have small gyroradii, and these will stay close to a magnetic surface, and only for these will the gravitational and centrifugal force dominate. Counter-rotating particles will have large gyroradii and large value of $\bar{\mu}$ (magnetic moment due to gyration); so that their motion is dominated by the mirroring effect of the magnetic field, and the equivalent potential with a minimum on the geomagnetic equator.

Our analysis applies only to the almost corotating particles, assumed to be collisionless. In general, these will be trapped on the field line at a point labeled s, and with some parallel component of the velocity v. The dust particle distribution function $f(s, v, t)$ is then described by Vlasov equation

$$\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial s} - \frac{\partial \psi}{\partial s}\frac{\partial f}{\partial v} + \gamma f = S(s, v), \tag{1}$$

where $S(s, v)$ is a source function for the trapped dust and γ is a loss term.

For simplicity, we make the following assumptions:

(a) The loss rate is proportional to the particle density, i.e., γ is constant.

(b) The source function S is independent of the particle position s on the field line, but the dependence on the parallel component of the velocity v is given by

$$S(s, v) = \begin{cases} S_0 = \text{constant} & \text{for } |v| \leq V_0, \\ 0 & \text{otherwise}; \end{cases} \tag{2}$$

(c) The maximum kinetic energy of the trapped particles is greater than the effective potential well depth, i.e., $V_0 \geq \sqrt{2\psi_0}$ (see below).

(d) The infall and loss of particles are slow processes, i.e., $\gamma/\omega \ll 1$, $S_0/\omega\,|f| \ll 1$, but the ratio $S_0/\gamma \equiv \Gamma$ is finite. Here ω is a frequency characteristic of the particle motion along the field line.

In addition, we will approximate the effective potential

$$\psi = -\left(\frac{GM_c}{R_e \sin^2\theta} + \tfrac{1}{2}\Omega^2 R_e^2 \sin^6\theta\right) \tag{3}$$

by a sine function of the form (Figures 3 and 4)

$$\psi \approx \psi_0 \sin\frac{3\pi}{4}\frac{s}{l}, \tag{4}$$

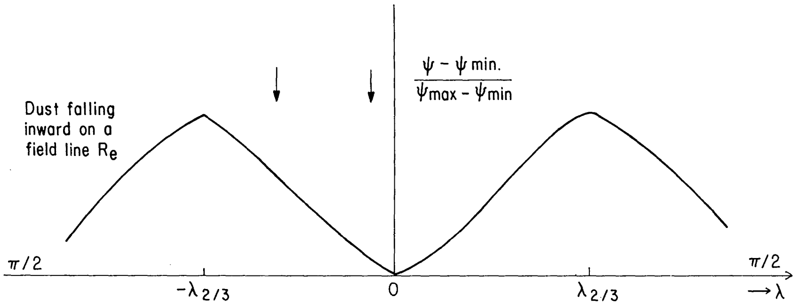


Fig. 3. Plot of

$$\frac{\psi - \psi_{\min}}{\psi_{\max} - \psi_{\min}} = \frac{\sin^2 \lambda + 3\left(\dfrac{L_1}{L}\right)^3 (1 - \sec^2 \lambda)}{1 + 3\left(\dfrac{L_1}{L}\right)^3 - 4\left(\dfrac{L_1}{L}\right)^{9/4}}$$

vs latitutde $\lambda = \frac{1}{2}\pi - \theta$ on a certain field line $L > L_i$.

$$L_1 = \left(\frac{2GM_c}{3\Omega^2}\right)^{1/3} \frac{1}{R_e} \quad \cos \lambda_{2/3} = \left(\frac{L_1}{L}\right)^{3/8} \quad L \equiv \frac{R_e}{R_c} \ .$$

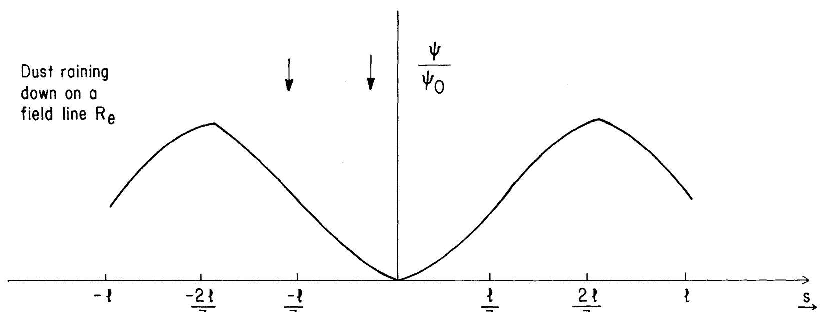


Fig. 4. Plot of $\psi/\psi_0 = \sin^2(3\pi/4)(s/l)$. s is the coordinate of a point on the field line relative to the equatorial plane. $l = l(R_e)$ is the coordinate of the point of the highest latitude.

where

$$\psi_0 \equiv \psi_{\max} - \psi_{\max} = \tfrac{1}{2}\Omega^2 R_e^2 \left[1 + 3\left(\frac{R_1}{R_e}\right)^3 - 4\left(\frac{R_1}{R_e}\right)^{9/4}\right], \tag{5}$$

and $l = l(R_e)$ is the coordinate of the highest latitude point on the field line.

The number density of the dust particles is obtained from the steady state solution of (1) as:

$$n_D(s) = \int f(\bar{s}, \bar{v})\, \mathrm{d}\bar{v} \, . \tag{6}$$

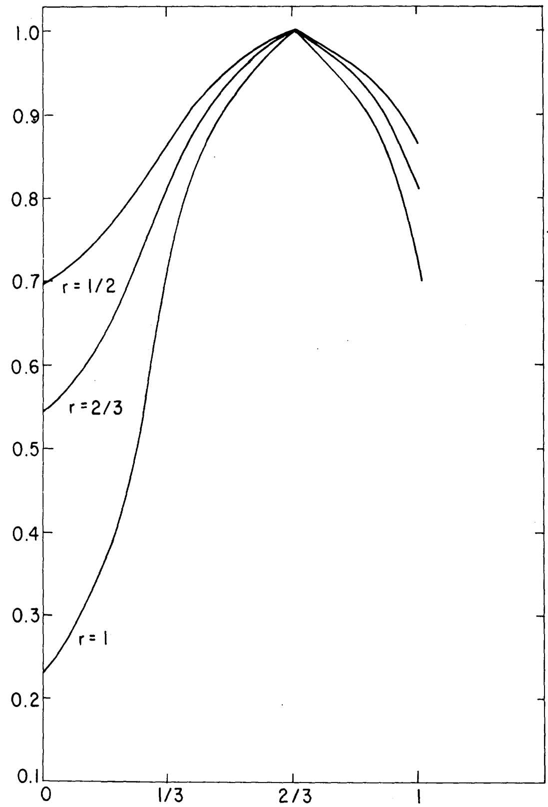


Fig. 5. Variation of the dust density $n_D(s/l, L)$ with latitude on a given line L, for different values of $r \equiv \psi_0/\frac{1}{2}V_0^2 \leq 1$. It is minimum at the equator $s = 0$ and maximum at the '$\frac{2}{3}$' points.

Figure 5 shows the variation of $n_D(s)$ with latitude along a certain field line R_e. If $v_0 \geq \sqrt{2\psi_0}$ then the charged particles will, on the average, have enough initial energy to leave the potential well at the equatorial plane where the speed is maximum, and overcome the hilltops of the potential ('$\frac{2}{3}$' points) where their speed is minimum. Hence, on a given field line, unlike the distribution of the ambient plasma which is in thermal equilibrium, and because the interparticle collisions are neglected, the density of the charged dust particles is minimum at the equatorial plane and maximum at the '$\frac{2}{3}$' points (hilltops of the potential). Because the equilibrium is unstable at the '$\frac{2}{3}$' points, the dust particles will either fall along the field line towards the central body, or else, if discharged, will leave the field line and follow Keplerian orbits in accordance with the Alfvén–Arrhenius mechanism.

2.2. Neutralization of the dust particles

The equilibrium charge on a dust particle is obtained from the current balance, i.e., $I_e + I_i = 0$ at the surface of the particle. The current (ion or electron) is the thermal flux

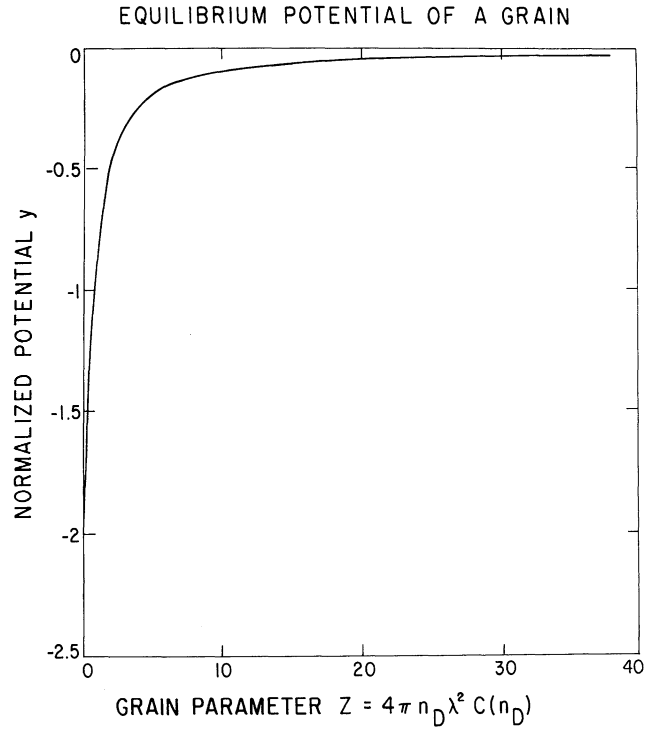


Fig. 6. Potential on a dust grain versus the grain parameter $Z \simeq 4\pi\lambda^2 n_D C$. C is the grain capacitance and n_D is the density of the dust particles.

modified by the negative potential of the particle. Under the conditions considered in this paper, contributions to the total current from such charging processes as photo-emission, field emission, and secondary emission are negligible. The equilibrium charge depends, among other things, on the properties of the ambient plasma and the dust concentration. This dependence is shown in Figure 6 (after Mendis *et al.*, 1985), where the normalized potential of a dust particle is plotted against the parameter $Z \cong 4\pi\lambda^2 n_D C(n_D)$, where λ is the Debye length, C the grain capacitance and n_D the density of the dust particles. As discussed in the previous section, the ratio n_D/n_p of dust to plasma density increases, hence, Z increases and as shown in Figure 6, the equilibrium potential on a dust particle decreases as we go from the equatorial plane to the '$\frac{2}{3}$' points along a certain field line. The equilibrium potential approaches zero as Z become greater than unity. Since in this case, Z can be approximately written as $Z \approx kT/(e^2/r_g)\,(n_D/n_p)$, we conclude that if at the '$\frac{2}{3}$' points $n_D/n_p > (e^2/r_g)/kT$, then a dust particle will decouple from the field line and move under the influence of the gravitational field alone.

2.3. Fraction of particles contribution to the '$\frac{2}{3}$' fall down process

As discussed above, our treatment is limited only to corotating particles for which the magnetic mirroring effect can be neglected. Therefore, only a fraction of an incoming

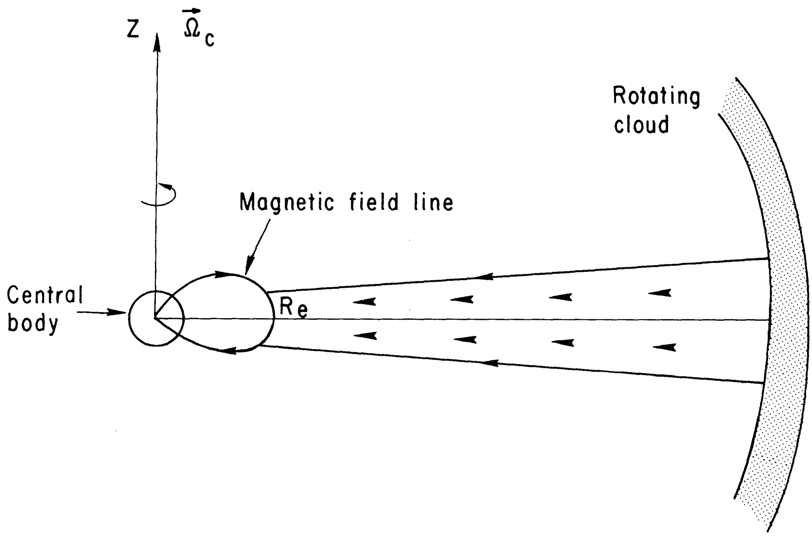


Fig. 7. Dust particles condensing out of a distant cloud rotating with constant angular momentum $\mathbf{L} = L_0\hat{Z}$. The Z-axis is taken along $\mathbf{\Omega}_c$ the angular velocity of the central body.

particle distribution will contribute to the '$\frac{2}{3}$' fall-down process. To estimate this fraction, we determine the rate $G(v_\parallel, v_\perp, \theta, R_e)\, dv_\parallel\, d(\frac{1}{2}v_\perp^2)\sin\theta\, d\theta$ of particles condensing out of a distant cloud rotating with constant angular momentum $\mathbf{L} = L_0\hat{z}$ (Figure 7). Such particles will be captured on the field line R_e, at colatitude between θ and $\theta + d\theta$, with parallel component of the velocity between $v_\parallel$ and $v_\parallel + dv_\parallel$, and with magnetic moment between $\bar{\mu}$ and $\bar{\mu} + d\bar{\mu}$ (where $\bar{\mu}B = \frac{1}{2}v_\perp^2$). These particles will be able to reach the '$\frac{2}{3}$' points if they have enough energy in the parallel motion to overcome the potential barrier, i.e. (if $\frac{1}{2}v_\parallel^2 > \psi_0$), and if no reflection, due to mirroring effect, occurs below the '$\frac{2}{3}$' points, i.e., if

$$\tfrac{1}{2}v_\perp^2 \leq \frac{B}{B_1 - B}\,[\tfrac{1}{2}v_\parallel^2 - \psi_0]\,,$$

where B is the magnetic field strength at the initial position of the particle and B_1 that at the '$\frac{2}{3}$' points. The fraction F of particles taking part in the '$\frac{2}{3}$' fall-down process is found by integrating G in velocity space over $v_\parallel$ and $v_\perp$ subject to the two conditions above. We will show that, if the incoming particles have the same Z-component of the angular momentum $L_0 \sim \Omega R_e^2$ then F will be expressed as

$$F \lesssim 1 - \sqrt{1 - \frac{B}{B_1}}\,.$$

As seen in Figure 10, in the region of interest where the effective potential has a minimum at the equator (i.e., for $R_e > R_1$), the fraction F is rather small ($<10\%$). However, if a uniform distribution of grains in the equatorial plane is unstable to filamentation, as is suggested by observation of the fine structure of ring systems and as was predicted by Baxter and Thompson (1971, 1973), then a small initial disturbance may be enough to produce a dramatic final effect.

2.4. Plasma Density and Temperature Limits

The upper limit on the plasma density is set by the requirement that the field $\delta\mathbf{B}$ induced by currents in the plasma be small in comparison with the dipole filed $\mathbf{B}_0$, i.e. $|\delta\mathbf{B}/\mathbf{B}_0| \ll 1$. As we will see, this condition can be cast in a simple and familiar form

$$\Omega R_0 \ll V_{\mathrm{A}}(R_2) \equiv \frac{B_2(R_2)}{\sqrt{4\pi\rho_2}} .$$

This means that the rotational speed at the inner edge R_0 must be much less than the Alfvén speed at the outer edge R_2 of the magnetosphere.

The assumption that the plasma is in thermal equilibrium means that the mean free path for Coulomb collisions must be less than some characteristic distance scale. It can easily be shown that this requirement imposes a lower limit on the plasma density

$$n_p(\mathrm{cm}^{-3}) \gg 3.8 \times 10^{11} \frac{T_{\mathrm{eV}}^2}{R_e(\mathrm{cm})} ,$$

where the typical distance scale is taken to be on the order of the equatorial distance R_e to a given field line.

Next, an upper limit on the plasma temperature is obtained by requiring that the two conditions above (small distortion of the dipole field and thermal equilibrium of the plasma) be simultaneously satisfied; while a lower limit is set by the requirement of quasi-neutrality. For a plasma in thermal equilibrium, the plasma density along a given field line R_e will be shown to be given by

$$\rho(R_e, \theta) \alpha \frac{\Omega R_e}{a} \exp - \left| \frac{\psi(R_e, \theta) - \psi(R_e, \frac{1}{2}\pi)}{a^2} \right| ,$$

where

$$a^2 = 2\frac{kT}{m_i} .$$

The above expression shows that, on a given field line, the plasma distribution becomes less and less uniform as the temperature decreases. Since $\rho(R_e, \theta_{2/3}) \leq \rho(R_e, \theta)$ quasi-neutrality holds at every point on the field line if it holds at the '$\frac{2}{3}$' points, i.e., $\rho_q(R_e, \theta_{2/3}) \ll (q_i/m_i)\rho(R_e, \theta_{2/3})$, $\rho_q(R_e, \theta_{2/3})$ being the net charge density at the '$\frac{2}{3}$' points and q_i and m_i the charge and mass of a positive ion. This condition sets a lower limit on the plasma temperature. For a plasma corotating with angular velocity $\Omega \sim 10^{-4}\ \mathrm{s}^{-1}$, detailed calculations will show that

$$10^{-1} \ll T_{(\mathrm{eV})} \ll 10^2 , \qquad 10T_{(\mathrm{eV})}^2 \ll n_p(\mathrm{cm}^{-3}) \ll 106 .$$

These are weak constraints and can easily be satisfied.

3. Dust Particles Distribution on a Given Field Line

To solve Vlasov equation:

$$\frac{\partial f}{\partial t} + v\,\frac{\partial f}{\partial s} - \frac{\partial \psi}{\partial s}\,\frac{\partial f}{\partial v} + \gamma f = S\,, \tag{1}$$

we introduce new variables (action, angle variables) ε and ϕ:

$$\varepsilon = \tfrac{1}{2}v^2 + \psi(s)\,, \tag{7}$$

$$\phi = \frac{\pi}{2} - \omega \int\limits_0^s \frac{\mathrm{d}\bar{s}}{\sqrt{2[\varepsilon - \psi(\bar{s})]}}\,. \tag{8}$$

If so, Vlasov equation (1) then becomes

$$\frac{\partial f}{\partial t} \pm \omega\,\frac{\partial f}{\partial \psi} + \gamma f = S(\varepsilon, \phi)\,; \tag{9}$$

ω being the angular frequency associated with the periodic motion in the effective potential ψ. The corresponding period T is given by

$$T(\varepsilon) = 4 \int\limits_0^{s_1} \frac{\mathrm{d}s}{\sqrt{2[\varepsilon - \psi(s)]}}\,; \tag{10}$$

s_1 being the positive root of the equation

$$\varepsilon = \psi(s_1)\,. \tag{11}$$

Because of (2), $S(\varepsilon, \phi)$ is written as:

$$S(\varepsilon, \phi) = \begin{cases} S_0 & \text{for} \quad \phi \varepsilon E(\varepsilon)\,, \\ 0 & \text{otherwise}\,; \end{cases} \tag{12}$$

where $E(\varepsilon)$ is a subinterval of $[0, 2\pi]$ such that

$$\varepsilon - \psi(s) < \tfrac{1}{2}V_0^2\,.$$

As shown in Appendix C, the solution of Equation (9) which satisfies the initial condition $f(\varepsilon, \phi, 0) = 0$ is

$$f(\varepsilon, \phi, t) = f(\varepsilon, t) = \Gamma(1 - e^{-\gamma t})\,\frac{\phi_0(\varepsilon)}{2\pi}\,. \tag{13}$$

In a steady state, this becomes

$$f(\varepsilon) = \Gamma\,\frac{\phi_0(\varepsilon)}{2\pi}\,, \tag{14}$$

where

$$\Phi_{0(\varepsilon)} = \int\limits_{\Phi' \varepsilon \mathrm{E}(\varepsilon)} \mathrm{d}\phi' = \begin{cases} 2\pi & \text{for } \varepsilon \leq \frac{1}{2}V_0^2\,, \\ 4\omega \displaystyle\int\limits_{s_0}^{s_1} \frac{\mathrm{d}\bar{s}}{\sqrt{2[\varepsilon - \psi(\bar{s})}} & \text{for } \varepsilon > \frac{1}{2}V_0^2\,; \end{cases}$$

s_1 being defined in (11) and s_0 the positive root of

$$\varepsilon - \tfrac{1}{2}V_0^2 = \psi(s_0)\,.$$

The number density of the dust particles is calculated from (14) as

$$n_D(s) = \int f(\bar{s}, \bar{v})\delta(\bar{s} - s)\,\mathrm{d}\bar{s}\,\mathrm{d}\bar{v} = \int f(\varepsilon)\delta[\bar{s}(\varepsilon, \phi) - s]J\,\mathrm{d}\varepsilon\,\mathrm{d}\phi\,, \tag{15}$$

where J (the absolute value of the Jacobian) is given by

$$J = \left|\frac{\partial(\bar{s}, \bar{v})}{\partial(\varepsilon, \phi)}\right| = \frac{1}{\omega(\varepsilon)} = \frac{T(\varepsilon)}{2\pi}\,; \tag{16}$$

$T(\varepsilon)$ being defined in (10). The number density $f(s)$ then becomes

$$n_D(s) = N \int \mathrm{d}\varepsilon\, \frac{f(\varepsilon)}{\sqrt{2[\varepsilon - \phi(s)]}}\,, \tag{17}$$

where $f(\varepsilon)$ is given in (14). In Equation (17) N is the number of zeros of $\bar{s}(\varepsilon, \theta) - s$.

The calculations of $T(\varepsilon)$, $f(\varepsilon)$, and $n_D(s)$, in the case where $\psi(s) = \psi_0 \sin^2(3\pi/4)(s/l)$, are straight forward, we omit the details and write the end results

$$T(\varepsilon) = \frac{16l}{3\pi\sqrt{2\psi_0}}\, K\left(\sqrt{\frac{\varepsilon}{\psi_0}}\right), \tag{18}$$

$$f(\varepsilon) = \begin{cases} \Gamma & \text{for } \varepsilon \leq \frac{1}{2}V_0^2\,, \\ \Gamma[1 - G(\varepsilon)] & \quad\;\; \varepsilon > \frac{1}{2}V_0^2\,; \end{cases} \tag{19}$$

$$\frac{n_D(s)}{n_0} = 2\sqrt{1 - r\cos^2 \frac{3\pi}{4}\frac{s}{l}} - \int\limits_1^{1+r} \frac{G(\frac{1}{2}V_0^2\chi)}{\sqrt{\chi - r\sin^2 \dfrac{3\pi}{4}\dfrac{s}{l}}}\,\mathrm{d}\chi\,, \tag{20}$$

where

$$n_0 \equiv N\Gamma V_0\,, \tag{21}$$

$$r \equiv \frac{\psi_0}{\frac{1}{2}V_0^2} \leq 1\,, \tag{22}$$

$$\chi \equiv \frac{\varepsilon}{\frac{1}{2}V_0^2}, \tag{23}$$

$$G(\varepsilon) \equiv 1 + \sqrt{\frac{\psi_0}{\varepsilon}}\,\frac{F\left(\alpha; \sqrt{\frac{\psi_0}{\varepsilon}}\right)}{K\left\{2\left[\left(\frac{\varepsilon}{\psi_0}\right)^{1/4} \Big/ 1 + \sqrt{\frac{\varepsilon}{\psi_0}}\right]\right\}}, \tag{24}$$

$$0 < \alpha < \frac{\pi}{2}, \qquad \sin\alpha = \sqrt{\frac{\varepsilon - \frac{1}{2}V_0^2}{\psi_0}}; \tag{25}$$

F and K are elliptic integrals of the first kind (K being the complete integral).

The variation of the dust particles density $n_D(s, R_e)$ with latitude on a given field line R_e, is shown in Figure 5 for $r = \frac{2}{3}$, and $\frac{1}{2}$. It is seen that for $r \equiv \psi_0/\frac{1}{2}V_0^2 \leq 1$, the density of the dust particles on a given field line R_e, increases from a minimum in the equatorial plane to a maximum at the '$\frac{2}{3}$' points, the points where the effective potential has an unstable equilibrium; then it decreases again as we go to higher latitudes. Comparison with Equation (49) and Figure 2(b) shows that the opposite is true for the plasma density.

4. Neutralization of the Dust Particles

Now the question naturally arises as to how the charge on a dust particle changes with latitude, i.e., as we go from the equator up to the '$\frac{2}{3}$' points. The electrostatic charging of dust particles in a plasma has been treated by many authors (Whipple, 1981; Axford and Mendis, 1976; Goretz and Ip, 1984; and Mendis *et al.*, 1985). It can be shown, under the conditions considered in this paper, that charging processes such as photo-emission, field emission thermo-ionic emission, and secondary emission can be neglected. Consequently, contributions to the total charging current come mainly from thermal fluxes of electrons and ions, i.e., $I = I_i + I_e$. A dust grain acquires its equilibrium charge $-q_e (q > 0)$ when

$$I = I_i + I_e = 0\,. \tag{26}$$

The electron and ion currents are given by (Mendis *et al.*, 1984)

$$\begin{aligned} I_e &= -4\pi r_g^2 n_e e \sqrt{\frac{kT}{2\pi m_e}}\, e^y, \\ I_i &= 4\pi r_g^2 n_p e \sqrt{\frac{kT}{2\pi m_i}}\,[1 - y], \end{aligned} \tag{27}$$

where $y = -q(e^2/kTC)$, C being the grain capacitance. In Equation (27) r_g is the grain radius, n_e the average electron density, and n_p is the average ion (plasma) density.

Mendis *et al.* (1985) wrote Equation (26) in the form

$$(1 - y)(1 - yZ) = \sqrt{\frac{m_p}{m_e}}\,(1 + yZ)e^{y}\,, \tag{26$'$}$$

where Z is a parameter defined by

$$Z(n_D) \approx 4\pi\lambda^2 n_D C(n_D)\,,$$

in which n_D is the dust particles density, C is the particle capacitance, and λ is the Debye length. The solution of Equation (26′) as a function of Z is reproduced in Figure 6. As we go from the equatorial plane to the '$\frac{2}{3}$' points along a certain field line, the ratio $n_D(s)/n^p(s)$ increases, consequently Z increases, and the normalized potential y on a dust particle drops significantly. As shown in Figure 6, the potential y approaches zero when $Z > 1$. This should occur when the average interparticle distance becomes comparable with the Debye length. In this case, it can be shown that (cf. Mendis *et al.*, 1985)

$$Z = \frac{kT}{e^2/r_g}\,\frac{n_D}{n_p}\,.$$

Hence, if at the '$\frac{2}{3}$' points

$$\frac{n_D}{n_p} \geq \frac{e^2/n_g}{kT} \approx 1.5 \times 10^{-3}\,\frac{1}{r_{g_\mu} T_{(\mathrm{eV})}}\,,$$

then a dust particle will almost completely be neutralized. As a result, it will subsequently decouple from the field line and follow a Keplerian orbit.

5. Fraction of Particles Contributing to the '$\frac{2}{3}$' Fall-Down Process

In the last two sections, we presented a model that described the role of dust in the '$\frac{2}{3}$' fall-down mechanism. In this section we turn to the crucial question of determining the fraction of incoming dust particles that can actually take part in this mechanism. Toward this end, we assume that particles condense out of a distant cloud rotating with constant angular momentum $\mathbf{L} = L_0\hat{Z}$ (Figure 7), and consider an initial distribution that describes dust particle all having the same z-component of the angular momentum L_0 and the same energy ε_0, i.e., $f_0 \alpha \delta(\varepsilon - \varepsilon_0)\delta(L_z - L_0)$. Using this initial distribution, we can determine the number of particles (per unit time)

$$G(v_\parallel, v_\perp, \theta, R_e)\,\mathrm{d}v_\parallel\,\mathrm{d}(\tfrac{1}{2}v_\perp^2)\sin\theta\,\mathrm{d}\theta\,,$$

which are captured on the field line R_e, at colatitude between θ and $\theta + \mathrm{d}\theta$, with parallel component of the velocity between $v_\parallel$ and $v_\parallel + \mathrm{d}v_\parallel$ and with magnetic moment between $\bar{\mu}$ and $\bar{\mu} + \mathrm{d}\bar{\mu}(\bar{\mu}B = \frac{1}{2}v_\perp^2)$. From the continuity equation

$$\frac{\partial n}{\partial t} + \nabla\cdot\int \mathbf{v}_0 f_0(\mathbf{v}_0)\,\mathrm{d}^3 v_0 = 0\,,$$

we obtain

$$N + \oint \mathrm{d}a_0 \int v_{0r} f_0(\mathbf{v}_0)\, \mathrm{d}^3 v_0 = 0$$

or

$$\mathrm{d}N = \pi f_0 v_0^3\, \mathrm{d}v_0 b\, \mathrm{d}b\, \mathrm{d}\psi_0 \sin\theta_0\, \mathrm{d}\theta_0\,, \tag{28}$$

where $N = \int n\, \mathrm{d}^3 r_0$, $\varepsilon = \frac{1}{2}v_0^2$, $L_z = bv_0 \sin\theta_0 \cos\psi_0$, b is the impact parameter and the angles ψ_0 and θ_0 are defined in Figure 8.

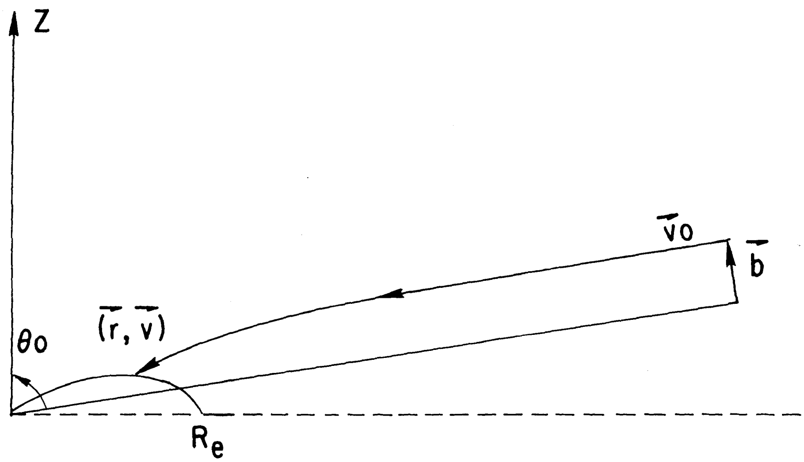


Fig. 8. Trajectory of a particle falling towards the central body with impact parameter $|\mathbf{b}|$ and initial velocity $\mathbf{v}_0$. The particle is captured on a certain field line R_e at position $\mathbf{r}$ and with velocity $\mathbf{v}$ where: $\mathbf{v} = (v_r, v_\theta, v_\psi)$; $\mathbf{v}_0 = (v_{0r}, 0, 0)$; $\mathbf{b} = (0, b \sin\psi_0, b\cos\psi_0)$.

If we assume that the particles are falling at low latitude ($\theta_0 \approx \frac{1}{2}\pi$) so that, in terms of the velocity $\mathbf{v} = (v_r, v_\theta, v_\psi)$ with which a particle is captured on a field line R_e, Equation (28) can be written as

$$\mathrm{d}N \simeq \pi R_e^2 f_0(\varepsilon, L_z) v_r\, \mathrm{d}v_r\, \mathrm{d}v_\theta\, \mathrm{d}v_\phi \sin\theta\, \mathrm{d}\theta\,. \tag{29}$$

In Equation (29) we have assumed that $\theta \approx \frac{1}{2}\pi$ and $r = R_e \sin^2\theta \approx R_e$. The conservation of energy and angular momentum allow us to write

$$\varepsilon = \tfrac{1}{2}V_0^2 = \tfrac{1}{2}(v_n^2 + v_\theta^2 + v_\phi^2) - \frac{k}{r}\,, \tag{30}$$

$$L_z = r\sin\theta v_\phi \equiv \rho v_\phi\,. \tag{31}$$

For a particle captured near the equatorial plane ($\theta \approx \frac{1}{2}\pi$), the kinetic energy in the motion parallel to the field line is $\frac{1}{2}v_\theta^2 \equiv \frac{1}{2}v_\parallel^2$, whereas that in the gyromotion is

$$\tfrac{1}{2}v_\perp^2 \equiv \bar{\mu}B = v_r^2 + (v_\phi - \rho\Omega)^2\,. \tag{32}$$

Using Equations (31) and (32) in (24) and (30), we obtain

$$\mathrm{d}N = \pi \frac{R_e^2}{\rho} f \,\mathrm{d}v_{\parallel}\, \mathrm{d}(\tfrac{1}{2}v_{\perp}^2)\, \mathrm{d}L_z \sin\theta \,\mathrm{d}\theta\,, \tag{33}$$

$$\varepsilon = \tfrac{1}{2}\left(v_{\parallel}^2 + v_{\perp}^2 - \frac{k}{r} - \Omega^2\rho^2\right) + \Omega L_z\,. \tag{34}$$

Integration of Equation (33) over L_z with $\rho^2\Omega - \rho v_\perp \le L_z \le \rho v_\perp + \rho^2\Omega$ gives

$$G(v_{\parallel}, v_{\perp}, R_e, \theta)\alpha\delta\left[v_{\parallel}^2 + v_{\perp}^2 - \rho^2\Omega^2 - \frac{2k}{r} + 2\Omega L_0 - 2\varepsilon_0\right] H\left[v_\perp - \left|\frac{L_0}{\rho} - \rho\Omega\right|\right], \tag{35}$$

where

$$H(x) = \begin{cases} 1 & \text{if} \quad x > 0\,,_1 \\ 0 & \text{if} \quad x < 0\,, \end{cases}$$

is the unit step function.

To calculate the amount of dust taking part in the '$\frac{2}{3}$' process, we note that a particle captured on a given field line R_e, at a colatitude $\theta(\sim\frac{1}{2}\pi)$ and with a parallel component $v_{\parallel}$ of the velocity, will be able to reach the '$\frac{2}{3}$' points if (i) it has enough kinetic energy in the parallel motion to overcome the potential barrier – i.e. if,

$$\tfrac{1}{2}v_{\parallel}^2 \ge \psi(R_e, \theta_{2/3}) - \psi(R_e, \theta) \equiv \psi_0\,;$$

(ii) no reflection, due to mirroring effect, occurs for $\theta < \theta_{2/3}$ – i.e., if

$$\tfrac{1}{2}v_{\perp}^2 \le \frac{B}{B_1 - B}\,[\tfrac{1}{2}v_{\parallel}^2 - \psi_0] \equiv \tfrac{1}{2}\bar{v}_2^2\,;$$

B_1 being the magnetic field strength at the '$\frac{2}{3}$' points. The fraction of particles, therefore, is

$$F \approx \frac{\displaystyle\int_{\sqrt{2\psi_0}}^{\infty} \mathrm{d}v_{\parallel} \int_0^{\bar{v}_\perp} \mathrm{d}v_\perp\, v_\perp\, G\left(v_{\parallel}, v_\perp, R_e, \frac{\pi}{2}\right)}{\displaystyle\int_0^{\infty} \mathrm{d}v_{\parallel} \int_0^{\infty} \mathrm{d}v_\perp\, v_\perp\, G\left(v_{\parallel}, v_\perp, R_e, \frac{\pi}{2}\right)}\,. \tag{36}$$

Substituting from Equation (35), we can write (36) as

$$F \approx \frac{\displaystyle\iint_{A_1} \delta[v_{\parallel}^2 + v_\perp^2 - W] v_\perp \,\mathrm{d}v_\perp\, \mathrm{d}v_{\parallel}}{\displaystyle\iint_{A_2} \delta[v_{\parallel}^2 + v_\perp^2 - W] v_\perp \,\mathrm{d}v_\perp\, \mathrm{d}v_{\parallel}}\,, \tag{37}$$

where $W \equiv 2\varepsilon_0 - 2\Omega L_0 + \Omega^2 R_e^2 + 2k/R_e$.

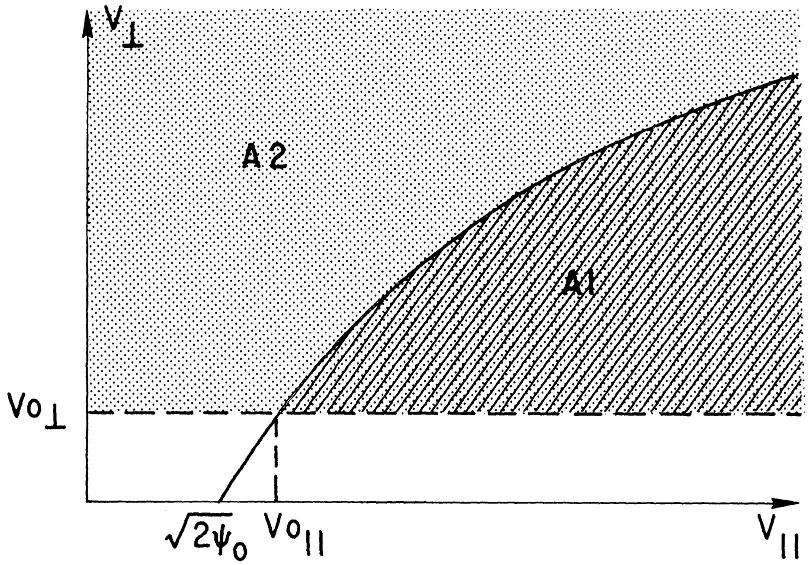


Fig. 9. Area A_2 is such that $v_\perp > |(L_0/R_e) - \Omega R_e|$. Area A_1 is between the curves $v_\perp = |(L_0/R_c) - \Omega R_e|$ and $v_\parallel^2 + [(B_1 - B)/B] V_\perp^2 = 2\psi_0$.

In Equation (37) the areas of integration A_1 and A_2 are shown in Figure 9. Since $v_\parallel^2 + v_\perp^2 \geq 0$, the conservation of energy implies that

$$\varepsilon_0 \geq \Omega L_0 - \tfrac{1}{2}\Omega^2 R_e^2 + \frac{2k}{R_e} . \tag{38}$$

Evaluation of Equation (37) yields

$$F \approx 1 - \left[1 - \frac{B}{B_1} \frac{w - (v_{0\parallel}^2 + v_{0\perp}^2)}{w - v_{0\perp}^2} \right]^{1/2} , \tag{39}$$

provided ε_0 satisfies constraint (38). In Equation (39), we have put

$$v_{0\perp}^2 \equiv \left(\frac{L_0}{R_e} - \Omega R_e \right)^2 ,$$

$$v_{0\parallel}^2 = 2\psi_0 + \frac{B_1 - B}{B} v_{0\perp}^2 ;$$

and, as before,

$$W = 2\left(\varepsilon_0 - \Omega L_0 + \tfrac{1}{2}\Omega^2 R_e^2 + \frac{k}{R_e} \right),$$

$$\psi_0 = \tfrac{1}{2}\Omega^2 R_e^2 \left[1 + 3\left(\frac{R_1}{R_e}\right)^3 - 4\left(\frac{R_1}{R_e}\right)^{9/4} \right]. \tag{5}$$

The magnetic field strength B_1, at the '$\frac{2}{3}$' points is given by

$$\frac{B_1}{B} = \frac{[4 - 3(R_1/R_e)^{3/4}]^{1/2}}{(R_1/R_e)^{9/4}}\,; \tag{40}$$

B being the magnetic field strength at the equator ($\theta = \frac{1}{2}\pi$) and $R_1 \equiv (2k/3\Omega^2)^{1/3} < R_e$.

The fraction of particles that can take part in the '$\frac{2}{3}$' fall-down process depends on the choice of ε_0 and L_0. Choosing $L_0 \gtrsim \Omega R_e^2$ is not physically unreasonable. In this case, (38) and (39) permit us to write

$$F \lesssim 1 - \sqrt{1 - \frac{B}{B_1}}\,, \quad \text{if } \varepsilon_0 > \tfrac{1}{2}\Omega^2 R_e^2 - \frac{k}{R_e} = \tfrac{1}{2}\Omega^2 R_e^2\left(1 - 3\,\frac{R_1^3}{R_e^3}\right),$$

$$F = 0\,, \qquad \text{otherwise}\,.$$

Figure 10 shows the variation of F with distance $R_e > R_1$ from the central body. It is clear that F increases as R_e decreases, this is because the '$\frac{2}{3}$' points get closer to the equatorial plane, and hence, more particles can overcome the potential barrier.

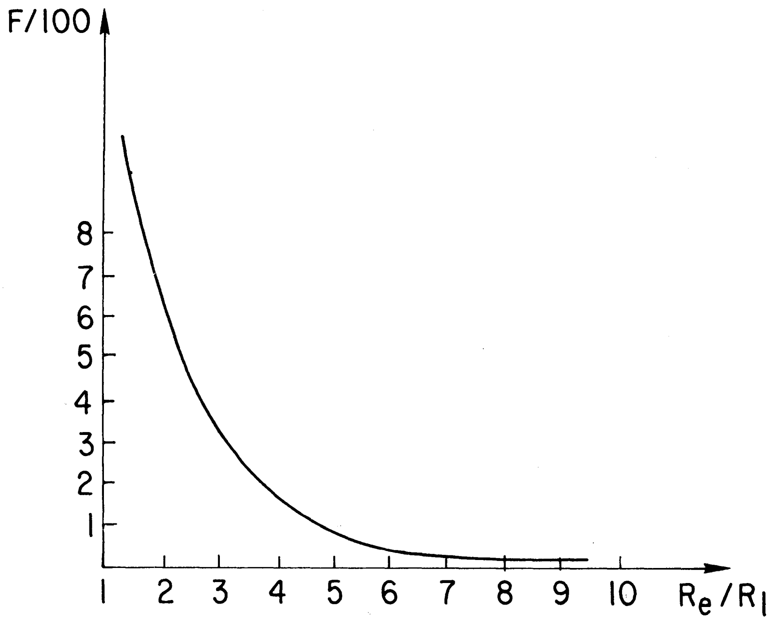


Fig. 10. Variation of F (fraction of dust particles contributing to the '$\frac{2}{3}$' fall-down mechanism) with equatorial distance.

Summarizing our results, we have found that it is necessary, for the '$\frac{2}{3}$' fall-down mechanism to work, to have a dust-plasma mixture. Because the plasma is in thermal equilibrium and the dust particles are assumed to be collisionless, we have shown that neutralization of dust is most likely to take place at the '$\frac{2}{3}$' points.

In addition, the requirements of thermal equilibrium, quasi-neutrality and small distortion of the dipole field must be satisfied. This will impose constraints on the plasma density and temperature. In the next section, we address these questions and determine what these constraints are.

6. Plasma Parameters Limits in the Gravitational and Magnetic Dipole Field of a Rotating Central Body

We apply the M.H.S. equations to an isothermal (rather cool) plasma cloud in equilibrium in the environment of a central body of mass M_c, rotating with angular velocity $\mathbf{\Omega}_c$ and having a magnetic dipole moment $\boldsymbol{\mu}$ parallel to $\mathbf{\Omega}_c$:

$$-\nabla_p - \rho\nabla\psi + \frac{\mathbf{J}\times\mathbf{B}}{c} = 0\,, \tag{41}$$

$$\nabla\times\mathbf{B} = \frac{4\pi}{c}\,\mathbf{J}\,, \tag{42}$$

$$\psi \equiv -\left(\frac{GM_c}{R} + \tfrac{1}{2}\Omega^2 \mathrm{R}^2 \sin^2\theta\right); \tag{43}$$

Ω being the angular speed of the plasma and ψ the effective potential (gravitational plus centrifugal) per unit mass at a point located at (R, θ) (Figure 1).

The above equations are valid if the following conditions are met:

(1) The plasma rotates rigidly, i.e., the local angular speed Ω of the plasma is constant.

(2) The system is quasi-neutral, i.e., $\rho_q \approx 0$ (ρ_q is the net charge density).

(3) The plasma is isotropic, i.e., there are many collisions during a characteristic time.

(4) The displacement current is negligible.

The pressure p and the mass density ρ are related by:

$$P = \rho a^2\,, \tag{44}$$

where

$$a^2 \equiv \frac{(1+Z_i)kT}{m_i\left(1+\dfrac{2m_e}{m_i}\right)} \approx \frac{(1+Z_i)kT}{m_i}\;; \tag{45}$$

where we have used quasi-neutrality, i.e., $\rho_q \approx 0$, and the fact that $Z_i m_e/m_i \ll 1$. m_e, m_i being the mass of electrons and ions, respectively; and Z_i is the charge of an ion in units of e. The magnetic field B can be written as

$$\mathbf{B} = \mathbf{B}_0 + \delta\mathbf{B}\,, \tag{46}$$

where $\mathbf{B}_0$ is the dipole field of the central body

$$B_{0R} = \frac{2\mu \cos\theta}{R^3}, \qquad B_{0\theta} = \frac{\mu \sin}{\mathrm{R}^3}; \tag{47}$$

and $\delta\mathbf{B}$ is the field produced by currents in the plasma. Thus

$$\nabla \times \mathbf{B}_0 = 0 \quad \text{and} \quad \nabla \times \delta\mathbf{B} = \frac{4\pi}{c}\,\mathbf{J}\,. \tag{48}$$

Using (44) and (48), Equation (46) can be solved to give

$$\rho(R_e, \theta) = g(R_e) \exp\left[-\frac{\psi(R_e, \theta)}{a^2}\right], \tag{49}$$

$$\mathbf{J}_\perp = -\frac{ca^2}{\mu} R_e^3 \sin^3\theta\, g'(R_e) \exp\left[-\frac{\psi(R_e, \theta)}{a^2}\right] \hat{\phi}\,, \tag{50}$$

for the plasma density $\rho(R_e, \theta)$ and current density $\mathbf{J}_\perp(R_e, \theta)$ at a point located on the field line R_e at latitude $\frac{1}{2}\pi - \theta g'(R_e) \equiv \mathrm{d}/\mathrm{d}R_e)g(R_e)$ with $g(R_e)$ being an arbitrary function of R_e to be determined.

6.1. Upper limit on the plasma density

The '$\frac{2}{3}$' fall-down mechanism requries that the magnetic field be almost dipolar. This is possible if the distortion of the dipole field $\mathbf{B}_0$ of the central body is negligible. This distortion is caused by a magnetic field $\delta\mathbf{B}$ which is produced by the current in the plasma.

Instead of considering $\delta B/B$, we consider an average value of the field distortion along a field line; the fractional change in mean field strength is

$$\frac{\displaystyle\int_{R_e} |B_0\, \delta B\, \mathrm{d}\tau|}{\displaystyle\int_{R_e} B_0^2\, \mathrm{d}\tau} \ll 1\,, \tag{51}$$

where the integration is over the volume enclosed between two magnetic surfaces characterized by R_e and $R_e + \delta R_e$, respectively (Figure 11). It can be shown that, in the worst case condition (51) reduces to (see Appendix A)

$$h \ll \Lambda \left(\frac{\mu}{\Omega R_0 R_2^2}\right)^2, \tag{52}$$

where

$$h\, \delta R_e \equiv \int_{R_e} \rho\, \delta\tau \tag{53}$$

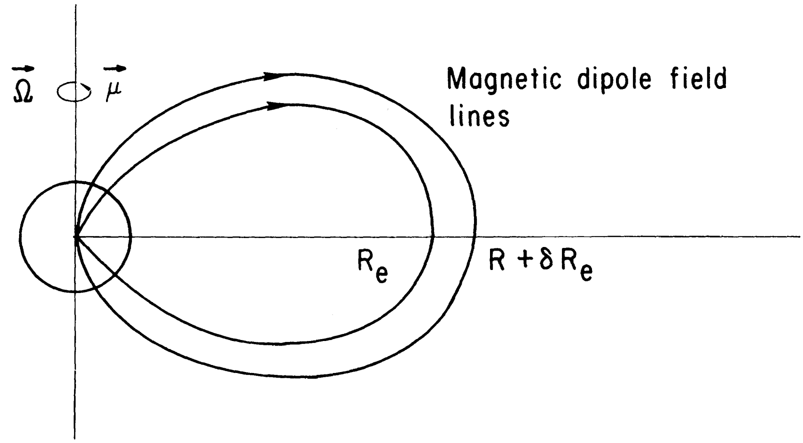


Fig. 11. A magnetic flux tube between R_e and $R_e + \delta R_e$.

is the mass of plasma in a flux tube δR_e. In deriving (52), we have assumed that

$$\frac{1}{h}\frac{\mathrm{d}h}{\mathrm{d}R_e} \approx 0 \tag{54}$$

and set

$$\frac{1}{\Lambda} \equiv \frac{1}{4}\frac{(1-\varepsilon\alpha)}{\sqrt{1-(R_0/R_2)}}\left\{\frac{32}{7}\left(1-\frac{3}{4}\frac{\varepsilon}{\alpha^2}-\frac{3}{4}\frac{\varepsilon^2}{\alpha}\right)+\frac{27\pi}{32}\left(1+\varepsilon\alpha-3\frac{\varepsilon^2}{\alpha}\right)\right\}, \tag{55}$$

with

$$\alpha \equiv \frac{R_0}{R_1}\left\{\sqrt{3}\left[\left(\frac{R_1}{R_0}\right)^3-\frac{1}{4}\right]^{1/2}-\frac{1}{2}\right\}, \tag{56}$$

$$\varepsilon = \frac{R_1}{R_2}, \qquad R_1 \equiv \left(\frac{2}{3}\frac{GM_c}{\Omega^2}\right)^{1/3}; \tag{57}$$

$R_0 \equiv (R_e)_{\min}$ and $R_2 \equiv (R_e)_{\max}$ denote the inner and outer edge of the rigidly rotating magnetosphere.

Now we can use the above result (52), the fact that $h \sim \rho R_e^2$ and that Λ is of the order of unity, to give an order of magnitude estimate of the upper limit on the plasma density. In the worst case, we must have

$$\rho \ll \left(\frac{B_c}{\Omega R_c}\right)^2 \frac{1}{L_0^2}\frac{1}{L_2^6}, \tag{58}$$

or, equivalently,

$$\Omega R_0 \ll B_{\mathrm{A}}(R_2)\,; \tag{59}$$

where

$$V_{\mathrm{A}}(R_2) \equiv \frac{B_2}{\sqrt{4\pi\rho_2}}$$

is the Alfvén speed in the equatorial plane at R_2, the outer edge of the magnetosphere.

Furthermore, we set

$$L_0 \equiv \frac{R_0}{R_c}\,, \qquad L_2 \equiv \frac{R_2}{R_c}\,, \tag{60}$$

$B_c \equiv \mu/R_c^3$ is the magnetic field strength at the equatorial surface of the central body. It is clear from (59) that requirement (51) is satisfied if the rotational speed at the inner edge is much less than the Alfvén speed at the outer edge of the magnetosphere.

6.2. Lower limit on the plasma density

Because the plasma is assumed to be in thermal equilibrium, a lower limit is set on the plasma density by the condition that the mean free path for Coulomb collisions ($\lambda_f(\mathrm{cm}) \approx 3.8 \times 10^{11}\, T_{(\mathrm{eV})}^2$) be less some characteristic distance scale.

If we take this typical distance to be on the order of R_e, the distance on the equatorial plane from the center of the central object to a given field line, then the above condition becomes

$$n(\mathrm{cm}^{-3}) \gg 3.8 \times 10^{11}\, \frac{\mathrm{T}_{(\mathrm{eV})}^2}{R_e(\mathrm{cm})}\,; \tag{61}$$

T being the plasma temperature in electron volts.

6.3. Upper limit on the plasma temperature

An upper limit is obtained by noting that conditions (58) and (61) must be simultaneously satisfied. This leads to

$$T_{(\mathrm{eV})} \ll 10^6\, \frac{B_c}{\Omega R_c^{1/2}}\, \frac{1}{L_0 L_2^3}\,. \tag{62}$$

6.4. Lower limit on the plasma temperature

The rotation of the magnetosphere induces an electrical field $\mathbf{E}$ given by

$$\mathbf{E} = -\frac{(\mathbf{\Omega} \times \mathbf{R}) \times \mathbf{B}}{c}\,. \tag{63}$$

This, in turn, is associated with a net charge density

$$\rho_q = \frac{\nabla \cdot \mathbf{E}}{4\pi} \, . \tag{64}$$

After substitution from (63) and (64), the quasi-neutrality requirement

$$|\rho_q| \ll \rho_i \approx \frac{q_i}{m_i}\,\rho \tag{65}$$

becomes

$$\rho(R_e, \theta) \gg \frac{\Omega m_i}{2\pi c q_i}\,|B_2(R_e, \theta)| \, . \tag{66}$$

From (49),

$$\rho(R_e, \theta) = g(R_e) \exp\left[-\frac{\psi(R_e, \theta)}{a^2}\right];$$

and from (A13)

$$g(R_e) = \frac{\sqrt{3}}{(2\pi)^{3/2}}\,\frac{h}{R_e^2}\,\sqrt{1 - \left(\frac{R_1}{R_e}\right)^3}\,\left(\frac{\Omega R_e}{a}\right) \exp\left[\frac{\psi(R_e, \frac{1}{2}\pi)}{a^2}\right],$$

we obtain

$$\rho(R_e, \theta) = \frac{\sqrt{3}}{(2\pi)^{3/2}}\,\frac{h}{R_e^2}\,\sqrt{1 - \left(\frac{R_1}{R_e}\right)^3}\,\left(\frac{\Omega E_e}{a}\right) \exp\left[-\left\{\frac{\psi(R_e, \theta) - \psi(R_e, \frac{1}{2}\pi)}{a^2}\right\}\right]. \tag{67}$$

Since $\rho(R_e, \theta)$ is minimum at the '$\frac{2}{3}$' points, the quasi-neutrality condition (66) is satisfied everywhere on the field line R_e if it is satisfied at the '$\frac{2}{3}$' points $\theta = \theta_{2/3}$. Hence,

$$\rho(R_e, \theta_{2/3}) \gg \frac{\Omega m_i}{2\pi c q_i}\,|B_2(R_e, \theta_{2/3})| \, . \tag{68}$$

This, according to (67), imposes a lower limit on the plasma temperature. In terms of the quantity h (average mass of plasma in a flux tube) defined in (53), the quasi-neutrality condition (68) becomes in the worst case (Appendix B)

$$h \gg \Lambda' \,\frac{a}{c}\,\frac{m_i \mu}{q_i}\,\frac{R_2}{(R_1^3 R_2)^{3/4}} \exp\left[\frac{1}{2}\left(\frac{R_2}{a}\right)^2 \nu\right], \tag{69}$$

where

$$\Lambda' \equiv 2\sqrt{\frac{2\pi}{3}}\,\frac{1 - \frac{3}{2}\varepsilon^{3/4}}{\sqrt{1 - \varepsilon^3}} \, , \tag{70}$$

$$\nu = 1 + 3\varepsilon^3 - 4\varepsilon^{9/4} \tag{71}$$

and, as before

$$\varepsilon \equiv \frac{R_1}{R_2}\,, \qquad R_1 \equiv \left(\frac{2GM_c}{3\Omega^2}\right)^{1/3}.$$

Requirements (69) and (52) are simultaneously satisfied if

$$\Lambda' \; \frac{a}{c} \; \frac{m_i \mu}{q_i} \; \frac{R_2}{(R_1^3 R_2)^{3/4}} \; \exp\left[\frac{1}{2}\left(\frac{\Omega R_2}{a}\right)^2 \nu\right] \ll \Lambda \left(\frac{\mu}{\Omega R_0 R_2^2}\right)^2,$$

which can be recast in the form

$$X\, e^{-X} \gg \delta\,, \tag{72}$$

where

$$X \equiv \left(\frac{\Omega R_2}{a}\right)^2 \nu \tag{73}$$

and

$$\delta \equiv \nu \left(\frac{\Lambda'}{\Lambda}\right)^2 \left(\frac{\Omega R_c}{c}\right)^4 \left(\frac{\Omega}{\omega g_c}\right)^2 \frac{L_0^4}{L_1^{9/2}} \; L_2^{21/2}\,; \tag{74}$$

where, as before,

$$\omega_{g_c} = \frac{q_i B_c}{m_i c}\,, \qquad B_c = \frac{\mu}{R_c}\,,$$

$$L_0 = \frac{R_0}{R_c}\,, \qquad L_1 = \frac{R_1}{R_c}\,, \qquad L_2 = \frac{R_2}{R_c}\,.$$

Using (72) and (73), we find that the thermal speed of ions

$$a \approx \left[\frac{(z_i + 1)kT}{m_i}\right]^{1/2}$$

must satisfy the inequality

$$a \gg \frac{\Omega R_2}{\sqrt{\dfrac{X_0}{\nu}}}\,, \tag{74'}$$

i.e.,

$$T_{(\mathrm{eV})} \gg 10^{-12} \; \frac{L_2^2}{2\,\dfrac{X_0}{\nu}} \; (\Omega R_c)^2\,, \tag{75}$$

where X_0 is the larger root of $X\,e^{-x} = \delta$. Hence, in order for a quasi-neutral plasma, which is in thermal equilibrium, to rotate rigidly in the slightly disturbed magnetic dipole field of a rotating central body, the plasma temperature and density must satisfy the inequalities

$$10^{-12}\,\frac{L_2^2}{2X_0/v}\,(\Omega R_c)^2 \ll T_{(\mathrm{eV})} \ll 10^6\,\frac{B_c}{\Omega R_c^{1/2}}\,\frac{1}{L_0^2 L_2^6}\,, \tag{76}$$

$$10^{11}\,\frac{T_{(\mathrm{eV})}^2}{R_c} \ll n_p(\mathrm{cm}^{-3}) \ll 10^{24}\left(\frac{B_c}{\Omega R_c}\right)^2\frac{1}{L_0^2 L_2^3}\,. \tag{77}$$

Observe that these are very weak constraints. As an illustration, we will apply the above results to Saturn. In this case we have:

$GM_c = 3.79 \times 10^{22}\,\mathrm{cm}^3\,\mathrm{s}^{-2}$,

$R_c = 6.0 \times 10^9$ cm,

$L_0 = 1.18$ (inner edge of the innermost ring),

$L_2 = 10$ (limit of plasma corotation),

$Z_i = 1$ (ions are mostly protons),

$B_c \sim G$ (magnetic field at the equatorial surface of Saturn during cosmogonic times),

$\Omega \approx 1.64 \times 10^{-4}\,\mathrm{s}^{-1}$ (plasma corotating with the planet).

Hence,

$$L_1 \approx 1.63\,, \qquad \varepsilon = \frac{L_1}{L_2} = 0.163\,, \qquad \alpha \approx 1.57\,,$$

$$v \approx 0.945\,, \qquad \frac{1}{\Lambda} = 1.48\,, \qquad \Lambda' = 1.12\,,$$

$$\omega_{g_c} = 9.56 \times 10^3\,s^{-1}\,, \qquad \delta = 6 \times 10^{-24}\,, \qquad X_0 \approx 57.5\,.$$

Using these results in (76) and (77), we can obtain an order of magnitude estimate of the range that the plasma temperature and number density can have

$$10^{-1} \ll T_{(\mathrm{eV})} \ll 10^2\,, \tag{78}$$

$$10T_{(\mathrm{eV})}^2 \ll n_p(\mathrm{cm}^3) \ll 10^6\,; \tag{79}$$

where $n_p = n_i \approx n_e$ being the plasma density in particles per cm^3.

We see that (78) and (79) give the conditions that the plasma density and temperature must satisfy in order for quasi-neutrality of the rigidly rotating plasma (in thermal equilibrium) to hold; and for the distortion of the dipole field of Saturn, caused by the plasma currents, to be negligible.

7. Discussions and Conclusions

In this paper we have reviewed the two-thirds fall-down mechanism and presented an approximate model for the dust distribution and neutralization at the '$\frac{2}{3}$' points along a given field line.

In this model, we have restricted ourselves to small gyroradii so that the charged dust particles are constrained to slide along the **B** field lines. Moreover, interparticle collisions were neglected and the initial kinetic energy of the infalling dust particles on a given field line was assumed to be large enough to surmount the hilltops of the potential ('$\frac{2}{3}$' points). Consequently, the density of the dust particles was found to be minimum at the bottom of the potential well (equatorial plane) and maximum at the '$\frac{2}{3}$' points.

The relationship between the charge on a dust particle and the ratio of dust to plasma density, n_D/n_p, was obtained from the current (ions + electrons) balance to the surface of the grain. In obtaining (26) we have neglected photoemission, field emission, secondary emission, etc., and considered only charging currents due to ions and electrons collection.

We have then studied the conditions under which rigid notation of a quasi-neutral plasma in thermal equilibrium is possible in a slightly perturbed magnetic dipole field.

Toward this end, we have determined the plasma distribution along a given field line. We have found that in thermal equilibrium, the plasma density is minimum at the '$\frac{2}{3}$' points and maximum at the equatorial plane.

The requirements of a quasi-neutrality, i.e., $\rho_q \ll \rho_{p_i}$, of thermal equilibrium and of slight distortion of the dipole field, i.e., $|\delta B|/|B_0| \ll 1$, impose limits on the plasma parameters. For Saturn, we have found that, for rigid corotation, the plasma temperature and density must satisfy

$$10^{-1} \ll T_{(\mathrm{eV})} \ll 10^2 , \qquad 10 T_{(\mathrm{eV})}^2 \ll n(\mathrm{cm}^{-3}) \ll 10^6 .$$

We conclude by noting that for a small fraction of the dust (the corotating component) the 'parallel' potential dominates and if $v_{\parallel}$ (the component of the grain velocity along the field line) has the right form of distribution there can be an effect, but that this effect is small ($<10\%$). However, if the formation of rings, satellite, etc., is dominated by inelastic collisions, then a small systematic disturbance of the kind we demonstrate may have a major effect on the final distribution of matter, even though most of the dust and plasma actually end up in the equatorial plane.

Moreover, if gyroradii are large, an important confining role may be played by the electrostatic field needed to maintain quasi-neutrality.

Acknowledgements

We should like to thank Drs. Hannes Alfvén, D. Asoka Mendis, and Elden C. Whipple for helpful discussions. Michel Azar wishes to acknowledge support from NASA under Grant NSG 7623.

Appendix A

Evaluation of

$$\frac{\left|\int B_0\,\delta B\,\mathrm{d}\tau\right|}{\int B_0^2\,\mathrm{d}\tau}\,. \tag{A1}$$

(A) Evaluation of $\int B_0^2\,\mathrm{d}\tau$ (see Figure 11). Substituting for $\mathbf{B}_0$ from (47) and using the fact that on a given field line R_e, $R = R_e \sin^2\theta$, we can write

$$\int B_0^2\,\mathrm{d}\tau = \mu^2 \int \frac{3\cos^2\theta + 1}{R^6} R^2 \sin\theta\,\mathrm{d}\theta\,\mathrm{d}R\,\mathrm{d}\phi =$$

$$= 2\pi\mu^2 \int\limits_{\xi}^{\pi-\xi} \mathrm{d}\theta \int\limits_{R_e\sin^2\theta}^{(R_e+\delta R_e)\sin^2\theta} \frac{3\cos^2\theta + 1}{R_e^4 \sin^2\theta} \sin\theta\,\mathrm{d}R =$$

$$= \frac{2\pi\mu^2}{R_e^4}\,\delta R_e \int\limits_{\xi}^{\pi-\xi} \frac{3\cos^2\theta + 1}{\sin^5\theta}\,\mathrm{d}\theta = \frac{2\mu^2}{R_e^4}\,\frac{\cos t\xi}{\sin^4\xi}\,\delta R_e\,,$$

where ξ is such that $\sin^2\xi \equiv R_0/R_e$, as before R_0 being the inner limit of the magnetosphere. The above result can be put in the form

$$\frac{1}{2\pi}\int\limits_{R_e} B_0^2\,\mathrm{d}\tau = \frac{2\mu^2}{R_0^2}\,\frac{\delta R_e}{R_e^2}\left(1 - \frac{R_0}{R_e}\right)^{1/2}. \tag{A2}$$

(B) Evaluation of $\left|\int \mathbf{B}_0 \cdot \delta\mathbf{B}\,\mathrm{d}\tau\right|$. We assume that the current density $\mathbf{J}$ produced in the plasma is toroidal and is therefore given by $J = J_\phi \hat{\phi}$, where

$$J_\phi = -\frac{ca^2}{\mu} R_e^3 \sin^3\theta g'(R_e) \exp\left[-\frac{\psi(R_e,\theta)}{a^2}\right]. \tag{50}$$

As a result of the azimuthal symmetry of the problem we may take the observation point to be in the $x - z$ plane ($\phi = 0$). Therefore, the integration

$$\mathbf{A} = \frac{1}{c}\int \frac{\mathbf{J}(\mathbf{x})\,\mathrm{d}^3x'}{|\mathbf{x} - \mathbf{x}|}$$

leaves only the y-component of $\mathbf{A}$, i.e., A_ϕ. Hence, $\mathbf{A} = A_\phi \hat{\phi}$

$$A = A_\phi \hat{\phi}\,, \tag{A3}$$

where

$$A_\phi = \frac{1}{c} \int \frac{J_\phi(R', \theta') \cos\phi' \, \mathrm{d}^3 x'}{|\mathbf{x} - \mathbf{x}|} \, ; \tag{A4}$$

with J_ϕ being given by (50). Now

$$\int \mathbf{B}_0 \cdot \delta\mathbf{B} \, \mathrm{d}\tau = \int \mathbf{B}_0 \cdot \nabla \times \mathbf{A} \, \mathrm{d}\tau = \int \nabla \cdot (\mathbf{A} \times \mathbf{B}_0) \, \mathrm{d}\tau = \oint (\mathbf{A} \times \mathbf{B}_0) \hat{n} \, \mathrm{d}a =$$

$$= \int_{R_e + \delta R_e} \mathbf{A} \times \mathbf{B}_0 \hat{n} \, \mathrm{d}a - \int_{R_e} \mathbf{A} \times \mathbf{B}_0 \hat{n} \, \mathrm{d}a =$$

$$= \delta R_e \frac{\partial}{\partial R_e} \int \mathbf{A} \times \mathbf{B}_0 \hat{n} \, \mathrm{d}a \, ,$$

where $\mathrm{d}a = R \sin\theta \, \mathrm{d}\phi \, \mathrm{d}S$ and $\mathrm{d}S = (B_0/B_0 \theta) R \, \mathrm{d}\theta$. The above expression can, therefore, be written as

$$\left| \int \mathbf{B}_0 \, \delta\mathbf{B} \, \mathrm{d}\tau \right| = 2\pi \, \delta R_e \left| \frac{\partial}{\partial R_e} \int \frac{\mu(3\cos^2\theta + 1)}{R_e \sin^2\theta} A_\phi(R_e, \theta) \, \mathrm{d}\theta \right| , \tag{A5}$$

where we have substituted for B_0^2, $B_{0\theta}$ and used $R = R_e \sin^2\theta$. By virtue of

$$A\phi(R, \theta) = \frac{1}{c} \int \frac{J_\phi(R', \theta') \cos\phi' \, \mathrm{d}^3 x'}{|\mathbf{x} - \mathbf{x}|} =$$

$$= \frac{1}{c} \int R'^2 J_\phi(R', \theta') \, \mathrm{d}R' \, \mathrm{d}\cos\theta' \, \times$$

$$\times \int_0^{2\pi} \frac{\cos\phi' \, \mathrm{d}\phi'}{\sqrt{R^2 + R'^2 - 2RR' (\cos\theta \cos\theta' + \sin\theta \sin\theta' \cos\phi)}} \, .$$

The ϕ' integration yields

$$\int_0^{2\pi} \frac{\cos\phi' \, \mathrm{d}\phi'}{\sqrt{R^2 + R'^2 - 2RR' (\cos\theta \cos\theta' + \sin\theta \sin\theta' \cos\phi')}} =$$

$$= \frac{4}{\sqrt{a + b}} \left[\frac{(2 - k^2) K(k) - 2E(k)}{k^2} \right] ,$$

where

$$a \equiv R^2 + R^2 - 2RR' \cos\theta \cos\theta' \, , \qquad b \equiv 2RR' \sin\theta \sin\theta' \, ,$$

$$k^2 \equiv \frac{2b}{1 + b} \leq 1 \, ;$$

K and E being the complete elliptic integrals of the first and second kind, respectively.

For small k^2, the above ϕ' integral reduces to $\pi b/2(a+b)^{3/2}$. $A_\phi(R, \theta)$, therefore, becomes

$$A_\phi(R, \theta) = -\frac{\pi a^2}{\mu} R \sin\theta \int_{R_0}^{R_2} dR_e' [g'(R_e')Q(R_e')R_e'^6], \tag{A6}$$

where we have used $R' = R_e' \sin^2\theta'$, substituted for J_ϕ from (50) and put

$$Q(R_e') \equiv \int_{\xi}^{\pi-\xi} \frac{\sin^{13}\theta' \exp\left[-\dfrac{\psi(R_e'), \theta')}{a^2}\right] d\theta'}{[R^2 + R_e'^2 \sin^4\theta' - 2RR_e' \sin^2\theta' \cos(\theta+\theta')]^{3/2}}, \tag{A7}$$

where as before $\sin^2\xi \equiv R_0/R_e$

$$\psi(R_e, \theta) = -\left[\frac{GM_c}{R_e \sin^2\theta} + \tfrac{1}{2}\Omega^2 R_e^2 \sin^6\theta\right].$$

$g(R_e)$ and, therefore, $g'(R_e) = dg(R_e)/dR_e$ can be expressed in terms of $h = \int_{R_e} \rho\, du$ ($h\, \delta R_e$ is the mass of plasma in the flux tube δR_e).

$$h = \int_0^{2\pi} \int_{\xi}^{\pi-\xi} \rho R \sin\theta\, d\phi\, dS = 2\pi \int_{\xi}^{\pi-\xi} \rho \frac{B_0}{B_{0\theta}} R^2 \sin\theta\, d\theta.$$

Substituting for B_0, $B_{0\theta}$ and using

$$\rho = g(R_e) \exp\left[-\frac{\psi(R_e, \theta)}{a^2}\right],$$

we obtain

$$g(R_e) = \frac{h(R_e)}{2\pi R_e^2 P(R_e)}, \tag{A8}$$

where

$$P(R_e) \equiv \int_{\xi}^{\pi-\xi} \sin^4\theta \sqrt{3\cos^2\theta + 1} \exp\left[-\frac{\psi(R_e, \theta)}{a^2}\right] d\theta; \tag{A9}$$

(A7) and (A9) are of the form

$$\int_{\xi}^{\pi-\xi} H(\theta)\, e^{f(R_e, \theta)}\, d\theta,$$

with

$$f(R_e, \theta) \equiv -\frac{\psi(R_e, \theta)}{a^2} = \frac{\left[\dfrac{GM}{R_e \sin^2\theta} + \frac{1}{2}\Omega^2 R_e^2 \sin^6\theta\right]}{a^2} .$$

We can evaluate the above integral using the method of steepest descent.

For $R_e \leq R_1$ the major contribution to the integral comes from $\theta = \xi$, $\pi - \xi$. For $R_e < R_1$, the major contribution comes from the equator $\theta + \frac{1}{2}\pi$ and from near the poles $\theta = \xi$, $\pi - \xi$. Moreover, for $R_e > R_1$, $f(R_e, \xi) < f(R_e, \frac{1}{2}\pi)$ if $R_e > \alpha R_1$ where

$$\alpha \equiv \frac{R_0}{R_1}\left\{-\frac{1}{2} + \sqrt{3}\left[\left(\frac{R_1}{R_0}\right)^3 - \frac{1}{4}\right]^{1/2}\right\} > 1 . \tag{A10}$$

After some manipulations, we can write $g'(R_e')$ and QR_e' as

$$Q(R_e') \approx \begin{cases} \dfrac{\gamma}{2}\left[\dfrac{1}{(R^2 + R_0^2 - 2RR_0\cos\theta)^{3/2}} + \dfrac{1}{(R^2 + R_0^2 + 2RR_0\cos\theta)^{3/2}}\right]\left(\dfrac{R_0}{R_e'}\right)^7 & \text{for } R_0 < R_e' \leq \alpha R_1 , \\ \dfrac{\exp\left[\dfrac{1}{a^2}\left(\dfrac{GM}{R_e'} + \frac{1}{2}\Omega^2 R_e'^2\right)\right]}{(R^2 + R_e'^2 + 2RR_e' \, sin\,\theta)^{3/2}} \left(\dfrac{\pi a^2}{\dfrac{e}{2}\,\Omega^2 R_e'^2 - \dfrac{GM}{R_e'}}\right)^{1/2} & \text{for } \alpha R_1 < R_3' \leq R_2 ; \end{cases} \tag{A11}$$

and

$$g'(R_e') \approx \begin{cases} \dfrac{h}{8\pi\gamma R_0^{5/2}} \dfrac{1}{R_e'^{5/2}} & \text{for } R_0 < R_e' \leq \alpha R_1 , \\ \dfrac{g(R_e'}{R_e'}\left\{\left(\dfrac{\Omega R_e'}{a}\right)^2\left[1 - \dfrac{3}{2}\left(\dfrac{R_1}{R_e'}\right)^3\right] + \dfrac{R_e'^3 - \frac{5}{2}R_1^3}{R_e'^3 - R_1^3}\right\} & \alpha R_1 < R_e' \leq R_e ; \end{cases} \tag{A12}$$

with

$$g(R_e'') \approx \frac{h}{2\pi R_e'^2}\left(\frac{\frac{1}{2}\Omega^2 R_e'^2 - GM/R_e'}{\pi a^2}\right)^{1/2} \exp\left[\frac{\psi(R_e, \frac{1}{2}\pi}{a^2}\right] \qquad \text{for } \alpha R_1 < R_e' \leq R_2 , \tag{A13}$$

$$\gamma \equiv \frac{\exp\left[\dfrac{GM}{a^2 R_0}\right]}{\dfrac{GM}{a^2 R_0}} ; \tag{A14}$$

where we have assumed here that $(1/h)\,(\mathrm{d}h/\mathrm{d}R_e) \approx 0$. With the help of (A11), (A12), (A13), and (A14), A_ϕ becomes

$$A_\phi(R,\theta) \simeq -\frac{\pi a^2}{\mu} R\sin\theta \left\{ \int_{R_0}^{\alpha R_1} \mathrm{d}R_e'[g'(R_e')Q(R_e')R_e'^6] + \right.$$

$$\left. + \int_{\alpha R_1}^{R_2} \mathrm{d}R_e'[g'(R_e')Q(R_e')R_e'^6] \right\} =$$

$$= -\frac{a^2 h}{8\mu}\frac{R_0^{9/2}}{R_1^{1/2}}\left[1 - \left(\frac{R_0}{\alpha R_1}\right)^{1/2}\right] R\sin\theta \left[\frac{1}{(R^2 + 2RR_0\cos\theta + R_0^2)^{3/2}} + \right.$$

$$\left. + \frac{1}{(R^2 - 2RR_0\cos\theta - R_0^2)^{3/2}}\right] + \frac{a^2 R\sin\theta}{2\mu} \times$$

$$\times \int_{\alpha R_1}^{R_2} \frac{hR_e'^3}{R^2 + R_e'^2 + 2RR_e'\sin\theta)^{3/2}} \left[\left(\frac{\Omega}{a}\right)^2 \frac{R_e'^3 - \frac{3}{2}R_1^3}{R_e'} + \right.$$

$$\left. + \frac{R_e'^3 - \frac{5}{2}R_1^3}{R_e' - R_1^3}\right] \mathrm{d}R_e' \,.$$

Using this result into (A5) we obtain

$$\frac{1}{2\pi}\left|\int \mathbf{B}_0 \cdot \partial\mathbf{B}\,\mathrm{d}\tau\right| \approx \tfrac{1}{2}(\Omega R_2)^2 h(\delta R_e)\,(1 - \alpha\varepsilon)\left[\frac{32}{7}\frac{R_e}{R_2}\left(1 - \frac{3}{4}\frac{\varepsilon}{a^2} - \frac{3}{4}\frac{\varepsilon^2}{\alpha}\right) + \right.$$

$$\left. + \frac{27\pi}{32}\left(1 + \varepsilon\alpha - e\,\frac{\varepsilon^2}{\alpha}\right)\right] + \frac{1}{2}\,a^2 h(\delta R_e) \times$$

$$\times \left\{\frac{16}{7}\frac{R_e}{R_1}\left[\sqrt{3}\tan^{-1}\frac{2/\varepsilon - \alpha}{\sqrt{3}\left[1 + \left(1 + \frac{2}{\varepsilon}\right)\left(\frac{1+\alpha}{3}\right)\right]} - \right.\right.$$

$$\left. - \frac{1}{2}\log\left[\left(\frac{1 - 1/\alpha}{1 - \varepsilon}\right)^3 \frac{1 - \varepsilon^3}{1 - 1/\alpha^3}\right] - 5\left(\frac{1}{\alpha} - \varepsilon\right)\right] +$$

$$+ \frac{27\pi}{16}\log\alpha\varepsilon\left(\frac{1 - \varepsilon^3}{1 - 1/\alpha^3}\right)^{1/2} +$$

$$+ \frac{3}{4}a^2 h(\delta R_e)\left(\frac{R_0}{R_1}\right)^{1/2}\left[1 - \left(\frac{R_0}{\alpha R_1}\right)^{1/2}\right], \qquad \text{(A15)}$$

where $\varepsilon \equiv R_1/R_2$ and α is defined in (A10). Since we are dealing with a cool plasma, we may assume that $a/\Omega R_2 \ll 1$ and consequently, we may neglect the last two terms on the right-hand side of (A15).

$$\frac{1}{2\pi}\left|\int \mathbf{B}_0 \cdot \delta\mathbf{B}\, \mathrm{d}\tau\right| \approx \tfrac{1}{2}(\Omega R_2)^2 h(\delta R_e)(1-\alpha\varepsilon)\left[\frac{32}{7}\frac{R_e}{R_2}\left(1-\frac{3}{4}\frac{\varepsilon}{\alpha^2}-\frac{3}{4}\frac{\varepsilon^2}{\alpha}\right)+ + \frac{27\pi}{32}\left(1+\varepsilon\alpha-3\frac{\varepsilon^2}{\alpha}\right)\right]; \tag{A16}$$

or, by dividing by $2\pi^{-1}\int B_0^2\, \mathrm{d}\tau$ given in (A2),

$$\frac{\left|\int \mathbf{B}_0 \cdot \delta\mathbf{B}\, \mathrm{d}\tau\right|}{\int B_0^2\, \mathrm{d}\tau} \approx (\Omega R_e)^2\left(\frac{R_0 R_2}{\mu}\right)^2 h\, \frac{(1-\alpha\varepsilon)}{\sqrt{1-R_0/R_e}}\left[\frac{32}{7}\frac{R_e}{R_2}\left(1-\frac{3}{4}\frac{\varepsilon}{\alpha^2}-\frac{3}{4}\frac{\varepsilon^2}{\alpha}\right)+ + \frac{27\pi}{32}\left(1+\varepsilon\alpha-3\frac{\varepsilon^2}{\alpha}\right)\right]. \tag{A17}$$

It is clear that this is maximum at $R_e = R_{e_{\max}} = R_2$; hence,

$$\left.\frac{\left|\int \mathbf{B}_0 \cdot \delta\mathbf{B}\, \mathrm{d}\tau\right|}{\int B_0^2\, \mathrm{d}\tau}\right|_{R_e} \ll 1 \quad \text{if} \quad \left.\frac{\left|\int \mathbf{B}_0 \cdot \mathbf{B}\right|}{\int B_0^2\, \mathrm{d}\tau}\right|_{R_0} \ll 1;$$

namely, if

$$\frac{1}{4}(\Omega R_2)^2\left(\frac{R_0 R_2}{\mu}\right)^2 h\, \frac{1-\alpha\varepsilon}{\sqrt{1-R_0/R_2}}\left[\frac{32}{7}\left(1-\frac{3}{4}\frac{\varepsilon}{\alpha^2}-\frac{3}{4}\frac{\varepsilon^2}{\alpha}\right)+ + \frac{27\pi}{32}\left(1+\varepsilon\alpha-3\frac{\varepsilon^2}{\alpha}\right)\right] \ll 1, \tag{A18}$$

which can be rewritten as

$$h \ll \Lambda\left(\frac{\mu}{\Omega R_0 R_2^2}\right)^2, \tag{A19}$$

where

$$\frac{1}{\Lambda} = \frac{1}{4}\frac{1-\alpha\varepsilon}{\sqrt{1-R_0/R_2}}\left[\frac{32}{7}\left(1-\frac{3}{4}\frac{\varepsilon}{\alpha^2}-\frac{3}{4}\frac{\varepsilon^2}{\alpha}\right)+ + \frac{27\pi}{32}\left(1+\alpha\varepsilon-3\frac{\varepsilon^2}{\alpha}\right)\right]; \tag{A20}$$

whereas before

$$\varepsilon \equiv \frac{R_1}{R_2}\,, \qquad R_1 = \left(\frac{2GM}{3\Omega^2}\right)^{1/3},$$

$$\alpha = \frac{R_0}{R_1}\left\{-\frac{1}{2} + \sqrt{3}\left[\left(\frac{R_1}{R_0}\right)^3 - \frac{1}{4}\right]^{1/2}\right\}.$$

Appendix B. Derivation of (69)

The requirement of quasi-neutrality $\rho_q \ll \rho_{q_i}$ is valid if

$$\rho_{\min} = \rho(R_e, \theta_{2/3}) \gg \frac{\Omega m_i}{2\pi c q_i}\,|B_z(R_e, \theta_{2/3})|\,, \tag{66}$$

where

$$|B_z(R_e, \theta_1) = \frac{2\mu}{(R_1^3 R_e)^{3/4}}\left[1 - \frac{3}{2}\left(\frac{R_1}{R_e}\right)^{3/4}\right], \tag{B1}$$

with $\sin\theta_{213} = (R_{213}/R_e)^{3/8}$. Using the results of Appendix A, we can write

$$\rho_{\min}(R_e) = \rho(R_e, \theta_{2/3}) \sim \exp\left[-\frac{m_i \psi_c(R_e, \theta)}{2kT}\right] =$$

$$= \frac{\sqrt{3}}{(2\pi)^{3/2}}\left(\frac{\Omega R_e}{a}\right)\frac{h}{R_e^2}\left(1 - \left(\frac{R_1}{R_e}\right)^3\right)^{1/2} \times$$

$$\times \exp\left\{-\frac{1}{2}\left(\frac{\Omega R_e}{a}\right)^2\left[1 + 3\left(\frac{R_1}{R_e^3}\right) - 4\left(\frac{R_1}{R_e}\right)^{9/4}\right]\right\}. \tag{B2}$$

With the help of (B1) and (B2), (66) becomes

$$\left(\frac{3}{2\pi}\right)^{1/2}\frac{cq_i}{2\mu m_i \Omega}\,\frac{(R_1^3 R_e)^{3/4}}{R_e^2}\,\frac{(1 - (R_1/R_e)^3)^{1/2}}{1 - \frac{3}{2}(R_1/R_e)^{3/4}}\,h\left(\frac{\Omega R_e}{a}\right) \times$$

$$\times \exp\left\{-\frac{1}{2}\left(\frac{\Omega R_e}{a}\right)^2\left[1 + 3\left(\frac{R_1}{R_e^3}\right) - 4\left(\frac{R_1}{R_e}\right)^{9/4}\right]\right\} \gg 1\,. \tag{B3}$$

Since the left-hand side of (B3) is a decreasing function of R_e, (B3) is valid at any value of R_e, if it is valid at $R_e = R_{e_{\max}} \equiv R_2$. Requirement (65) is, therefore, equivalent to

$$\frac{1}{2}\left(\frac{3}{2\pi}\right)^{1/2}\frac{(1 - \varepsilon^3)^{1/2}}{1 - \frac{3}{2}\varepsilon^{3/4}}\left(\frac{c}{a}\right)\frac{q_i}{m_i}\,\frac{(R_1^3 R_2)^{3/4}}{\mu R_2} \times$$

$$\times \exp\left\{-\frac{1}{2}\left(\frac{\Omega R_2}{a}\right)^2 [1 + 3\varepsilon^3 - 4\varepsilon^{9/4}]\right\} h \gg 1\,. \tag{B4}$$

Appendix C

Solution of

$$\frac{\partial f}{\partial t} \pm \omega \frac{\partial f}{\partial \psi} + \gamma f = S(\varepsilon, \phi). \tag{9}$$

Taking the Fourier transform of (9), we obtain

$$\frac{\partial f_n}{\partial t} + (\gamma \pm in\omega) f_n = S_n, \tag{C1}$$

where

$$f_n(\varepsilon, t) = \frac{1}{2\pi} \int_0^{2\pi} f(\varepsilon, \psi, t)\, e^{-in\phi}\, \mathrm{d}\phi, \tag{C2}$$

$$S_n(\varepsilon) = \frac{1}{2\pi} \int_0^{2\pi} f(\varepsilon, \psi)\, e^{-in\phi}\, \mathrm{d}\phi \tag{C3}$$

and

$$f(\varepsilon, \psi, t) = \sum_{n=-\infty}^{+\infty} f_n(\varepsilon, t)\, e^{in\phi}, \tag{C4}$$

$$S(\varepsilon, \psi) = \sum_{n=-\infty}^{+\infty} S_n(\varepsilon)\, e^{in\phi}. \tag{C5}$$

The solution of Equation (C1) which satisfies the initial condition $f_n(\varepsilon, 0) = 0$ is of the form

$$f_n(\varepsilon, t) = \frac{S_n(\varepsilon)}{\gamma \pm in\omega} \left[1 - e^{-(\gamma \pm in\omega)t}\right], \tag{C6}$$

where, according to (12),

$$S_n(\varepsilon) = \frac{S_0}{2\pi} \int_{\psi' \varepsilon E(\varepsilon)} e^{-in\psi'}\, \mathrm{d}\psi'. \tag{C7}$$

Substitution of (C6) and (C7) into (C4) yields

$$f(\varepsilon, \phi, t) = \frac{1}{2\pi} \sum_n \left\{ \frac{S_0}{\gamma \pm in\omega}\, e^{in\phi}\left[1 - e^{-(\gamma \pm in\omega)t}\right] \int_{\phi' \varepsilon E(\varepsilon)} e^{-in\phi'}\, \mathrm{d}\phi' \right\}. \tag{C8}$$

By taking account of (d), we can write the final result as

$$f(\varepsilon, \phi, t) = f(\varepsilon, t) = \Gamma(1 - e^{-\gamma t})\, \frac{\phi_0(\varepsilon)}{2\pi}, \tag{C9}$$

where

$$\phi_0(\varepsilon) \equiv \int_{\phi' \varepsilon E(\varepsilon)} \mathrm{d}\phi' = \begin{cases} 2\pi & \text{for} \quad \varepsilon \leq \frac{1}{2}V_0^2\,, \\ \dfrac{4\omega \displaystyle\int_{s_0}^{s_1} \mathrm{d}\bar{s}}{\sqrt{2[\varepsilon - \psi(\bar{s})]}} & \text{for} \quad \varepsilon > \frac{1}{2}V_0^2\,; \end{cases} \tag{C10}$$

s_1 being defined in (11) and s_0 the positive root of

$$\varepsilon - \tfrac{1}{2}V_0^2 = \psi(s_0)\,.$$

References

Alexander, A. F. O'D.: 1953, *Brit. Astron. Assoc. J.* **64**, 26.

Alexander, A. F. O'D.: 1962, *The Planet Saturn, a History of Observation, Theory and Discovery,* Macmillan, New York.

Alfvén, H. and Arrhenius, G.: 1976, *Evolution of the Solar System*, NASA SP-345.

Baxter, D. and Thompson, W. B.: 1971, 'Jet Stream Formation Through Inelastic Collisions', in T. Gehrels (ed.), *Physical Studies of Minor Planets*, NASA SP-207.

Baxter, D. and Thompson, W. B.: 1973, *Astrophys. J.* **183**, 323.

Goretz, C. K. and Ip, W.-H.: 1984, *Geophys. Res. Letters* **11**, 349.

Mendis, D. A. and Axford, W. L.: 1974, *Rev. Earth Planetary Sci.* **2**, 419.

Mendis, D. A., Northrop, T. G., and Whipple, E. C.: 1985, *J. Geophys. Res.* **40**, 7405.

Whipple, E. C.: 1981, *Rep. Progress Phys.* **44**, 1197.

THE ROLE OF MAGNETOHYDRODYNAMICS IN HELIOSPHERIC SPACE PLASMA PHYSICS RESEARCH*

MURRAY DRYER, ZDENKA KOPAL SMITH,

Space Environment Laboratory, National Oceanic and Atmospheric Administration, Boulder, U.S.A.

and

SHI TSAN WU

Center for Space Plasma and Aeronomic Research, University of Alabama, Huntsville, U.S.A.

(Received 12 August, 1987)

Abstract. Magnetohydrodynamics (MHD) is a fairly recent extension of the field of fluid mechanics. While much remains to be done, it has successfully been applied to the contemporary field of heliospheric space plasma research to evaluate the 'macroscopic picture' of some vital topics via the use of conducting fluid equations and numerical modeling and simulations. Some representative examples from solar and interplanetary physics are described to demonstrate that the continuum approach to global problems (while keeping in mind the assumptions and limitations therein) can be very successful in providing insight and large scale interpretations of otherwise intractable problems in space physics.

1. Introduction

Looking back in time to the dawn of the space age, one is intrigued by the controversies that existed at the time. Although today's problems are different, controversies still persist despite the accumulation of data of all kinds since the 1950's. One famous example of an early space-age controversy is the Gold–Parker debate (Gold, 1959; Jastrow, 1959) on the structure of transient disturbances in interplanetary space that can produce modulation of solar and galactic cosmic-ray particle arrival at Earth. Part of the debate predicted that global magnetic 'tongues' should be ejected from solar flares, and that, later, these disturbances should engulf the Earth. Another part of the debate predicted a particle evaporation from the Sun, a phenomenon that would become known as the 'solar wind'. This outward-flowing plasma, together with the entrained magnetic fields, is drawn directly from the Sun and may be deformed by a propagating heliospheric disturbance. Jastrow (1959), with additional insight, allowed that each point of view probably contained some elements of the truth. A new, related controversy was also introduced whereby the continuum approach of the solar wind (Parker, 1958) was challenged by the exospheric (or particle) approach to the solar plasma outflow.

We refer to these controversies because it is our belief that the latter one, in particular, continues in more advanced forms to the present day. We refer, for example, to one view of space plasma research that is directed to the search for the understanding of fundamental plasma kinetics by focusing on local, *in situ*, processes. This research has

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 407–425.

advanced because of the ever-increasing temporal and spatial resolution in the measurement of plasma, magnetic, and electric field data. Since this view is generally incapable of explaining larger-scale (larger, say, than the particle gyro-radii) global processes, the continuum or collective interaction approach has some appeal for the elucidation of the 'macroscopic picture' of heliospheric processes that range in scale from the immediate vicinity of the Sun to the whole solar system. Specifically, by taking the first three moments of the Vlasov equation, one arrives at the well-known equations of motion (plus the assumed closure equation for a constitutive plasma equation) that become solvable, given a set of appropriate boundary conditions.

Progress towards the attainment of certain goals espoused by the kinetic theorists has been made possible by the recent development of particle simulation codes. For example, the various physical processes within various categories of parameters (plasma beta, magnetic field orientation relative to the shock plane, and Mach number) associated with electrostatic and ion acoustic collisionless shock waves have been delineated. However, particle codes are restricted to spatially-limited physical regimes (cf. shock structure, reconnection zones, double layers); i.e., they are incapable of examining the spatially comprehensive 'macroscopic picture'.

Progress has also been made toward satisfying the objectives of the MHD perspective. Various approaches have been tried. One such approach studied the problem morphologically utilizing the results of Pudovkin and Chertkov (1976) and Ivanov *et al.* (1980). These investigations suggested an association (as an extention of the Gold (1959) thesis) between the polarity of predominant magnetic topologies in the vicinity of the flaring regions on the Sun and the subsequent interplanetary magnetic field (IMF) polarity observed at the Earth. The morphological model also related the Sun's magnetic topology with oblate global interplanetary shock shapes. The interpretation of the IMF polarity has, however, not been generally accepted (cf. Tang *et al.*, 1985). Complementary, kinematic studies (cf., Hakamada and Akasofu, 1982; Akasofu *et al.*, 1985) have also been made for interplanetary disturbances. The latter kinematic studies, while providing some insight for complex three-dimensional MHD studies still in development, will not be considered here because of their non-MHD formulations.

Cuperman and Dryer (1985) investigated the use of equations derived from higher moments of the Vlasov equation (with appropriate Fokker–Planck formalism). These workers were interested in deriving generalized, non-Maxwellian, non-equilibrium expressions (reducible to the classical Spitzer–Braginskii expressions) for thermal conductivities of electrons and protons in a collisionless plasma. This usage of the higher moment equations, however, has not been extended to the general class of MHD solutions for the 'macroscopic picture'.

During the past 20 years, substantial improvements have been achieved in the MHD class of solutions that use the classical three-moment (conservation of mass, momentum, and energy) equations in addition to the equation of state and Maxwell's equations. Numerical techniques (Rubin and Burstein, 1967; Brackbill and Pracht, 1973; Han *et al.*, 1982; Hu and Wu, 1984; Wu and Wang, 1987) and ever-increasing computer storage and speed improvements have produced an expanding group of applications to solar, interplanetary and magnetospheric physics (cf. Brecht *et al.*, 1982; Forbes and Priest, 1983).

The purpose of this paper is to demonstrate the application of MHD numerical simulation to some large-scale, global heliospheric problems in space plasma physics. A few examples of these applications will be presented.

2. Discussion

The MHD method has been used in a number of studies in recent years thereby demonstrating its vitality and utility in heliospheric space plasma physics research. The discussion below is divided into two parts: (I) Solar Aspects, and (II) Interplanetary Aspects.

(I) Solar aspects

The state-of-art numerical simulations of coronal mass ejections (CME's) and their extension to the initiation of propagating interplanetary disturbances are divided into three distinct approaches (cf. the review by Dryer, 1982), each with varying degrees of development:

(a) The MHD approach is generally considered to be established in its 'first generational' aspects. The initially-quiescent atmosphere has been assumed to be in the lowest energy state prior to an externally imposed perturbation. 'Second generational' aspects have recently been initiated and, because of their advanced numerical development, will be summarized in some detail.

(b) The second approach assumes that the initially-quiescent atmosphere is in a higher energy state and is subject to some form of instability.

(c) The third approach assumes that an impermeable, extraneous diamagnetic plasmoid (or an ensemble of cylindrical magnetically-confined plasmoids) is destabilized by photospheric motion and moves outward due to MHD buoyancy effects.

These three approaches, with main emphasis on (a), are summarized as follows.

(a) The first approach (the MHD paradigm) might be referred to as the 'coronal atmospheric response' to various perturbations at the base. Given a reasonable representation of the pre-event atmosphere and magnetic topology, the initial boundary value problem in an ionized plasma is solved producing a series of compressions, rarefactions, MHD waves and, with sufficiently strong forcing functions, even shocks that travel into interplanetary space. The early 'first generational' work has been confined to solving the response of an ideal plasma (neutral composition, no dissipation except at shocks) in two dimensions permeated initially with potential and force-free magnetic fields. This work contains both boundary perturbations as well as instantaneous energy release within the atmosphere (Nakagawa *et al.*, 1978, 1981; Wu *et al.*, 1978, 1982a; Steinolfson *et al.*, 1978; Dryer *et al.*, 1979; Dryer and Maxwell, 1979). It is of interest to note that two of these papers (Steinolfson *et al.*, 1978; Dryer *et al.*, 1979) inadvertently introduced (via an initial field rotation) a violation of the solenoidality condition (Low, 1986, private communication). However, in subsequent 'second generational' work, Wu *et al.* (1987c) demonstrated, via an improved state-of-art code (Hu and Wu, 1984; Wu *et al.*, 1983a; Wu and Wang, 1987), the coronal responses of cases when solenoidality ($\nabla \cdot \mathbf{B} = 0$) is deliberately violated. Comparing these results to those obtained correctly (i.e., when solenoidality is *not* violated), Wu *et al.* (1987c) showed that the error thus introduced into the energy density and plasma density profiles is small and, hence, does not invalidate the more general earlier conclusions concerning physical interpretations of mass and wave motion that follow solar flares.

Let us examine the MHD paradigm applied to certain aspects of solar and planetary physics. When one adds phenomenological Joule dissipation terms to the energy and induction equations, the MHD approach can simulate reconnection associated with opposing magnetic topologies within active regions on the Sun (Wu *et al.*, 1986) as well as within planetary magnetospheric tails (Forbes and Priest, 1983). The actual physics is approximated by a model that causes dramatic resistivity increases by as much as 4–5 orders of magnitude (Wu *et al.*, 1987b) when the induced electrical current increases (due to enhanced plasma density fluctuations and magnetic field gradients) beyond a critical value, say 10^{-3} A m^{-2}. Wu *et al.* (1986) use this result and show an example of an initially-quiescent, multipole potential field that is magnetically stressed when the footpoints are set into motion (at 1 km s^{-1}) toward the neutral line and subsequently experiences reconnection. Figure 1 shows the build-up of the excess energy (per unit depth in this 2D (two-dimensional) calculation) in the magnetic and thermal modes as a function of time. This figure (lower panels) also indicates the reconnection at an altitude of $\sim$ 4000 km above the photosphere, the motion of magnetic field lines toward and away from this zone, and the plasma acceleration upward and away from the reconnecting area.

We carry the above scenario to the next reasonable step. Following reconnection, some magnetic energy has been converted to kinetic and thermal energies (over and above the initial amount within the flaring zone). Consequently, we may devise another initially quiescent stratified coronal layer which, for simplicity, also contains an initially potential magnetic field that will be disturbed when this thermal pressure pulse is imposed. Clearly, the actual physics is much more complicated by asymmetries relative to existing loops (Harrison *et al.*, 1985) and/or with combined mass and magnetic flux from below the photosphere. Nevertheless, we may illustrate an approximate global coronal response by re-examining a simple pressure pulse model introduced some ten years ago ('first generation' CME modeling). Figure 2 shows the spatial distribution of the excess density distribution in the corona subjected to a 40-fold temperature increase over a 4° extent in arc length at its base. The pulse was applied for 10 min. The 'snapshot' shown in the figure at t = 6 min demonstrates the re-distribution of coronal material wherein a compression (solid lines) along the outer rim is followed by a rarefaction (dotted lines). Other features are noted in the figure caption.

In another example, Figure 3 demonstrates the spatial and temporal coronal density response (Wu, 1985) when sub-photospheric mass is assumed to be injected, symmetrically, into the two legs of an initially-quiescent loop system. The lower panels show the development and propagation of a fast MHD shock into the interplanetary medium at

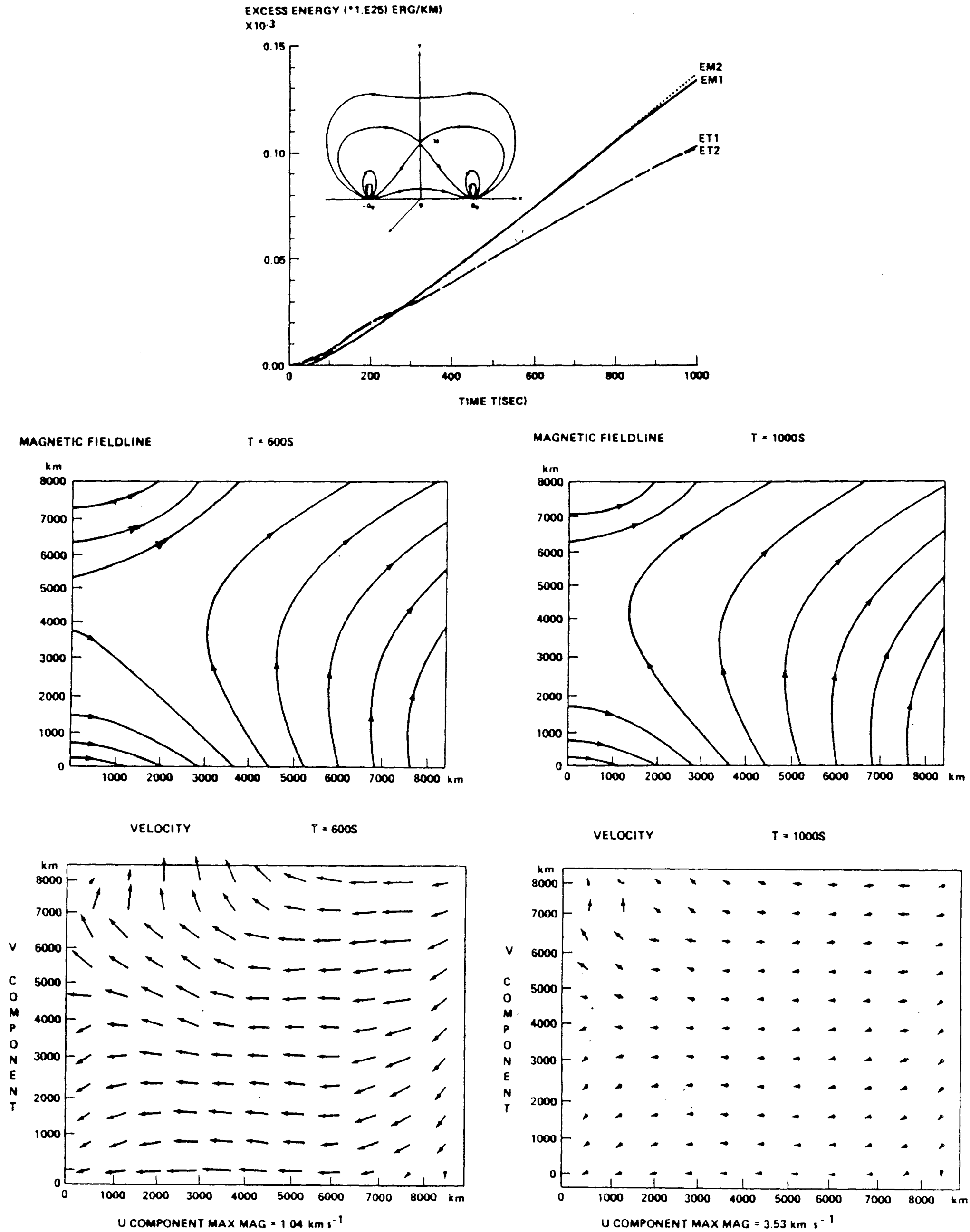


Fig. 1. Evolution of magnetic and thermal energy due to a 1 km s^{-1} converging motion of the footpoints of an initial multipole potential field in the solar atmosphere. The initial plasma beta (ratio of thermal to magnetic pressure) is 0.1 at $x = y = 0$, and the resistivity is increased whenever and wherever the enhanced current exceeds 10^{-3} A m^{-2} (assumed to be a 'critical' value). Lower panels show the magnetic field line positions and velocity vectors at two times (t = 600 and 1000 s) as a function of position (y = altitude). The reconnecting area is seen to be in the vicinity of $y \approx 4000$ km. (Wu *et al.*, 1986.)

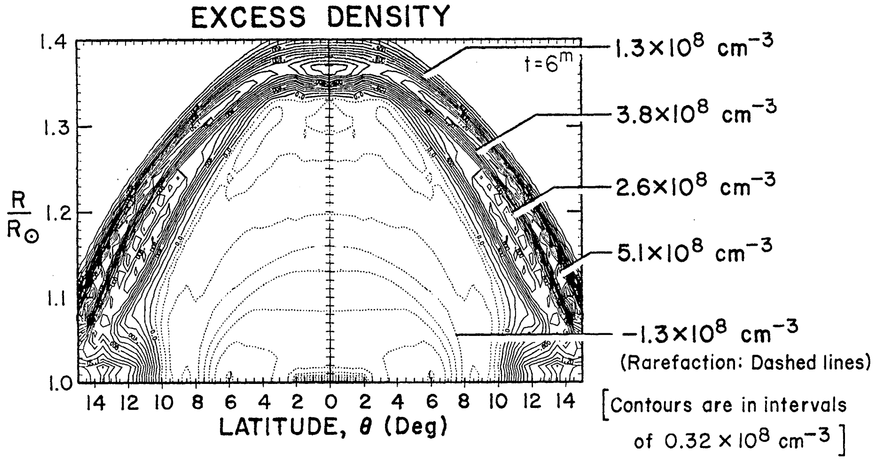


Fig. 2. MHD simulation of the excess density produced by a temperature pulse of 40×10^6 K for 10 min at the base of a representative corona model. The latitude θ refers to the angular distance from the center of the simulated flare pulse. The heliocentric distance R is normalized by the solar radius $R_\odot$ (where $R_\odot = 6.95 \times 10^5$ km). The 'snapshot' shown here is at an elapsed time of $t = 6$ min after initiation of the pulse. A rarefaction follows the density compression. Although the model is simplified by using two dimensions, it is believed that coronal transients from energetic flares are primarily bubble-shaped, i.e., three-dimensional, entities. A fast MHD shock wave was formed just ahead of the simulated density enhancement in this particular example because the assumed thermal energy input (presumed to have been derived from the originally-stored free (magnetic) energy) was particularly large. This shock then expands, bubble-like, into the corona, running ahead of the original 'piston', becoming a full-fledged interplanetary shock wave that propagates through the heliosphere, 'riding over' or superimposed upon the previously-emitted solar wind. (Dryer, 1982.)

the now-larger distances (6 $R_\odot$). The closed loop system continues to generate several, interacting, slow MHD shock waves. Injection of mass from the photosphere need not be symmetric as noted by Wu *et al.* (1982b) who also suggest that this kind of event may trigger neighbouring events not associated with flares or erupting prominences.

The above MHD paradigm discussion implicitly contains the requirement for an initially-quiescent (or better still, a moving, but steady-state) coronal atmosphere. Also, the temporal boundary conditions must assure the compatibility of: (a) divergence-free condition, and (b) the proper specification of the correct number of arbitrarily-assigned dependent variables plus those that must be calculated from the compatibility relationships. The former topic has been discussed in a number of different contexts (cf. Yeh and Dryer, 1985), and the latter topic has been developed in detail by Hu and Wu (1984), Wu and Wang (1987), and Wu *et al.* (1983a, b). As for the quiescent atmosphere, it is possible that some adaptation of the work accomplished by Pneuman and Kopp (1971) and Yeh and Pneuman (1977) may be adopted (Cuperman, 1986, private communication).

Since this first approach which is basically an MHD formalism has now been developed to the point wherein direct comparisons are possible with coronagraph observations, an initial study has been performed with a limited subset (5) of the

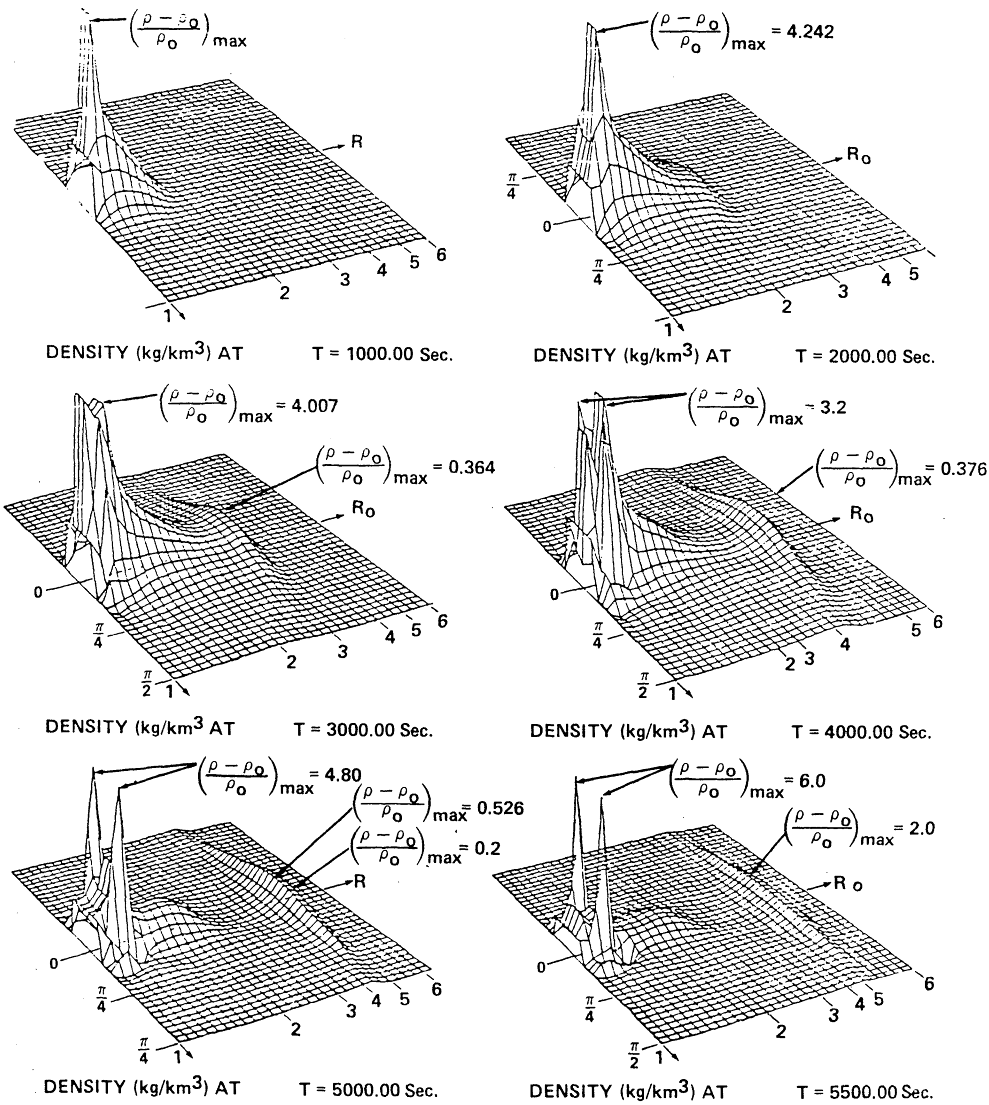


Fig. 3. Temporal and spatial development of an MHD-simulated coronal mass ejection (CME). The fractional density change, $(\rho - \rho_0)/\rho_0$, is plotted (from upper left, temporally, to ower right) as a function of solar latitude... relative to the center of a closed magnetic loop topology... and solar radius R_0 ($= R/R_{\odot}$). The assumed input 'cause' is mass injection from the photosphere into a loop system. As the fast MHD shock develops and moves into interplanetary space, the loop system undergoes interaction of confined slow MHD shocks with the spike-like peaks. (Wu, 1985.)

77 SKYLAB cases of CME's (Sime *et al.*, 1984). The conclusions reached by these authors have proved to be controversial (Dryer and Wu, 1985) and, thereby, indicated the urgent need for further comparisons of a larger data base using MHD simulations. It is important to note that this approach may be highly appropriate for solar flare applications.

(b) The second approach does not solve time-dependent equations as in the MHD case discussed above but obtains instead an analytic solution to stationary force balance equations. However, it is related to the MHD method in the sense that the dynamic coronal response is inferred. It differs (Low, 1981) from the above MHD model in that free energy is assumed to exist in a more realistic non-potential magnetic field. It also assumes that an instability, triggered by photospheric field line motions, would be sufficient to provide the impetus for a CME that would balloon upwards. Thus far, such studies have been limited to quasi-static evolution of coronal magnetic fields. Self-consistent dynamical motion has not been considered in terms of numerical simulations. This approach is possibly appropriate for CME's that are associated with neither flares nor prominence eruptions.

(c) The third approach is also a quasi-static force balance analysis which directs attention to an extraneous, diamagnetic plasma body (loop or bubble) that is immersed within (and possibly moving relative to) the ambient, outward-moving solar wind. This 'projectile' is acted upon by an MHD buoyancy force which results from the inhomogeneity of thermal and magnetic stresses that act upon this compressible extraneous body (Yeh, 1985). As in the case of the second approach, this method is primarily analytic and is not amenable in its present form to dynamical numerical simulation. This approach is probably more appropriate for prominence eruptions (with or without accompanying solar flares). These eruptions are called *disparition brusque* or 'disappearing filaments' when they occur on the solar disc and 'EPL' (eruptive prominence, limb) when they are observed at or near the plane-of-sky near the solar limb. This approach has formed the rationale for an increasing number of morphological models that resemble the case of planetary magnetospheric interaction with the solar wind. As noted above, these concepts represent a reincarnation of the Gold thesis.

A great deal of work has been done coordinating interdisciplinary studies of solar events and their consequences at multi-spacecraft positions in the ecliptic plane. A summary of some of this activity, under the aegis of the STIP project (formerly in SCOSTEP, presently in COSPAR), is described by Dryer and Shea (1986) and Shea and Dryer (1988). Our particular interest relates to the use of numerical simulations that start at or near the Sun and that are driven by forces assumed to exist in the lower corona (such as flare-generated shocks and disappearing filaments). These assumptions are often aided by observations of solar activity from slow drift type II radio bursts that provide shock wave velocity estimates; white light CME observations that provide angular extent and kinematic information; and solar flare diagnostics such as the soft X-ray flux from GOES and the UVSP, HXIS, and XRP data from SMM that provide temporal duration and thermal pressure pulse information. CME simulations of coronal disturbances that follow flares have been described in the references noted earlier. In some cases, the coronal response can be initiated by pre-flare mechanisms that are, as yet, not understood (Harrison *et al.*, 1985). Here again, the MHD approach can be used to describe the coronal response if sufficient diagnostic input conditions are available. It is important to point out that the MHD approach, by virtue of its initial boundary value nature, may be specifically tailored, according to the inferred or assumed physical cause, to a particular event by specifying the conditions observed to have prevailed during the initial time period.

(II) INTERPLANETARY ASPECTS

The modeling of steady-state structures quickened after the discovery of corotating interaction regions (CIRs) by Smith and Wolfe (1976, 1979) who used their solar wind plasma analyzer and magnetometer data from Pioneer 10 and Pioneer 11 to describe the interaction of corotating streams. Studies of multi-spacecraft testing of non-MHD models in this context were initiated with the evaluation of a 1D (one-dimensional) time-dependent gas dynamic model (Gosling *et al.*, 1976; Hundhausen and Gosling, 1976). These studies used data from an Earth-orbiting spacecraft as the primary input and Pioneer 10 and 11 data as the test-bed along a nearly-heliocentric radial with the Earth but at larger distances. The test, limited to velocity comparisons, was quite successful in demonstrating evolving streams. Testing of a $1\frac{1}{2}$D (one spatial dimension in radius and two components of velocity and IMF, respectively, in the ecliptic plane) time-dependent MHD code was performed, using Pioneer 11 data (input) and Pioneer 10 data (output), again along a nearly-radial line. The test, limited to velocity (V), density (n), temperature (T), and azimuthal component (B_ϕ) of the IMF (Dryer *et al.*, 1978; Smith *et al.*, 1981) was successful with V, n, and B_ϕ but less so in T (Figure 4). The generic build-up (i.e., model development) of CIRs and their potential interactions at large distances is demonstrated in Figure 5 and described in more detail in the figure caption.

Recent tests that use this same approach in a 1D time-dependent MHD (B_ϕ only) method-of-characteristics model (Whang, 1984; Whang and Burlaga, 1985a, b, 1986) have also been successful in describing the most important structures and their magnitudes, this time with the use of Voyager 1 and 2 and Pioneer 11 data. A 3D, time-independent MHD model was also applied to the case of nearly-radially aligned spacecraft so that the calculation was actually performed in the radial (r)-azimuthal (ϕ) plane. Steady corotating flow was assumed, thereby allowing transformation of the simulated output to radius (only) and time (Pizzo, 1982, 1983; Burlaga *et al.*, 1985) as in the earlier works.

The following time-dependent MHD models have been developed: a 2D model (Wu *et al.*, 1979; Dryer *et al.*, 1980; D'Uston *et al.*, 1981) and a $2\frac{1}{2}$D model (Wu *et al.*, 1983). In the latter model, the components of V and B perpendicular to the ecliptic plane are also explicitly considered. A multi-spacecraft test (using Prognoz 1 and Pioneer 9 data for comparison) was performed with the 2D model (D'Uston, 1981) for the August 1972 events. The $2\frac{1}{2}$D model has been tested (V and $|B|$ only) against ISEE-3 data during the August 1979 epoch (STIP Interval VII) with some success in predicting the major structures and amplitudes (Dryer *et al.*, 1986a) but also incurring some timing discrepancies.

Part of the August 1979 simulation is shown in Figure 6 to demonstrate the model's ability to simulate complex, compound interactions of flare-generated disturbances comingled with corotating streams as well as with other flare-generated flows (Smith

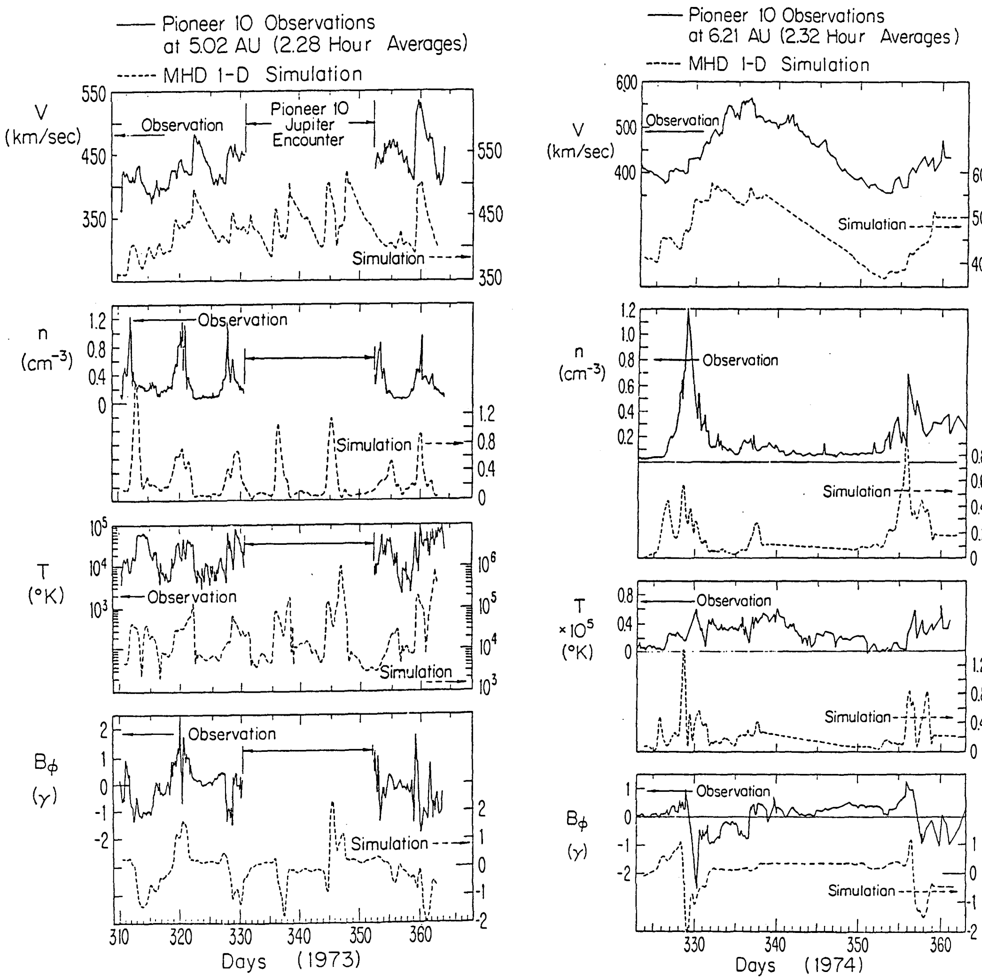


Fig. 4. *Left*: Corotating interaction regions (CIRs) observed by Pioneer 10 prior to, and following, its encounter with Jupiter's magnetosphere in 1973. Plotted here are: the solar wind velocity V; the solar wind proton density n; the proton temperature T; and the azimuthal component of the interplanetary magnetic field B_ϕ. MHD simulation is also shown; input conditions of the simulation were provided by using data from Pioneer 11, which was located directly sunward at 2.8 AU. The forward and reverse shocks that bound each CIR are seen most clearly in the density (n) measurements (solid lines). The MHD simulation (dashed lines) shows good agreement with the observations, thereby providing confidence that Jupiter continued to be buffeted by two CIRs during the time that Pioneer 10 was inside the magnetosphere. *Right*: the MHD comparison was repeated for 1974 when Pioneer 11 passed through the Jovian magnetosphere and Pioneer 10 was downstream, again radially-aligned, at 6.2 AU. The solar wind input data for the model were taken from the Pioneer 11 data prior to, and following, its passage through the Jovian system. (Dryer *et al.*, 1978; Smith *et al.*, 1981.)

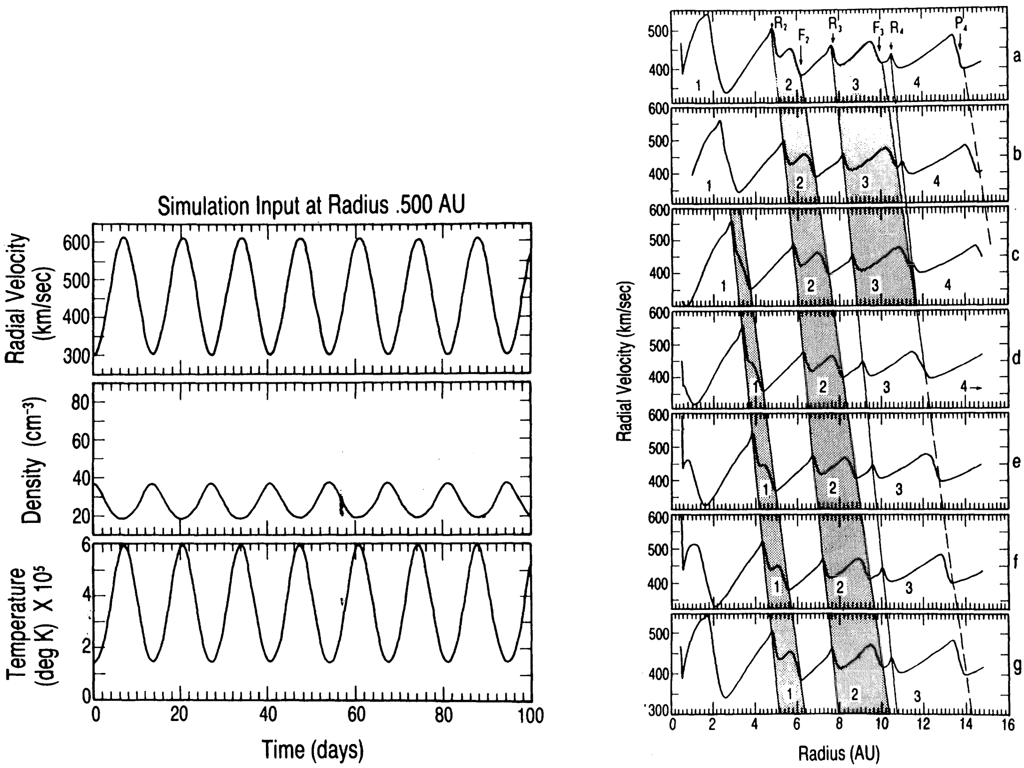


Fig. 5. The fluid nature of the solar wind can be well demonstrated by solving mathematical MHD equations. *Left*: in this example of interacting solar wind streams, a set of parameters (velocity, density, and temperature) is assumed at 0.5 AU to represent the flow from coronal holes. Two high-speed streams per solar rotation are used. The time-dependent equations are then solved to determine the evolution of these streams at the larger heliocentric distances. *Right*: the resulting corotating interaction regions predicted by solution of MHD equations are shown starting at a time greater than two solar rotations. The simulated velocity evolution is shown here at successive times during one-half of a solar rotation between 0.5 and 15 AU. Starting with the top panel (a), we first direct our attention to the steepening stream No. 1 between 1 and 2 AU. Moving to panel (b) we see that stream No. 1 steepens enroute to 3 AU. In panel (c), the shading is added to help us detect the additional steepening, due to collective fluid behaviour, into leading (forward) and following (reverse) shock waves. By the end of the half-solar rotation, the forward shock of steam No. 1 reaches 6.4 AU (panel (g)). This panel is identical to panel (a), so this now is re-numbered as F_2 and stream No. 1 in panel (g) becomes Stream No. 2 in panel (a). In the next half-solar rotation, stream No. 2 reaches 10.4 AU and is then re-numbered as stream No. 3 in panel (a), and so on, until multiple interactions take place resulting in a merged interaction region at the larger distances in this example. (From Smith *et al.*, 1985.)

et al., 1986). Upstream inhomogeneities over a wide range of scale lengths can undoubtedly be expected to affect large segments of shock surfaces (Chao, 1984).

Interplanetary simulations for corotating interaction regions and solar flares have been described in a series of general studies by Dryer *et al* (1984), Dryer and Smart (1984), Smith *et al* (1985), Smith *et al* (1986), and Gislason *et al* (1984). Several innovative studies were also incorporated by combining the $2\frac{1}{2}$D model calculations (the

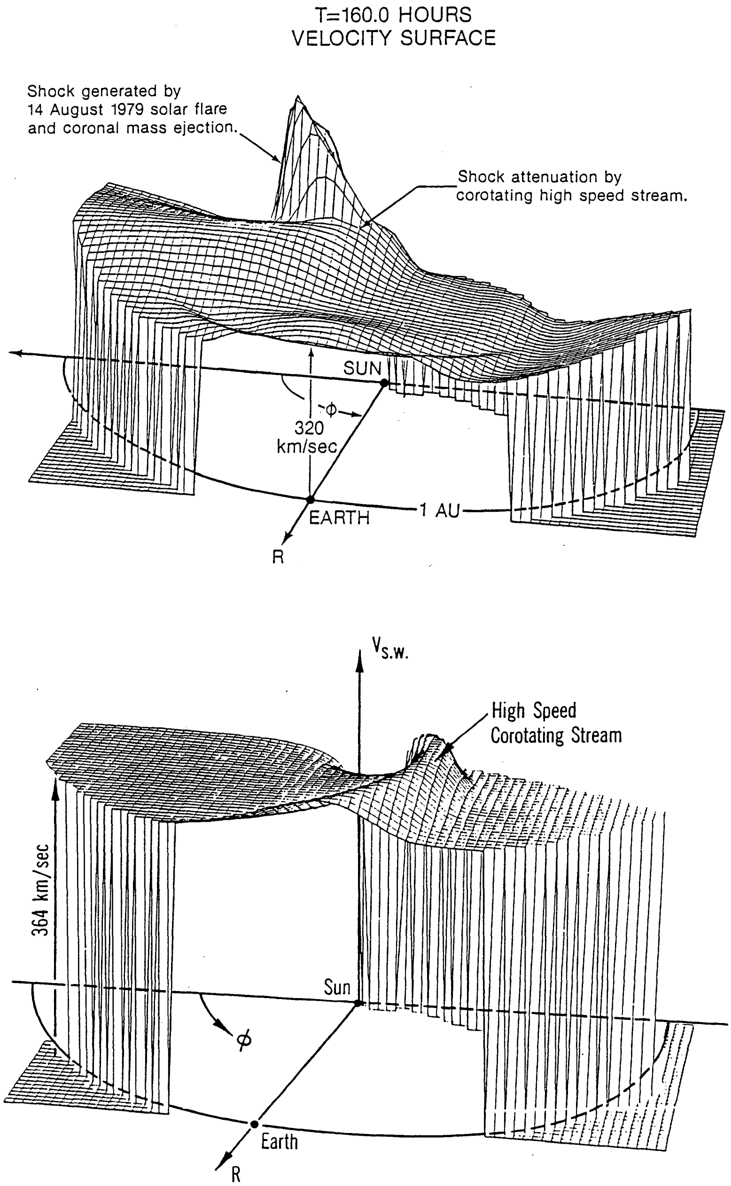


Fig. 6. The MHD simulation of the compound effects during the August 1979 series of solar and interplanetary events. *Lower panel*: a simulation of a typical high speed corotating stream that is in steady-state within the corotating frame of reference. This particular stream was assumed to emanate from a coronal hole with angular width of 35°. *Upper panel*: solar wind velocity in the ecliptic plane during the passage over the Earth of a forward, fast-mode, MHD wave (produced by an eruptive prominence four days earlier). Also shown is a forward, fast-mode MHD shock wave from an E73° flare. (Simulated date: 14 August 1979.) (Dryer and Smart, 1984; Dryer *et al.*, 1986b, c.)

$\frac{1}{2}$ refers to the fact that all three velocity and IMF components are explicitly considered in only the 2D ecliptic plane) with planetary spacecraft ionospheric diagnostics at both Earth and Venus to invoke the properties of common solar activity at different spatial locations (Taylor *et al.*, 1984, 1985).

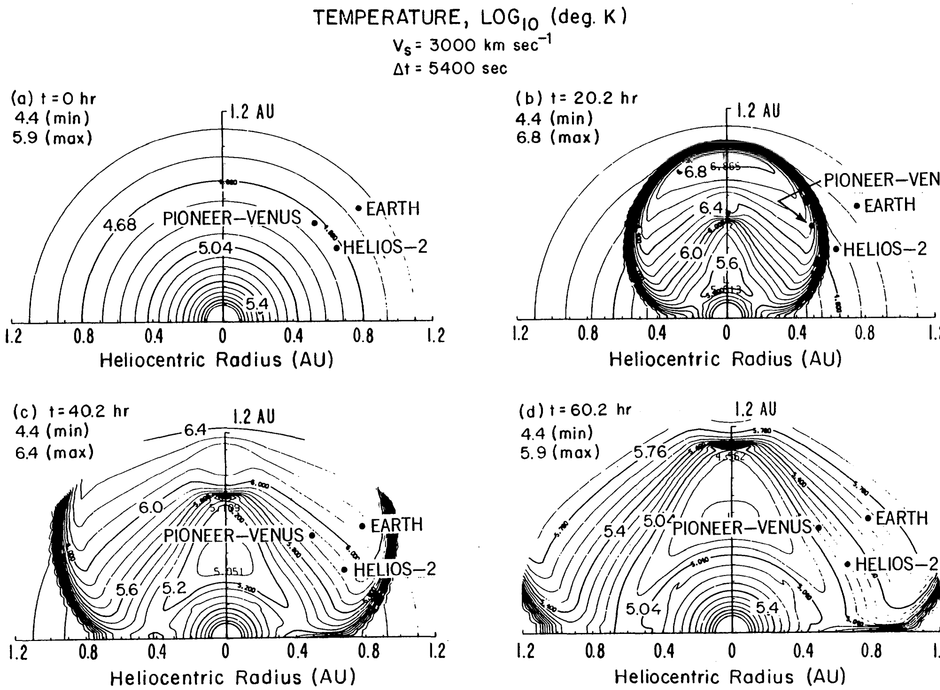


Fig. 7. MHD simulation ($2\frac{1}{2}$D) of a flare-generated shock wave from an H-alpha classified 2B flare at S 16° W 50° on 11 December, 1978, 1833 UT. Contours of solar wind temperature (log base 10) are used to show the temporal and spatial progression of the fast forward MHD shock and the fast reverse shock that is confined to a narrow heliolongitudinal band centered on the flare's meridian. Progression of the forward shock, in sequential order past Venus, Helios-2, and Earth, is clearly suggested. (H. A. Taylor, Jr., 1985, private communication.)

An example of this kind of multi-spacecraft *and* multi-planet response, together with the use of the $2\frac{1}{2}$D MHD model, is indicated in Figure 7. The reasonably close association of Earth, Pioneer–Venus–Orbiter (often referred to as Pioneer 12 in orbit around Venus), and the Helios-2 spacecraft is indicated in the half-ecliptic plane. The MHD simulation, initiated with a shock wave that was assumed to have a piston-driven velocity of 3000 km s^{-1} for 5400 s at 18 $R_{\odot}$, was oriented so that the flare's meridian is directly upward in the figure. The observed responses at the indicated locations were indeed noted: (1) the Cytherean ionopause was compressed; (2) Helios-2 observed the shock; and (3) the ionosphere at Earth also was dramatically compressed (H. A. Taylor, Jr., 1985, private communication).

Other intervals of exceptional solar and geomagnetic activity (August 1979 and February 1986) were studied by Dryer and Smith (1986, 1987), Dryer *et al* (1986a), Garcia (1986), and Garcia and Dryer (1987). These afore-mentioned studies were performed with the $2\frac{1}{2}$D model that used solar observations for input as described above. In a complementary study (Dryer *et al.*, 1986b, c), of several flares during the February 1986 epoch, a 3D model (described in more detail below) was used to demonstrate the development, propagation, and expansion of a magnetic cloud (or bubble) that was formed as a result of numerical reconnection. An example of a 3D MHD modeled 'bubble' or 'cloud' (as suggested by the observations of Burlaga *et al.*, 1981) is shown in Figure 8 (Wu *et al.*, 1987a). This preliminary result suggests that the diffusivity produced by the numerical algorithm produced an effect that is equivalent to the physics of a realistic resistivity in the solar wind near or within the heliospheric current sheet. Further studies with mathematical approximations for non-infinite electrical conductivity must be performed as part of this program in the future.

A fully three-dimensional, time-dependent, interplanetary MHD simulation code (Dryer *et al.*, 1986d; Han *et al.*, 1988a, b; and Wu *et al.*, 1987a) has also been developed recently as just noted above. This code has been used to demonstrate several important features: (1) 3D shock propagation and IMF (interplanetary magnetic field) distortion following simulated flares and CMEs from any heliospherical location out to 1.1 AU; and (2) magnetic bubble (or 'cloud') propagation to 1.1 AU through an interplanetary medium that contains a canonical flat heliospheric current sheet. These exploratory demonstrations are presently limited to an ideal plasma and reconnections produced by numerical diffusivity (as noted above) associated with opposed IMF polarities at the current sheet (as in earlier-noted magnetospheric MHD simulations). Also, the input functions are inserted at the lower boundary of 18 solar radii ($R_\odot$) which is well within the field of view of an NRL proposed coronagraph and the complementary proposed radio swept frequency experiment of the Meudon group for SOHO. Numerical simulations for the coronal interplanetary medium are presently being extended to include the following additional physical features: (1) more realistic curved heliospheric current sheet with self-consistent volumetric currents; (2) improved adaptive gridding (Panitchob, 1987) to better define high-gradient surfaces such as shocks and contact surfaces ('pistons'); (3) dissipative processes such as resistivity, thermal conduction, and viscosity; and (4) initialization of disturbances at the base of the corona rather than the present 18 $R_\odot$ surface.

This 3D MHD model will be able to examine important spatial configurations involving latitudinal effects such as those arising at the Earth's location (say) as a result of a high latitude flare, an irregularly shaped coronal hole, transient coronal hole apparitions, eruptive prominences, etc. Multi-spacecraft tests of this model have not been made as yet.

A kinematic technique in 3D, time-dependent, structure has been developed by Hakamada and Akasofu (1982) and tested (V only) in a multi-spacecraft mode in the ecliptic plane (Akasofu *et al.*, 1985). Although this method does not explicitly consider the simultaneous solution of the conservation equations, it can be calibrated (Olmsted

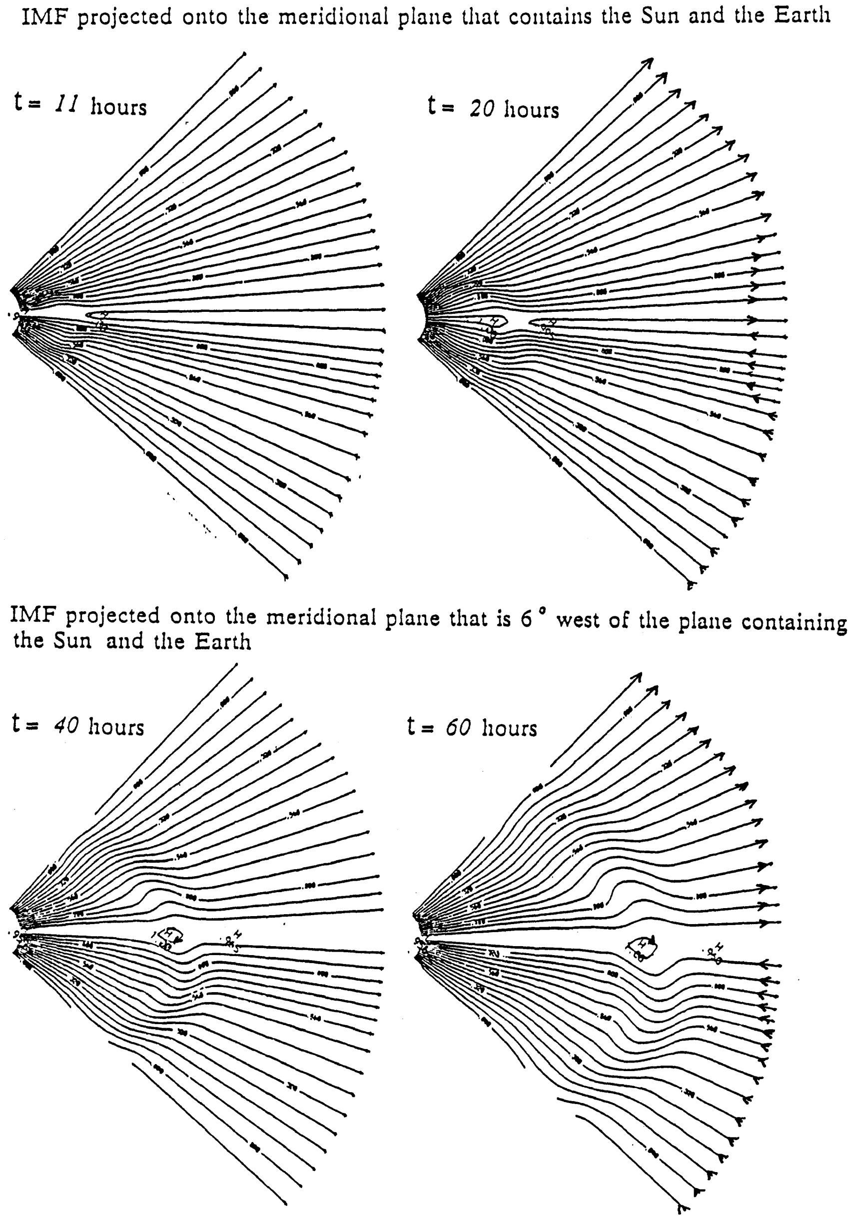


Fig. 8. Projections of the interplanetary magnetic field (IMF) onto two meridional planes as simulated by the 3D Interplanetary Global Model. The Sun is at the left side of each panel; the Earth is near the right side (1.1 AU) within the current sheet that separates outward polarity (northern hemisphere) from inward polarity (southern hemisphere). Generation of a magnetic cloud (or 'bubble') by strong pressure gradients produced by two solar flares in the southern hemisphere is indicated by this simulation. (Wu *et al.*, 1987a.)

and Akasofu, 1985) with the use of known $1\frac{1}{2}$D MHD solutions (Dryer and Steinolfson, 1976; Smith *et al.*, 1985) for the case of corotating streams. Nevertheless, as suggested earlier, it may be useful to provide insight but is otherwise scientifically limited.

3. Concluding Remarks

The solar wind and heliosphere constitute a rich resource for the study of an accessible 'laboratory' of space plasmas and MHD simulation studies. Both steady- and time-dependent observations of our closest star's neighbourhood have enriched our understanding of both laboratory and astrophysical plasmas. Numerical MHD simulations have provided a first-order understanding of the physical processes that take place in this huge 'laboratory'. The next step is a requirement to complete the modeling of the various processes required for the full chain of solar-terrestrial physics from the Sun to the Earth's atmosphere.

The ULYSSES spacecraft, together with ecliptic baseline observations by WIND, SOHO, and PHOBOS would be invaluable for providing a test bed to evaluate models such as the 3D model described herein. Remote sensing of the interplanetary (Tappin *et al.*, 1984) and coronal environment will, of course, add much to the verification of disturbances such as shocks. Furthermore, the use of soft X-ray imaging and white light imaging will be important elements for the determination of input forcing functions at the Sun. These studies will be especially important in the study of solar wind-cometary interactions where the association of solar effects (flares, sector boundaries, etc.) have often proved to be ambiguous in the past with respect to cometary tail disconnection events, tail kinking, and possibly, even cometary brightness variations.

We anticipate, therefore, that numerical simulations and further comparisons with multidisciplinary, correlative observations will be possible with the wide range images (1.1 to 30 $R_\odot$) from a proposed SOHO coronagraph, the low-frequency (metric through kilometric) radio experiment, the *in situ* SOHO and WIND plasma and magnetic field experiments, and the magnetospheric response provided by the CLUSTER experiments. This full complement of interdisciplinary observational and theoretical approaches, applied to actual solar-terrestrial events, would provide a quantum jump in our understanding of the transfer of energy, mass, and momentum from the Sun to the Earth as envisioned by the SCOSTEP program, STEP (Solar-Terrestrial Energy Programme), and the COSPAR program, IHS (International Heliosphere Study).

Acknowledgements

We wish to thank Dr H. A. Garcia for reading and commenting on the manuscript with many helpful suggestions. Partial support for this work was provided by the Air Force Geophysics Laboratory Order No. GLH7–6017 (MD), NASA Grant NAGW-9 (STW), NASA-MSFC Grant No. NAG8–053 (STW), and NOAA/SORANR60072 (STW).

References

Akasofu, S.-I., Fillius, W., Sun, W., Fry, C., and Dryer, M.: 1985, *J. Geophys. Res.* **90**, 8193.

Brackbill, J. V. and Pracht, W. E.: 1973, *J. Comp. Phys.* **13**, 455.

Brecht, S. H., Lyon, J. G., Fedder, J. A., and Hain, K.: 1982, *J. Geophys. Res.* **87**, 6098.

Burlaga, L., Pizzo, V., Lazarus, A., and Gazis, P.: 1985, *J. Geophys. Res.* **90**, 7377, 1985.

Burlaga, L., Sittler, E., Mariani, F., and Schwenn, R.: 1981, *J. Geophys. Res.* **86**, 6673.

Chao, J.-K.: 1984, *Adv. Space Res.* **4**, 327.

Cuperman, S. and Dryer, M.: 1985, *Phys. Rev.* **A31**, 3470.

Dryer, M.: 1982, *Space Sci. Rev.* **33**, 233.

Dryer, M. and Maxwell, A.: 1979, *Astrophys. J.* **231**, 945.

Dryer, M. and Shea, M. A.: 1986, *Adv. Space Res.* **6**(6), 343.

Dryer, M. and Smart, D. F.: 1984, *Adv. Space Res.* **4**, 291.

Dryer, M. and Smith, Z. K.: 1986, 'MHD Simulation of Multiple Interplanetary Disturbances During STIP Interval VII (August 1979)', in V. E. Stepanov and V. N. Obridko (eds.), *Solar Maximum Analysis*, Irkutsk, USSR, June 17–22, 1985, VNU Science Press, Utrecht, pp. 369–380.

Dryer, M., and Smith, Z. K.: 1987, 'Multi-Spacecraft Testing of Time Dependent Interplanetary MHD Models for Operational Forecasting of Geomagnetic Storms', in *Proceedings of the Workshop on the Multi-Comet Mission*, 17–18 September, 1986, Greenbelt, Maryland, NASA Goddard Space Flight Laboratory (in press).

Dryer, M., and Steinolfson, R. S.: 1976, *J. Geophys. Res.* **81**, 5413.

Dryer, M., and Wu, S. T.: 1985, *J. Geophys. Res.* **90**, 559.

Dryer, M., Smith, Z. K., Wu, S. T., Han, S. M., and Yeh, T.: 1986a, 'MHD Simulation of the "Geo-effectiveness" of Interplanetary Disturbances', in Y. Kamide and J. A. Slavin (eds.), *Solar Wind-Magnetosphere Coupling*, Terra Scientific Publ. Co., Tokyo, pp. 191–207.

Dryer, M., Smith, Z. K., Detman, T. R., and Yeh, T.: 1986b, *MHD Simulation of the Interplanetary Environment in the Ecliptic Plane During the 3–9 February 1986 Solar and Geomagnetic Activity*, Air Force Geophysics Laboratory, AFGL-TR-86-0189.

Dryer, M., Smith, Z. K., Garcia, H. A., and Han, S. M.: 1986c, *EOS* **67**(44), 917.

Dryer, M., Wu, S. T., and Han, S. M.: 1986d, 'Three-Dimensional, Time-Dependent, MHD Model of a Solar Flare-Generated Interplanetary Shock Wave', in R. G. Marsden (ed.), *The Sun and the Heliosphere in Three Dimensions*, D. Reidel Publ. Co., Dordrecht, Holland, pp. 135–140.

Dryer, M., Smith, Z. K., Smith, E. J., Mihalov, J. D., Wolfe, J. H., Steinolfson, R. S., and Wu, S. T.: 1978, *J. Geophys. Res.* **83**, 4347.

Dryer, M., Wu, S. T., Gislason, G., Han, S. M., Smith, Z. K., Wang, J. F., Smart, D. F., and Shea, M. A.: 1984, *Astrophys. Space Sci* **105**, 187.

Dryer, M., Wu, S. T., and Han, S. M.: 1980, *Geofis. Intern.* **19**, 1.

Dryer, M., Wu, S. T., Steinolfson, R. S., and Wilson, R. M.: 1979, *Astrophys. J.* **227**, 1059.

D'Uston, C.: 1981, 'Dynamique des Perturbations Interplanetaires engendrées par les Eruptions Solaires', Ph.D. Thesis, Université Paul Sabatier, Toulouse, France.

D'Uston, C., Dryer, M., Han, S. M., and Wu, S. T., 1981, *J. Geophys. Res.* **86**, 525.

Forbes, T. G. and Priest, E. R.: 1983, *J. Geophys. Res.* **88**, 863.

Garcia, H. A.: 1986, *Bull. Am. Astron. Soc.* **18**, 699.

Garcia, H. A. and Dryer, M.: 1987, *Solar Phys.* **109**, 119.

Gislason, G., Dryer, M., Smith, Z. K., Wu, S. T., and Han, S. M.: 1984, *Astrophys. Space Sci.* **98**, 149.

Gold, T.: 1959, *J. Geophys. Res.* **64**, 1665.

Gosling, J. T., Hundhausen, A. J., and Bame, S. J.: 1976, *J. Geophys. Res.* **81**, 2111.

Hakamada, K. and Akasofu, S. I.: 1982, *Space Sci. Rev.* **31**, 3.

Han, S. M., Wu, S. T., and Dryer, M.: 1988a, *Comp. Fluids* **16**, 81.

Han, S. M., Wu, S. T., and Dryer, M.: 1988b, 'A Transient, Three-Dimensional MHD Model for Numerical Simulation of Interplanetary Disturbances', in M. A. Shea and D. F. Smart (eds.), *STIP Symposium on Retrospective Analyses*, Book Crafters Publ. Co., Chelsea, Michigan (in press).

Han, S. M., Wu, S. T., and Nakagawa, Y.: 1982, *Comp. Fluids* **10**, 159.

Harrison, R. A., Waggett, P. W., Bentley, R. D., Phillips, K. J. H., Bruner, M., Dryer, M., and Simnett, G. M.: 1985, *Solar Phys.* **97**, 387.

Hu, Y. Q. and Wu, S. T.: 1984, *J. Comp. Phys.* **55**, 33.

Hundhausen, A. J. and Gosling, J. T.: 1976, *J. Geophys. Res.* **81**, 1436.
Ivanov, K. G., Mikerina, N. V., and Evdokimova, L. V.: 1980, 'Classification and Investigation of Solar Flare Situations Conformably to Interplanetary and Magnetospheric Disturbances', in M. Dryer and E. Tandberg-Hanssen (eds.), *Solar and Interplanetary Dynamics*, D. Reidel Publ. Co., Dordrecht, Holland, pp. 421–424.
Jastrow, R.: 1959, *J. Geophys. Res.* **64**, 1647.
Low, B. C.: 1981, *Astrophys. J.* **251**, 352.
Nakagawa, Y., Wu, S. T., and Han, S. M.: 1978, *Astrophys. J.* **219**, 314.
Nakagawa, Y., Wu, S. T., and Han, S. M.: 1981, *Astrophys. J.* **244**, 331.
Olmsted, C. S. and Akasofu, S. I.: 1985, *Planetary Space Sci.* **33**, 831.
Panitchob, S.: 1987, 'An Adaptive Grid Finite Difference Method for Time-Dependent Magnetohydrodynamics Flows and its Astrogeophysical Applications', Ph.D. Thesis, University of Alabama in Huntsville, U.S.A.
Parker, E. N.: 1958, *Astrophys. J.* **224**, 631.
Pizzo, V. J.: 1982, *J. Geophys. Res.* **87**, 4374.
Pizzo, V. J.: 1983, 'Quasi-Steady Solar Wind Dynamics, in M. Neugebauer (ed.), *Solar Wind Five*, NASA Conf. Publ. CP 2280, pp. 675–691.
Pneuman, G. W. and Kopp, R. A.: 1971, *Solar Phys.* **18**, 258.
Pudovkin, M. I. and Chertkov, A. D.: 1976, *Solar Phys.* **50**, 213.
Rubin, E. L. and Burstein, S. Z.: 1967, *J. Comp. Phys.* **2**, 178.
Shea, M. A. and Dryer, M.: 1988, 'Overview of STIP Intervals I–XIV', in M. A. Shea and D. F. Smart (eds.), *STIP Symposium on Retrospective Analyses*, 10–12 June, 1985, Les Diablerets, Switzerland, BookCrafters Publ. Co. Chelsea, Michigan, U.S.A. (in press).
Sime, D. G., MacQueen, R. M., and Hundhausen, A. J.: 1984, *J. Geophys. Res.* **89**, 2113.
Smith, E. J. and Wolfe, J. H.: 1976, *Geophys. Res. Letters* **3**, 137.
Smith, E. J. and Wolfe, J. H.: 1979, *Space Sci. Rev.* **23**, 217.
Smith, Z. K., Dryer, M., Fillius, R. W., Smith, E. J., and Wolfe, J. H.: 1981, *J. Geophys. Res.* **86**, 6773.
Smith, Z. K., Dryer, M., and Han, S. M.: 1986, *Astrophys. Space Sci* **119**, 337.
Smith, Z. K., Dryer, M., and Steinolfson, R. S.: 1985, *J. Geophys. Res.* **90**, 217.
Steinolfson, R. S., Wu, S. T., Dryer, M., and Tandberg-Hanssen, E.: 1978, *Astrophys. J.* **225**, 259.
Tang, F., Akasofu, S.-I., Smith, E., and Tsurutani, B.: 1985, *J. Geophys. Res.* **90**, 2703.
Tappin, S. J., Hewish, A., and Gapper, G. R.: 1984, *Planetary Space Sci.* **32**, 1273.
Taylor, H. A., Jr., Cloutier, P. A., Dryer, M., Suess, S. T., Barnes, A., Wolff, R. S., and Stern, A. A.: 1984, *Adv. Space Res.* **4**, No. 7, 343.
Taylor, H. A., Jr., Cloutier, P. A., Dryer, M., Suess, S. T., Barnes, A., and Wolff, R. S.: 1985, *Earth, Moon, and Planets* **32**, 275.
Whang, Y. C.: 1984, *J. Geophys. Res.* **89**, 7367.
Whang, Y. C. and Burlaga, L. F.: 1985a, *J. Geophys. Res.* **90**, 10765.
Whang, Y. C. and Burlaga, L. F.: 1985b, *J. Geophys. Res.* **90**, 221.
Whang, Y. C. and Burlaga, L. F.: 1986, 'The Coalescence of Two Merged Interaction Regions Between 6.2 and 9.5 AU: September, 1979, Event', NASA/Goddard SFC preprint, June, 1986.
Wu, S. T.: 1985, 'Computer-Simulated Active Solar Phenomena', in C. de Jager and Chen Bio (eds.), *Proceedings of Kunming Workshop on Solar Physics and Interplanetary Travelling Phenomena*, Science Press, Beijing, Volume 1, pp. 56–72.
Wu, S. T. and Wang, J. F.: 1987, 'Numerical Tests of a Modified Full Implicit Continuous Eulerian (FICE) Scheme with Projected Normal Characteristic Boundary Conditions for MHD Flows', *J. Computer Methods in Applied Mechanics and Engineering*, Elsevier Science Publishers B.V., North-Holland, **64**(1–3), 267 - 282.
Wu, S. T., Bao, J. J., and Wang, J. F.: 1986, *Adv. Space Res.* **6**(6), 53.
Wu, S. T., Han, S. M., and Dryer, M.: 1979, *Planetary Space Sci.* **27**, 255.
Wu, S. T., Dryer, M., and Han, S. M.: 1983, *Solar Phys.* **84**, 395.
Wu, S. T., Dryer, M., and Han, S. M.: 1987a, 'Three-Dimensional, Time-Dependent MHD Simulation of Traveling Interplanetary Phenomena with Solar Origin', *Proceedings of the Third International School of Space Simulation*, 22–27 June 1987, Beaulieu, France (in press).
Wu, S. T., Song, M. T., and Dryer, M.: 1987b, *Astrophys. Space Sci.* (in press).
Wu, S. T., Dryer, M., Steinolfson, R. S., and Tandberg-Hanssen, E.: 1987c, *Astrophys. J.* (submitted).

Wu, S. T., Dryer, M., Nakagawa, Y., and Han, S. M.: 1978, *Astrophys. J.* **219**, 324.
Wu, S. T., Nakagawa, Y., Han, S. M., and Dryer, M.: 1982a, *Astrophys. J.* **262**, 369.
Wu, S. T., Hu, Y. Q., Wang, S., Dryer, M., and Tandberg-Hanssen, E.: 1982b, *Astrophys. Space Sci.* **83**, 189.
Wu, S. T., Wang, S., Dryer, M., Poland, A. I., Sime, D. G., Wolfson, C. J., Orwig, L. E., and Maxwell, A.: 1983a, *Solar Phys.* **85**, 351.
Wu, S. T., Hu, Y. Q., Nakagawa, Y., and Tandberg-Hanssen, E.: 1983b, *Astrophys. J.* **266**, 866.
Yeh, T.: 1985, *Solar Phys.* **95**, 83.
Yeh, T. and Dryer, M.: 1985, *Astrophys. Space Sci.* **117**, 165.
Yeh, T. and Pneuman, G. W.: 1977, *Solar Phys.* **54**, 419.

ALFVÉN WAVES IN A SPACE PLASMA AND ITS ROLE IN THE SOLAR WIND INTERACTION WITH COMETS*

A. A. GALEEV and R. Z. SAGDEEV

Space Research Institute of USSR Academy of Sciences, Moscow, U.S.S.R.

(Received 3 September, 1987)

Abstract. The importance of Alfvén wave generation in interacting plasmas is discussed in general and illustrated by the example of solar wind interaction with cometary plasma. The quasi-linear theory of Alfvén wave generation by cometary ions at distances far from the cometary nucleus is reviewed. The incorporation of a diabatic plasma compression effects into this theory modifies the spectrum of Alfvén waves and the integral intensity of magnetic field fluctuations previously published. These results are in quantitative agreement with the *in situ* observations near the comets Giacobini–Zinner and Halley. However, the polarization of quasi-linearly excited waves needs further detailed comparison with observations.

1. Introduction

Alfvén waves (Alfvén, 1942) are one of three branches of low-frequency magnetohydrodynamic waves in a uniform magnetized plasma. However, their role in the energetics and dynamics of a space plasma happens to be very specific. First of all, this is due to the fact that the binary collisions in a rarefied space plasma are so rare that the damping of magnetohydrodynamic waves comes mainly from the resonant interaction of waves with plasma particles. For the typical ratios of the space plasma pressure to the magnetic field pressure $\beta \sim 1$ there are a large number of ions in a plasma with the velocities of the order of the phase velocities of magnetohydrodynamic waves. Thus the slow and fast magnetosonic waves experience strong Landau damping due to Cherenkov reasonance with these ions. Alfvén waves propagating exactly along the magnetic field lines interact with plasma particles only via the cyclotron resonance and the damping of low-frequency Alfvén waves is very weak, since the number of resonance particles is exponentially small. As a result, the Alfvén waves could be excited in a space plasma more easily than the magnetosonic waves. Moreover, due to a weak damping these waves propagate to large distances from the region of their generation and thus could provide the energy transport from one region to another.

Precisely because of the properties described above the Alfvén waves play a key role in many phenomena in a space plasma. The discovery of large amplitude Alfvén waves propagating from the solar corona to the solar wind (Belcher and Davis, 1971) has stimulated the idea that these waves play an important role in the energetics of the solar wind. In particular, a number of specialists (see Hollweg, 1985; and references given therein) hold that the ponderomotive force produced by these waves provides additional acceleration to the solar wind while the dissipation of the waves results in additional plasma heating; these are necessary for an agreement of the two fluid model of the

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

thermal expansion of solar corona (Hurtle and Sturrock, 1968) with the observations of solar wind.

Alfvén waves play the principal role in the Fermi acceleration of charged particles leaking easily forward from behind the shock front and having an anisotropic velocity distribution (Bell, 1978). The waves excited upstream of the shock front are then convected downstream with the plasma flow. Since Alfvén waves are convected by upstream plasma faster than downstream, counterstreaming motion of upstream and downstream waves results. The charged particles reflected from these counterstreaming waves are accelerated in a manner described by the Fermi acceleration of the first kind. Since the mechanism of injection of particles into the acceleration is not yet specified the self-consistent theoretical description of the Fermi acceleration by shocks is still absent. Therefore, the energy spectrum of accelerated particles and the frequency spectrum of excited Alfvén waves are calculated in the test particle approximation where the quasi-linear theory provides an adequate plasma description (see comprehensive review by Forman and Webb, 1985).

The role of Alfvén waves in the solar wind interaction with cometary atmospheres is even more important. Here the beam of ions created by the ionization of atoms and molecules of cometary gas in the flow of solar wind plasma excites Alfvén waves in the large area upstream of the comet (Smith *et al.*, 1986; Sagdeev *et al.*, 1986). The rate of cometary gas ionization is much lower than the growth rate of the beam instability and of the relaxation rate of the beam. Thus the solar wind together with the beam of cometary ions relaxes quickly to the stable state. Moreover, far from the comet the energy density of the excited Alfvén waves happens to be much lower than the densities of magnetic field energy and thermal plasma energy. The latter permits the use of the quasi-linear description of a plasma to solve the problem of solar wind loading by cometary ions (Sagdeev *et al.*, 1986). In contrast to the problem of particle acceleration by shocks here the density and the velocity distribution of the cometary ions exciting Alfvén waves are easily found from the known spatial distribution of cometary gas. In view of the fact that the quasi-linear theory of the solar wind loading by cometary ions is the most complete one in comparison with the other two important problems mentioned above, we demonstrate the abilities of a quasi-linear theory of plasma for this particular example.

2. The Theory of Alfvén Wave Generation by Cometary Ions

For the sake of illustration we restrict ourselves to the simplest case of Alfvén wave generation in a solar wind interacting with a cometary gas at far distances from the comet where the amplitudes of Alfvén waves are so low that one can use the quasi-linear theory of plasma. The geometry under consideration is shown in Figure 1. Here we assume that the solar wind flows with velocity $u(x)$ in a cometary system of coordinate along the x-axis directed towards the nucleus of a comet. The vector $\mathbf{B}$ of the interplanetary magnetic field forms the angle α with the direction of plasma flow. The cometary gas evaporated from the cometary nucleus by the solar radiation expands into

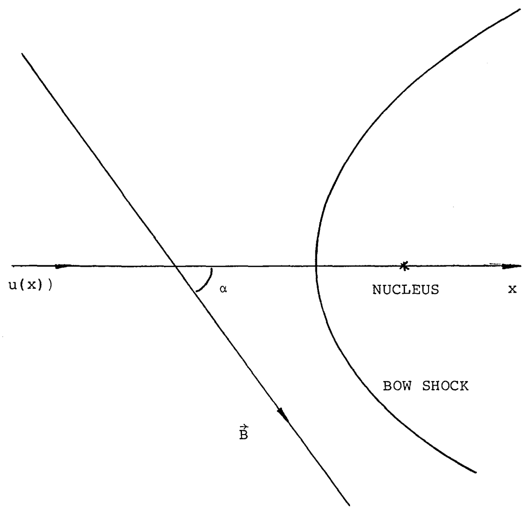


Fig. 1. The geometry of solar wind intraction with comets. Here α is the angle between the velocity vector **u** of solar wind and the magnetic field vector **B**.

the interplanetary space with a velocity of the order of ~ 1 km s^{-1} that is much lower than the solar wind velocity. Therefore, the initial velocity of a cometary ion born in a solar wind flow can be considered equal to zero. Under the action of the self-consistent electric field $\mathbf{E} = -[\mathbf{u} \times \mathbf{B}]c^{-1}$ this ion drifts across the interplanetary magnetic field with the plasma at the velocity $u \sin \alpha$ and rotates along the cyclotron orbit with the same velocity (Landau and Lifshitz, 1960). However, the ion velocity along the magnetic field is equal to zero in the cometary system of coordinates. Thus in the system of coordinate moving with the solar wind this ion moves with the velocity $v_{\parallel} = -u \cos \alpha$ along the magnetic field and performs the cyclotron rotation with the velocity $v_{\perp} = u \sin \alpha$. As a result a beam of cometary ions is formed in a plasma moving along spirals with the velocity equal to the local plasma velocity $u(x)$ relative the comet.

Before performing accurate calculations of Alfvén wave generation by such beam of ions it is useful to present energy conservation arguments that permit a qualitative picture of the development of beam instability to be drawn.

2.1. The Energy Conservation Arguments

As is known the reaction of excited waves on the velocity distribution of the plasma particles can be reduced to a diffusion equation for the resonant particles in velocity space (Sagdeev and Galeev, 1969). In the case of Alfvén waves such diffusion can be described without using the quasi-linear equation for the ambient particle distribution function. Here we concentrate on the simplest case when only the waves propagating exactly along the magnetic field lines are excited. This case can be justified by the fact that the growth rate reaches its maximum for the waves propagating along the magnetic field lines (Winske *et al.*, 1985). In Alfvén waves the ratio of the wave electric field to the wave magnetic field is equal to the ratio of the phase velocity of wave to the speed of light; thus this ratio is the same for waves of different frequencies if only these waves propagates in the same direction. Therefore, in the system of coordinates moving with these waves the wave electric field is zero and as a result the energy of particles interacting with waves is conserved.

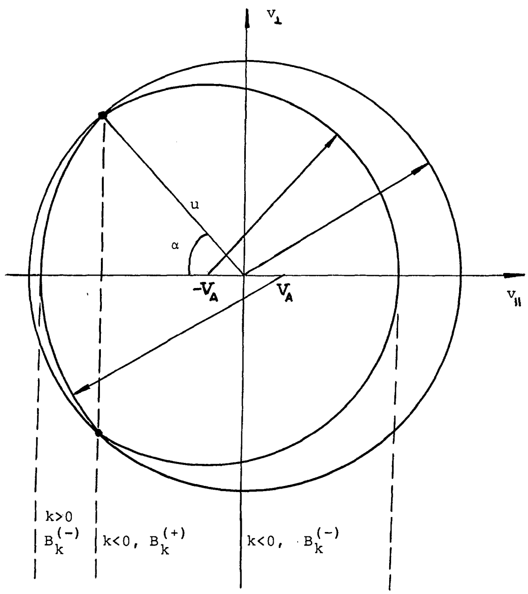


Fig. 2. The cross-section of two hemispheres s_1 and s_2 of quasi-linear diffusion of cometary ions in a velocity space under the action of Alfvén waves excited by the cometary ions born in a solar wind flow.

Figure 2 shows graphically how the velocity of cometary ions changes under the action of waves. Initially the ions produced by the ionization of the cometary gas form in velocity space a ring with radius $u \sin\alpha$ in the plane perpendicular to the $v_\parallel$ axis and shifted along this axis by the value $v_\parallel = -u\cos\alpha$ relative the origin of coordinate system. According to the afore-mentioned properties of Alfvén waves these ions interacting with waves propagating in the positive direction (in the direction of the vector **B**) diffuse from the initial ring along the surface of a sphere with the center on the $v_\parallel$ axis at the point $v_\parallel = V_A$ (V_A is the Alfvén velocity). The ions interacting with the waves propagating in the opposite direction diffuse along the sphere with the center at the point $v_\parallel = -V_A$. Let us now draw the sphere of constant energy $v_\parallel^2 + v_\perp^2 = u^2$ through the initial ring of cometary ions in the velocity space (Figure 2). Then we easily see that the diffusion of cometary ions into the velocity region $v_\parallel > -u\cos\alpha$ results in a decrease of ion energy if these ions interact with the waves propagating in the negative direction relative to the **B**-field vector. The particle energy released in the process of diffusion ($\sim m_i u V_A$) is transferred to the excited Alfvén waves. Similarly when particles diffuse into the region of negative velocities $v_{\parallel} < -u\cos\alpha$ the waves propagating in the opposite (positive) direction are excited. The meridional cross-section of the spherical surfaces along which the injected cometary ions diffuse in the velocity space is shown in Figure 2. The value of the wave vector and the sign of polarization of waves generated by the ions with the velocity $v_{\parallel}$ are found from the condition of cyclotron resonace between waves and particles

$$\omega_k - kv_{\parallel} \pm \omega_{ci} = 0\,, \tag{1}$$

where $\omega_{ci} = e_i B/m_i c$ is the cyclotron-frequency of cometary ions with charge e_i and mass m_i, and the wave frequency ω_k is considered to be positive. The signs '+' and '−' refer to the right and left polarization of the waves, respectively (the magnetic field vectors $B_k^{(+)}$ and $B_k^{(-)}$ rotate in the directions of cyclotron motion of electrons and ions, respectively). The polarization of excited waves is also indicated in Figure 2.

Thus the solar wind plasma together with the cometary ions born in it relaxes to the stable state via excitation of Alfvén waves followed by the scattering of the beam of cometary ions by the excited waves. The fraction of the energy of cometary ions transferred to excited Alfvén waves can be easily found assuming that as a result of scattering by waves the cometary ions injected on the ring in the velocity space are spread uniformly over the two hemispherical surfaces intersecting each other at this ring. The relative fraction of energy transferred to waves is given by the expression

$$\begin{aligned} \varepsilon(\alpha) &= 1 - \left(\int v^2\,\mathrm{d}s_1 + \int v^2\,\mathrm{d}s_2\right)\Big/ u^2\left(\int \mathrm{d}s_1 + \int \mathrm{d}s_2\right) \approx \\ &\approx (V_{\mathrm{A}}/u)\,(1+\cos^2\alpha)\,, \end{aligned} \tag{2}$$

where $v^2 = v_{\parallel}^2 + v_{\perp}^2$ and we neglected higher-order terms in the small parameter $V_{\mathrm{A}}/u \ll 1$ while integrating over the hemispheres s_1 and s_2. In the next sections we obtain the more general result taking into account the injection of cometary ions into the solar wind not only due to the local ionization of cometary gas, but also due to their diffusion along magnetic field lines from the downstream regions taking into account the adiabetic heating of ions.

2.2. The Analysis of Quasi-Linear Equations

As we have shown in Section 2.1, the velocity distribution $f_i(\mathbf{v})$ of cometary ions has to be calculated taking into account the resonant interaction of ions with the excited Alfvén waves. The corresponding equation for this distribution function has a form

$$(u + v_{\parallel}\cos\alpha)\frac{\partial f_i}{\partial x} + \frac{v_{\perp}}{2B}\frac{\mathrm{d}B}{\mathrm{d}x}\left[u\frac{\partial}{\partial v_{\perp}} - \cos\alpha\left(v_{\perp}\frac{\partial}{\partial v_{\parallel}} - v_{\parallel}\frac{\partial}{\partial v_{\perp}}\right)\right]f_i =$$

$$= \frac{\pi e_i^2}{m_i^2}\int\frac{\mathrm{d}k}{2\pi}\sum_{\pm}\left[\left(1 - \frac{kv_{\parallel}}{\omega_k}\right)\frac{1}{v_{\perp}}\frac{\partial}{\partial v_{\perp}}v_{\perp} + \frac{kv_{\perp}}{\omega_k}\frac{\partial}{\partial v_{\parallel}}\right]|B_k^{\pm}|^2 \times$$

$$\times \left(\frac{\omega_k}{kc}\right)^2 \delta(\omega_k - kv_{\parallel} \pm \omega_{ci}) \left[\left(1 - \frac{kv_{\parallel}}{\omega_k}\right)\frac{\partial}{\partial v_{\perp}} + \frac{kv_{\perp}}{\omega_k}\frac{\partial}{\partial v_{\parallel}}\right] f_i +$$

$$+ (\dot{N}/\pi)\,\delta(v_{\perp}^2 - u^2 \sin^2 \alpha)\,\delta(v_{\parallel} + u\cos\alpha)\,. \tag{3}$$

The first term on the left-hand side describes the hydrodynamic convection of cometary ions by plasma and its free motion along magnetic field lines. The second accounts for the adiabatic heating and cooling of cometary ions while they move from the region of weak magnetic field into the region of stronger field and *vice versa* (Luhmann, 1976). The collision term on the righe-hand side of the equation describing the quasilinear diffusion of ions in the velocity space is taken from the book of Sagdeev and Galeev (1969) and is written in the coordinate system of solar wind plasma. Here as in the original paper by Sagdeev *et al.* (1986) the spectrum of Alfvén waves is assumed to be one-dimensional in accordance with the fact that the growth rate of the instability reaches its maximum for Alfvén waves propagating along the magnetic field lines. The last term in Equation (3) describes the injection of cometary ions into the solar wind with the rate $\dot{N}$ defined by the gas production by a comet Q, the velocity V_g of spherical expansion of a gas and the time of gas ionization τ

$$\dot{N} = Q/4\pi v_g r^2 \tau\,, \tag{4}$$

where r is the distance to the nucleus of a comet ($r = -x$ according to the Figure 1).

As follows from the preceding section , for any longitudinal velocity of cometary ions falling approximately in the range $[-u, +u]$ one finds a resonant Alfvén wave; as a result of velocity space diffusion the cometary ions are uniformly distributed over the surfaces of two hemispheres shifted relative to the origin of the system of coordinates by the value of the phase velocity of the Alfvén waves. In a super-Alfvénic solar wind upstream of the comet ($u/V_A \sim 10$ at the 1 AU distance from the Sun) these hemispheres coincide with the surface of constant energy with the accuracy of $(V_A/u) \sim 0.1$. Therefore, in the first approximation the velocity distribution of cometary ions can be considered as isotropic in velocity space. However, we cannot neglect the small anisotropic correction to this distribution that is always present because of the processes of cometary ion production in the solar wind flow, free flow of ions along the magnetic field lines and the adiabatic heating (cooling) of these ions. We calculate this small correction from Equation (3) using perturbation theory. For the sake of convenience of calculations we rewrite Equation (3) in the spherical system of coordinates in the velocity space with the polar axis directed along the magnetic field

$$(u + v\mu\cos\alpha)\frac{\partial f_i}{\partial x} - \frac{v}{2B}\frac{\mathrm{d}B}{\mathrm{d}x}\cos\alpha(1-\mu^2)\frac{\partial f_i}{\partial \mu} + \frac{u}{2B}\frac{\mathrm{d}B}{\mathrm{d}x}(1-\mu^2)\times$$

$$\times\left(v\frac{\partial}{\partial v} - \mu\frac{\partial}{\partial \mu}\right) f_i = \mathrm{St}_{QL}(f_i) + (\dot{N}/2\pi v^2)\,\delta(v-u)\,\delta(\mu + \cos\alpha)\,, \tag{5}$$

where

$$\mathrm{St}_{QL}(f_i) = \frac{\pi e_i^2}{m_i^2} \int \frac{\mathrm{d}k}{2\pi} \left[\frac{1}{v^2}\frac{\partial}{\partial v} v^2 \left(\frac{\omega_k}{kv}\right) + \frac{1}{v}\frac{\partial}{\partial \mu}\left(1 - \frac{\omega_k \mu}{kv}\right)\right] \times$$
$$\times \, |B_k^\pm|^2 (1-\mu^2)(v/c)^2 \, \delta(\omega_k - kv\mu \pm \omega_{ci}) \times$$
$$\times \left[\left(1 - \frac{\omega_k \mu}{kv}\right)\frac{\partial}{v\,\partial\mu} + \left(\frac{\omega_k}{kv}\right)\frac{\partial}{\partial v}\right] f_i ;$$

where $\mu = \cos\theta$; θ being the pitch-angle of the cometary ion.

Let us represent now the distribution function of cometary ions in the form

$$f_i(x, \mathbf{v}) = f_0(x, v) + f_1(x, v, \mu) . \tag{6}$$

In accordance with the results of the preceding section the correction f_1 to the isotropic distribution function f_0 is small as $(V_A/u) \sim 0.1$. Therefore, while calculating this correction the collision term has to be expanded on this small parameter and only the terms to the same order have to be left: i.e.,

$$\mathrm{St}_{QL}(f_i) \simeq \frac{\partial}{\partial \mu}(1-\mu^2)\, \nu_{\mu\mu} \frac{\partial f_1}{\partial \mu} + \frac{\partial}{\partial \mu}(1-\mu^2)\, \nu_{\mu v} v \frac{\partial f_o}{\partial v} , \tag{7}$$

where

$$\nu_{\mu\mu} = \frac{\pi e_i^2}{m_i^2 c^2} \int \frac{\mathrm{d}k}{2\pi} \sum_{\pm} |B_k^\pm|^2 \, \delta(\omega_k - kv\mu \pm \omega_{ci}) \left(1 - \frac{\omega_k \mu}{kv}\right)^2$$

$$\nu_{\mu v} = \frac{\pi e_i^2}{m_i^2 c^2} \int \frac{\mathrm{d}k}{2\pi} \sum_{\pm} |B_k^\pm|^2 \, \delta(\omega_k - kv_\mu \pm \omega_{ci}) \left(\frac{\omega_k}{kv}\right)\left(1 - \frac{\omega_k \mu}{kv}\right) . \tag{8}$$

Taking into account only the variable in μ terms on the left hand side of Equation (5), i.e., the terms with the average values equal to zero, and neglecting the terms of the order of $(f_1/f_0) \ll 1$ we obtain the expression for the derivative of the distribution function of cometary ions on the pitch-angle that enters the expression for the growth rate of Alfvén waves

$$\nu_{\mu\mu} \frac{\partial f_1}{\partial \mu} = - \frac{\dot{N}}{2\pi v^2} \frac{\delta(v-u)}{(1-\mu^2)} [\eta(\mu + \cos\alpha) - (1+\mu)/2] -$$
$$- (v/2)\cos\alpha \frac{\partial f_0}{\partial x} + (\mu u v/6) \frac{\mathrm{d}\ln B}{\mathrm{d}x} \frac{\partial f_0}{\partial v} - \nu_{\mu v} v \frac{\partial f_0}{\partial v} , \tag{9}$$

where $\eta(\mu) = 1$ for $\mu \geq 0$ and $\eta(\mu) = 0$ for $\mu < 0$. While integrating Equation (5) over μ we have chosen the integration constant in such a way that the function $\partial f_1/\partial \mu$ has no singularities at $\mu = \pm 1$. The anisotropy of cometary ions is related to the following effects: the velocity anisotropy of produced ions (the first term on the right-hand side of Equation (9)), the free motion of ions along the magnetic field lines (the second term), the adiabatic change of the ion energy while ions move along field lines from the plasma regions with the different magnetic field strength (the third term) and finally the deviation of surfaces along which the quasi-linear diffusion of ions takes place from the surface of constant energy (the last term).

The isotropic part of the distribution function of cometary ions entering the above written expressions satisfies the equation

$$u\frac{\partial f_0}{\partial x} + (uv/3)\frac{\mathrm{d}\ln B}{\mathrm{d}x}\frac{\partial f_0}{\partial v} = \frac{\dot{N}}{4\pi v^2}\,\delta(v-u)\,. \tag{10}$$

This equation is obtained by the averaging of Equation (5) and neglecting the terms of higher order.

With the help of the expression (9) for the anisotropic part of the distribution function one can calculate the growth rate of Alfvén waves excited by cometary ions. For the sake of simplicity we use the energy conservation law in the system 'cometary ions + Alfvén waves':

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int (m_i v^2/2) f_i\,\mathrm{d}\mathbf{v} + \sum_{\pm}\int \frac{\mathrm{d}k}{2\pi}\frac{|B_k^{\pm}|^2}{4\pi}\right] - \dot{N}m_i u^2/2 \equiv$$

$$\equiv 2\pi\int_0^{\infty} v^2\,\mathrm{d}v \int_{-1}^{+1}\mathrm{d}\mu (m_i v^2/2)\,\mathrm{St}_{QL}(f_i) + \sum_{\pm}\int\frac{\mathrm{d}k}{2\pi}2\gamma_k^{\pm}\frac{|B_k^{\pm}|^2}{4\pi} = 0. \tag{11}$$

As a result we obtain the well known expression for the growth rate (cf. Wentzel, 1974)

$$\gamma_k^{\pm} = \frac{4\pi^3\omega_k e_i^2}{km_i}\int_0^{\infty} v^2\,\mathrm{d}v\int_{-1}^{+1}\mathrm{d}\mu(1-\mu^2)(v/c)^2\,\delta(\omega_k - kv\mu \pm \omega_{ci}) \times$$

$$\times\left[\left(1-\frac{\omega_k\mu}{kv}\right)\frac{1}{v}\frac{\partial f_1}{\partial\mu} + \frac{\omega_k}{kv}\frac{\partial f_0}{\partial v}\right]. \tag{12}$$

With the help of Equation (9) we rewrite this in the form

$$\gamma_k^{\pm} = \frac{8\pi^3}{|B_k^{\pm}|^2}\int_0^{\infty}\mathrm{d}v\,m_i v^4\int_{-1}^{+1}\mathrm{d}\mu\left[\frac{e_i^2}{2m_i^2c^2}|B_k^{\pm}|^2\,\delta(\omega_k - kv\mu \pm \omega_{ci})\right]\times$$

$$\times\frac{(1-\mu^2)}{(\nu_+ + \nu_-)}\left(\frac{\omega_k}{kv}\right)\left(1-\frac{\omega_k\mu}{kv}\right)\left\{\frac{\dot{N}}{4\pi v^2}\frac{\delta(v-u)}{(1-\mu^2)}[1+\mu-2\eta(\mu+\cos\alpha)] + \right.$$

$$\left. + \left(\frac{v}{2}\right)\cos\alpha\frac{\partial f_o}{\partial x} - \left(\frac{\mu uv}{6}\right)\frac{\mathrm{d}\ln B}{\mathrm{d}x}\frac{\partial f_0}{\partial v} + 2\nu_{\mp}\left(\frac{\omega_k}{k}\right)\frac{\partial f_0}{\partial v}\right\}, \tag{13}$$

where the values $\nu_\pm$ are given by the expressions

$$\nu_\pm = \frac{\pi e_i^2}{m_i^2 c^2} \int \frac{\mathrm{d}k}{2\pi} \, |B_k^\pm|^2 \, \delta(\omega_k - kv\mu \pm \omega_{ci}) \, . \tag{14}$$

Let us note that according to the discussion in the preceding section the cometary ions with given values of v, μ excite in the linear approximation the Alfvén waves with only one direction of the wave vector k and one polarization. Therefore, the last term within curly brackets in the expression (13) turns to be zero. This is due to the fact that the scattering of cometary ions in the system of coordinates moving with the wave excited by them does not change the energy of the ions. It is possible, however, that for some reasons (for example, due to the induced scattering of linearly excited waves into the region of phase space where the waves weakly damp in the linear approximation) there are Alfvén waves propagating in opposite directions and falling into the resonance with the same cometary ions. Then these waves experience the nonlinear damping described by the last term in Equation (13). The resonant cometary ions in their turn experience the Fermi acceleration of the second type due to the scattering by the counter-streaming Alfvén waves.

2.3. Growth rate and spectrum of excited waves

Here we restrict ourselves to the calculation of the growth rate of Alfvén waves at far distances from the comet where the number density of cometary ions as well as the growth rate is small and therefore we can neglect the nonlinear interaction of excited low amplitude Alfvén waves. To perform this calculation we have to find first the distribution function of cometary ions from Equation (10). In its turn this requires the solution of a set of hydrodynamic equations for the cometary component of plasma that could be derived from the same equation (Wallis and Ong, 1975; Galeev *et al.*, 1985). For an arbitrary angle α it is not easy to find the spatial distribution of the magnetic field. Therefore, as it is done usually (Forman and Webb, 1985; and references quoted therein) we assume here that in the case of the one-dimensional flow the magnetic field is frozen into the solar wind plasma and approximately conserves its direction, i.e., $u(x)B(x) = \text{const}$. Then using the mass conservation equation

$$\mathrm{d}(\rho u)/\mathrm{d}x = \dot{N} m_i \, , \tag{15}$$

and the new variables (u, $\chi \equiv v^3 u$) instead of the old ones (x, v) we find the following solutions of Equation (10) (Galeev *et al.*, 1987):

$$f_0(u, \chi) = \frac{3}{4\pi} \int_{u_\infty}^{u} \delta(\chi/u' - u'^3) \frac{\mathrm{d}(\rho' u')}{\mathrm{d}u'} \, \mathrm{d}u' \, , \tag{16}$$

where the mass density $\rho(x)$ and the hydrodynamical velocity of plasma $u(x)$ are single values functions of the coordinate x, and u_∞ is the value of the unperturbed solar wind velocity far from the comet.

While calculating the derivative of this distribution function entering the expression (13) for the growth rate we neglect terms that do not have δ-function singularities. This is justified in the case of large distances from the comet when the velocity decrease of the solar wind is negligible: $|u_\infty - u| \ll u_\infty$. As a result the expression (13) is reduced to a simple form

$$\gamma_k^\pm = \frac{2\pi^2 \omega_k \omega_{ci}}{k^3} \frac{\dot{N} m_i}{|B_k^\pm|^2} \left\{ \left(1 \pm \frac{\omega_{ci}}{k u_\infty}\right) - 2\eta \left(\cos\alpha \pm \frac{\omega_{ci}}{k u_\infty}\right) + \right.$$
$$\left. + \frac{3}{8}\left(\cos\alpha \mp \frac{\omega_{ci}}{k u_\infty}\right)\left(1 - \frac{\omega_{ci}^2}{k^2 u_\infty^2}\right)\right\}. \tag{17}$$

Now we see that this result has verified the main conclusion of our qualitative consideration in Section 2.1, that the ions with the given resonant velocity $v\mu = \pm\, \omega_{ci}/k$ generate only Alfvén waves with the definite direction of propagation and polarization indicated in Figure 2. This is because the main contribution to the Alfvén wave generation comes from the locally produced ions (the first two terms in the curly brackets of Equation (17) and the contributions from the effects of the free motion of ions along the magnetic field lines and the adiabatic heating (cooling) are smaller. Nevertheless, the correct numerical value of the growth rate has to be determined taking into account all three effects.

In order to compare the theory with the *in situ* measurements of Alfvén wave intensity near the comets Giacobini–Zinner (Smith *et al.*, 1986) and Halley (Galeev *et al.*, 1986; Johnstone *et al.*, 1987), it is useful to calculate also the integral intensity and frequency spectrum of the Alfvén waves. To do this we use the quasi-linear equation for the wave growth in the form

$$[u + (\omega_k/k)\cos\alpha] \frac{\mathrm{d}}{\mathrm{d}x} |B_k^\pm|^2 = 2\gamma_k^\pm\, |B_k^\pm|^2 . \tag{18}$$

The integral energy density of all linearly growing Alfvén waves is found by integration over the wave vectors of the equation

$$\sum_\pm \int \frac{\mathrm{d}k}{2\pi} \frac{|B_k^\pm|^2}{4\pi} \approx \varepsilon(\alpha) \frac{m_i u_\infty^2}{2} \int_{-\infty}^{x} \frac{\dot{N}\,\mathrm{d}x}{u_\infty} , \tag{19}$$

where

$$\varepsilon(\alpha) = (V_{\mathrm{A}}/u)\,[1 + \tfrac{7}{4}\cos^2\alpha - \tfrac{1}{4}\cos^4\alpha + \tfrac{1}{16}\sin^4\alpha] ,$$

is the relative fraction of the energy of cometary ions transferred to Alfvén waves. This fraction is larger than that calculated in Section 2.1 (see Equation (2)) because there we did not take into account the fact that besides the locally produced cometary ions, some ions produced closer to the comet can leak forward along the magnetic field lines, thus increasing the local density of cometary ions. Naturally, the contribution of this effect to the wave excitation (part of the second term and the third term in the expression for ε) turns out to be zero in the case of plasma flow perpendicular to the magnetic field. Finally, the effect of adiabatic heating of ions on the wave growth is numerically small and turns out to be zero in the case of the plasma flow along the magnetic field (see the last term in the expression for ε).

Similarly one can find the spectral energy density of Alfvén waves from Equation (18). In the short wave limit ($k \gg \omega_{ci}/u_\infty$), the spectrum is described by the power law

$$\frac{|B_k^\pm|^2}{8\pi^2} = (1 - \tfrac{3}{8}\cos\alpha) \int_{-\infty}^{x} m_i \dot{N}\, \mathrm{d}x \left(\frac{\omega_{ci} V_\mathrm{A}}{2u_\infty k^2}\right). \tag{20}$$

We see that short waves of different polarization have the same amplitude and as indicated in Figure 2 they propagate in the same direction. Therefore, these waves can be considered linearly polarized with the accuracy to small terms of the order of $\sim \omega_{ci}/ku_\infty \ll 1$.

In situ measurements of magnetic field oscillations near the comet Giacobini–Zinner (Tsurutani and Smith, 1968a) have discovered that the intensity of the magnetic field fluctuations is proportional to the number density of injected cometary ions. This result agrees with the theory (see Equation (19)) and was confirmed during the 'Vega-1' encounter with the comet Halley (Galeev *et al.*, 1986). Moreover, the energy density of Alfvén waves according to the measurements aboard 'Giotto' (Johnstone *et al.*, 1987) agrees numerically with our theoretical estimate (19). The question of power-law index for the power-law spectrum of the excited waves in the short-wave range seems to be more complicated since it is difficult to make distinction in measurements between the two close values $\frac{5}{3}$ and 2 corresponding to the cases of Kolmogoroff turbulence and weak plasma turbulence considered above. Nevertheless, we can state that the measurements (Tsurutani and Smith, 1968a; Johnstone *et al.*, 1987) do not contradict our theory. The theoretical explanation of the observed linear polarization of long wavelength Alfvén waves (Tsurutani and Smith, 1986b) needs further study.

3. Conclusion

The theory of Alfvén wave generation by cometary ions presented here agrees, in general, well with *in situ* measurements. However, we would like in conclusion to turn our attention to the important porblem related to the strong Alfvén turbulence near comets. The matter is that already during the encounter with the comet Giacobini–Zinner high fluxes of energetic ions were detected in the large region filled by Alfvén waves (Hynds *et al.*, 1986). It was natural to assume that this fact can be explained by the Fermi acceleration f cometary ions in a turbulent solar wind plasma. However, the first calculations of such acceleration (Gribov *et al.*, 1986) encountered serious difficulties. As we have mentioned this Fermi acceleration of the second kind by Alfvén waves is possible only if waves propagating in opposite directions and both falling into resonance with accelerated ions are present in a plasma (see also Forman nd Webb, 1985).

However, according to the linear theory the given group of ions resonates only with the waves propagating in one direction. The numerical computation of induced scattering of Alfvén waves by plasma ions show that the intensity of waves generated as a result of such scattering and propagating in the opposite direction relative the original wave is very low (Galeev *et al.*, 1987). Therefore, the quasi-linear theory of Alfvén wave generation by cometary ions can be considered only as the first step to the development of a more general nonlinear theory.

Acknowledgements

The authors thank Professors V. D. Shapiro, V. I. Shevchenko, and K. Szegö for numerous fruitful discussions of problems considered here.

References

Alfvén, H.: 1942, *Arkiv. Mat. Astron. Fyzik* **29B**, No. 2.
Belcher, J. W. and Davis, L., Jr.: 1971, *J. Geophys. Res.* **76**, 3534.
Bell, A.R.: 1978, *Monthly Notices Roy. Astron. Soc.* **1182**, 147.
Forman, M. A. and Webb, G. M.: 1985, in R. G. Stone and B. T. Tsurutani, (ed.), *Collisionless Shocks in the Heliosphere: A Tutorial Review*, AGU, Washington, p. 91.
Galeev, A. A., Gribov, B. E., and Gringauz, K. I. *et al.*: 1986, *Geophys. Res. Letters* **13**, 801.
Galeev, A. A., Polyudov, A. N., Sagdeev, R.Z. *et al.*: 1987, *Zh. Eksperim. Teor. Fiz.* **92**, 2090.
Gribov, B. E., Kecskemety, K., and Sagdeev, R. Z. *et al.*: 1986, *Proc. 20th ESLAB Symp. on the Exploration of Halley's Comet, Heidelberg, 27–30 October*, ESA SP-250, p. 271.
Hollweg, J. V.: 1985, in B. Buti (ed.), *Advances in Space Plasma Physics*, World Scientific, Singapore, p. 77.
Hynds, R. J., Cowley, S. W., and Sanderson, T. R. *et al.*: 1986, *Science* **232**, 361.
Johnstone, A. D., Glassmeier, K. H., Acuna, M. *et al.*: 1987, *Astron. Astrophys.* (in press).
Luhmann, J.: 1976, *J. Geophys. Res.* **81**, 2089.
Landau, L. D. and Lifshitz, E. M.: 1960, *Field Theory*, Fizmatgiz, Moscow, pp. 73 and 80.
Sagdeev, R. Z. and Galeev, A. A.: 1969, *Nonlinear Plasma Theory*, Benjamin, New York, p. 55.
Sagdeev, R. Z., Shapiro, V. D., Shevchenko, V. I., and Szego, K.: 1986, *Geophys. Res. Letters* **13**, 85.
Smith, E. J., Tsurutani, B. T., and Slavin, J. A. *et al.*: 1986, *Science* **232**, 382.
Tsurutani, B. T. and Smith, E. J.: 1986a, *Geophys. Res. Letters* **13**, 259.
Tsurutani, B. T. and Smith, E. J.: 1986b, *Geophys. Res. Letters* **13**, 263.
Wentzel, D. G.: 1974, *Ann. Rev. Astron. Astrophys.* **12**, 71.
Winske, D., Wu, S. S., Li, Y. Y., Mou, Z. Z., and Guo, S. Y.: 1985, *J. Geophys. Res.* **90**, 2713.

MODELING A COMETARY NUCLEUS*

K. SZEGÖ

Central Research Institute for Physics, Budapest, Hungary

(Received 30 June, 1987)

Abstract. Considerations are summarized concerning the physical properties of and plasma phenomena around a cometary nucleus aiming at a new model of the nucleus and its interaction with the solar wind.

1. Introduction

In March 1986 a flotilla of spacecrafts encountered Comet Halley when it was close to the equatorial plane on its way out to the distant regions of the solar system. The encounters made a new benchmark in space research. Not only the *in situ* investigations of the small bodies of the solar system developed in a spectacular way, but the well-coordinated efforts of many space agencies and astronomers' organizations opened up new horizons in space research and, one hopes, also triggered future cooperative ventures.

While the experimentalists were busy putting together the devices, theoreticians completed their share to be prepared for the date with the comet. The full pre-encounter theory was beautifully summarized by Mendis *et al.* (1985). If nothing else is stated, we always refer to this paper. In many details the theory proved to be true, but fundamental differences were also detected.

In what follows some of those differences will be summarized and efforts will be made to fit them into a next generation model. It is unusual to publish preliminary considerations and not nicely polished results. I was always taught that there is no difference between a physicist and an acrobat; the performance should go with ease and smile, the sweat is not for the public. But may be this special occasion allows us to have a look behind the scenes.

2. The Cometary Nucleus

One new observational evidence has changed a paradigm. Earlier it was thought that the density of the nucleus is about 1 g cm^{-3} on the basis that the most abundant material in it is water. The pioneering work of Rickman (1986) revealed that the nucleus might be considerable less dense. The density derived from non-gravitational forces turned out to be about 0.1–0.3 g cm^{-3}. This was reanalysed by Sagdeev *et al.* (1987a) leading to an average density about 0.6 g cm^{-3} which is still less than anticipated. Even if the uncertainties in the non-gravitational parameters are higher than it was estimated (Rickman, 1986), the nucleus material now seems to be less dense than previously

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 439–449.

believed. Greenberg and Grim (1986) suggested a fluffy structure of silicate and other grain components where the voids are filled with volatiles. This is consistent well with the observations. Whipple (1987) called this type of nucleus a dirty snowdrift to distinguish it from the previous dirty snowball model. We conclude that the material is more fluffy, less dense, and probably more friable than it was thought to be.

Earlier the surface was considered to be homogeneous in accordance with the snowball model. In the vertical direction the core-mantle structure was accepted. After the missions the author of the present paper is convinced that there are two kinds of cometary surfaces, at least in the case of P/Halley. The first is a thicker, hot, inactive crust, impervious to volatiles; the second is a fresh cometary material rich in volatiles. This is the basic source of the jets. The albedo of the surface is very low (Sagdeev *et al.*, 1986a).

The complete set of equations describing the thermodynamics of the core-mantle model and estimating the gas and dust outflow was first solved in the friable sponge model of Horanyi *et al.* (1984). This model incorporated the energy balance on the surface for solar irradiation, thermal radiation, and heat transferred to the mantle; the heat conduction in the mantle assuming thermal balance between the core material and the outflowing gas temperature; and energy balance at the core-mantle boundary where the gas sublimation takes place. The core acts as a heat sink. The friability of the mantle material allows dust be dragged by the outflowing gas in quantities proportional to the outflow velocity. The outflow velocity differs from the thermal velocity of the gas (Gombosi *et al.*, 1986) if the process is considered similar to gas discharged from a reservoir to low-pressure external medium. Some of the assumptions can be challenged or improved (Gombosi *et al.*, 1986; Ip and Rickman, 1986). But one of the basic results, the relationship between the mantle thickness and the emitted amount of gas, surface and core-mantle boundary temperature is to a good approximation valid in our opinion. These curves are shown in Figure 1.

As it can be seen from this figure, the inactive places should be hot, about 400 K. Such a high surface temperature was actually measured by IKS aboard VEGA (Emerich *et al.*, 1986); from these measurements the authors inferred a hot object of 5 km radius; but they are able to find solution with different vertical and longitudinal sizes, or even disintegrated hot spots are possible. The heat must come from the cometary surface. In principle other factors may influence the measurements, e.g., the dust temperature in the atmosphere. Whereas it is true that the grains leaving the surface are heated up during the first few ten meters according to coupled dust-gas hydrodynamical equations (Gombosi *et al.*, 1982), later on they are cooled down if only the hydrodynamics is considered. However, the Sun heats the grains too. To estimate this temperature we may use the same heat balance as for the surface. As there is no outflowing heat, the grains should be slightly warmer than the inactive surface. But the temperature of small, absorbing grains depends also on their size, since if the size is much smaller than the wavelength, the size determines their emissivity. As shown by Hanner (1986), 0.1 μ grains can be 300–400 K hotter than larger ones. However, from the low optical thickness of the dust attested by the imaging experiments (Keller *et al.*, 1987; Sagdeev

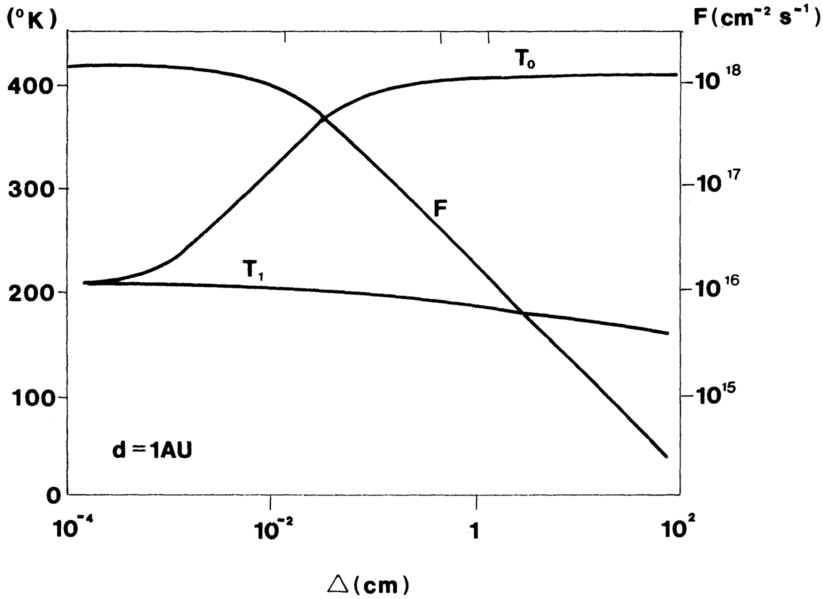


Fig. 1. The surface temperature (T_0) and core/mantle temperature (T_1) as well as the production rate (F) are shown as functions of the mantle thickness at 1 AU.

et al., 1985) this contribution should not be important. Hence, we assume that IKS really measured the inactive layer temperature and size.

It is known from the literature that there are possibilities for building up an inactive layer. This was studied by Moore *et al.* (1983) experimentally. They exposed cometary type ice mixtures to conditions persisting in the Oort cloud, inlcuding the simulation of cosmic-ray radiation. They concluded that 1% of the material was converted into a nonvolatile residue. After the first few apparition this can turn to be, e.g., the inactive crust. The procedure was discussed by Johnson *et al.* (1986) also.

It was Ip (1985) who first concluded on the basis of the GIOTTO HMC data that the active surface is only 15% of the total. The 3-dimensional analysis of the most prominent jet features seen on VEGA-2 images also resulted in the finding that the total surface of the most active jet sources is small. From the images it is also evident that the general activity is negligible on the dark hemisphere. This establishes the low heat conductivity of the surface which is consistent with the fluffy material referred to above. Whereas it is true that the most active dust jets are well localized, the source of the other brightness features on the VEGA images is not known. Those features cannot be just a part of the coma since what we call coma in ground-based observation is not visible on the VEGA images. So there is some unlocalized activity on the Sunward surface which may consist of small jets or may be of a different nature. Their source extension is unknown.

We may try to attack this problem from the other end, i.e., to try to estimate the active part of the surface. During the VEGA-1 encounter 1.3×10^{30} mol s^{-1} gas production rate was reported with a factor of two error bar (Gringauz *et al.*, 1985). At the GIOTTO and VEGA-2 encounter IUE measurements indicated about half of the previous number (Festou *et al.*, 1986), 5.2×10^{29} mol s^{-1} at GIOTTO, 5.6×10^{29} at VEGA-2 for water production. Water is about 80% of the total gas production. To account for those values on the basis of the values shown in Figure 1, a 50–100 km^2 active surface is required for zero-mantle thickness. These are actual surface areas, not the perpendicular-to-Sun cross sections.

The approximate shape of the comet is also known (Sagdeev *et al.*, 1986b); for our consideration it can be approximated as a truncated cone, capped at each end by hemispheres. The long axis is 16 km, and the radii of the hemispheres are 4 and 2.5 km, respectively. The total volume is 500 km^3, and the total surface is 350 km^2. The maximal and minimal perpendicular cross sections of the object are about 100 and 50 km^2, respectively. The angle between the Sun-nucleus line and the major axis was 22, 110, 130 deg, respectively, for the VEGA-1, VEGA-2, and GIOTTO fly-by based on the analysis of Sagdeev *et al.* (1987b).

However, the uncertainty of a factor of two in the number above starts to hurt if we would like to make definite conclusions. The inactive surface size has an uncertainty of a factor of two. The amount of emitted gas strongly depends on the mantle thickness, even a mm thick mantle suppresses it to half the value, the activity pattern of the nucleus shows strong variability, etc. So we need more input.

This may come from the investigation of the brightness distribution of the jets and other features. For this purpose the VEGA-2 imaging data serve as a unique data set. During the VEGA-2 encounter several large size images were exposed from a distance closer than 30000 km from the nucleus. On these images extended jets and other brightness features are distinguishable (Smith *et al.*, 1986). These images were used to reconstruct the spatial orientation of the most prominent jets. Recently, the brightness distribution of the jets was also analysed (Szegö *et al.*, 1987). The preliminary analysis suggests that grain disintegration takes place in the vicinity of the nucleus. As the disintegration cannot be caused by collisions, it was assumed that the grains leaving the surface are still rich in volatiles. This does not contradict the observation of PUMA aboard VEGA (L. M. Mukhin, private communication). As the grains are heated up, the volatiles evaporate and this process leads to the disintegration. If it is so, we have to conclude that at least the jet sources are not covered by dust, that there the naked pristine ice is exposed to the solar radiation. A rough estimate of the source surface, based on the study of Smith *et al.* (1986) for the VEGA-2 fly-by, can be about 5% of the surface, 20 km^2. This can account for 2×10^{29} mol s^{-1} gas production, 20% of the total. If for the inactive surface we accept a medium value, about 50 km^2, the 'normal' activity can spread over 100 km^2 of the dayside surface. The above-mentioned brightness analysis (Szegö *et al.*, 1987) allows the preliminary conclusion that there is no grain disintegration in the Sunward, non-jet type activity. This is really not a solid conclusion yet because it is very difficult to differentiate jet and non-jet contributions close to the nucleus. But providing it is so, the non-jet activity comes from a surface covered by thin dust. A quickly changing thin dust layer is very convenient to explain the strong variability of the activity.

It is an important but yet unanswered question whether the jet sources are fixed relative to the nucleus or whether they may change position. Temporal changes in the activity has been reported (Kaneda *et al.*, 1986). But as we know, the rotation of this irregularly shaped nucleus is complicated. It has been proven that the long axis makes one rotation in each 54 hr about a spin axis which is inclined approximately 80 deg to it (Sagdeev *et al.*, 1987a, b). This rotation period is also in agreement with the curvature of the dust jets (S. Larson, private communication). Evidence is growing, however, that the overall brightness shows a 7.4 d period (Millis and Schleier, 1986; Festou, 1987). The motion accommodating these two observations is not yet completely understood. This is why it is difficult to tell whether the changes in the individual jet source activity are connected with the variable illumination or whether there are pockets of jet material which become empty after a while. Observations of pure gas jets were also reported (A'Hearn *et al.*, 1986). Whether they come from the surface or are emitted by dust in flight is not yet clear.

To summarize, the irregularly-shaped, very dark of the surface nucleus is covered partly by inactive, meter-thick layers or spots formed probably even in the Oort cloud; there are active spots where pristine ice-dust composite material or probably only gas is emitted when the Sun shines above; and at least there is the most abundant surface part where the core-mantle model gives a good description with a quickly changing mantle on the top.

After the physical description chemistry ought to be discussed but due to the limited knowledge of the author about this field it has to be emitted in spite of the many interesting results collected by ground-based and space-borne observations. The most important conclusion is that, by composition, P/Halley belongs to our solar system.

3. Plasma Environment

Several surprises came from the plasma investigations. It was recognized early (Schmidt and Wegman, 1982) that mass loading, i.e., photoionized gas picked up by the solar wind plays an important role around comets. But the intensive MHD turbulence extending over many times of the nucleus-bow shock distance, and the acceleration of heavy ions due to it is new and not completely understood. The bow shock itself (in the quasi-parallel case) is much more diffusive than the planetary cases and it is similar to cosmic-ray shock, or it may even be a new type of shock. The downstream region was much more structured than anticipated before.

The dust envelope extends to about 1.5×10^5 km from the nucleus. Depending on the gas activity, this envelope can be outside (e.g., for P/Giacobini–Zinner) or inside the bow shock (e.g., P/Halley). The dust certainly can be electrically charged. However, since the dust-to-gas mass ratio is < 0.3, and the atomic weight of the dust material is heavier, the overall density of dust is at least an order of magnitude less than the gas density. In spite of the fact that several boundaries were observed in the dust distribution (Vaisberg *et al.*, 1986) and some of those coincide with plasma boundaries, at this stage we think that dust does not play a decisive role in the plasma phaenomena.

Even a brief summary of all the plasma phaenomena would exceed the scope of this paper. This is why we shall concentrate only to the upstream region and the bow shock.

The first surprise came when ICE encountered Comet Giacobini–Zinner. Aboard ICE flew a very sensitive wave analyzer package. In a vast region around the comet exceptionally intense plasma waves were detected, extending more than 4×10^6 km around the nucleus. Close to the nucleus the electromagnetic wave level were 'as intense as any signals previously recorded by this wave instrument' (Scarf *et al.*, 1986). It was concluded that 'many important cometary processes are influenced by microscopic plasma physics phenomena to an extent not anticipated in fluid dynamics or magnetohydrodynamic models'. The further fly-by missions confirmed this on a scale proportional to gas production rate. The intensive waves were accompanied by fluxes of energetic particles, with energies higher than anticipated before. (To underline how unexpected this was, the author can recall debates from the time of the payload selection for the VEGA mision when the usefulness of an energetic particle analyzer was questioned on a cometary mission.)

Before the missions the general belief was that cometary gas will be ionized and these ions will be picked up by the frozen-in magnetic field of the solar wind. In the solar wind frame the motion of the particle is a gyration with speed $E/B = V \sin \theta$, where θ is the angle between the magnetic field and the solar wind direction. The gyrocenter moves toward the Sun at a speed of $V \cos \theta$. In the cometary frame the particle moves away from the Sun, its energy varies between 0 and $2MV^2 \sin^2 \theta$, M being the ion mass. For a water ion, the maximal energy corresponds to 36 keV. The necessary time for the pick up is about 100 s (Wu *et al.*, 1986). The small speed the ions retain along the field line was though to be unimportant.

After the missions, as we learned, this small speed is the culprit for the most of the fascinating results. As the initial velocity distribution of the cometary ions is anisotropic, ion-cyclotron resonance instability is developing. This excites Alfvén waves moving almost along the magnetic field lines. The resonant interaction of these waves with particles isotropises their velocity via pitch angle scattering. The particle scattering to small pitch angles is due to the normal Doppler cyclotron resonance

$$\omega - kv_{\|} = \omega_{\mathrm{Hi}} ,$$

which corresponds to left-hand circularly polarized waves. The scattering to large angles takes place via anomalous Doppler resonance

$$\omega - kv_{\|} = -\omega_{\mathrm{Hi}} ,$$

and yields right-hand polarized waves. We assume that there is a balance between isotropization due to pitch angle scattering and the creation of new cometary ions that maintain a certain amount of anisotropy. The weak turbulence theory of this phenomena was developed by Sagdeev *et al.* (1986). According to this the characteristic frequency for water ions is of the order of $\frac{1}{100}$ Hz, fully confirmed by observations. The Alfvén waves, continuously excited by the anisotropy, are convected towards the comet by the solar wind. The dependence of the wave energy density, $\Sigma\, |B_k^2|/4\pi$, on the distance x from the cometary nucleus was calculated by Galeev *et al.* (1986). This quantity $\mathrm{d}W/\mathrm{d}x$ is proportional to the growth of the Alfvén waves

$$QMV_{\mathrm{A}} \exp(-rV_g\tau)/r^2V_g\tau\,,$$

which is stabilized by the saturation of wave growth due to induced scattering of these waves on solar wind protons

$$(T_p/mu_\infty^2)\,(\omega_{\mathrm{Hi}}/V_{\mathrm{A}})\,(W^2/B_0^2)\,.$$

The experimental results agree well with this estimate. The experimental spectral density for $k > \omega_{\mathrm{Hi}}/u$, given by

$$|B_k^2| \sim 1/k^{5/3}$$

(Tsurutani and Smith, 1986), is also close to the consequences of the theory above. However, the experimental spectral density also contains a long wave length part $k < \omega_{\mathrm{Hi}}/u$. The most probable mechanism of excitation for the long wavelength part of the spectra is the firehose instability. This develops if $p_\parallel > p_\perp + B_0^2/4\pi$. This instability is possible sufficiently close to the bow shock, about 2–3 $\times$ 10^6 km from the nucleus in the case of P/Halley. As measurements aboard SUISEI show (Mukai *et al.*, 1986) in that part of the upstream region $p_\parallel/p_\perp > 1.5$ which makes this scenario feasible.

As was mentioned already, in many experiments particle acceleration was observed. The most probable mechanism for this is stochastic Fermi acceleration as concluded by Ip and Axford (1986), having investigated all the possible mechanisms. The kinetic equation for the cometary ion-distribution function upstream was discussed in Sagdeev *et al.* (1986c). The change of the distribution function is caused partly by diffusion and by the continuous creation of new ions. The diffusion coefficient describes the particle scattering on Alfvén waves, and is proportional to

$$D \sim \int \mathrm{d}k\, |B_k^{\pm}|^2\, (1 - \cos^2\theta)\delta(\omega_k - k_\parallel \pm \omega_{\mathrm{Hi}})\,,$$

where B_k is the magnetic field spectra density. Assuming Alfvén waves going in both directions, from the kinetic equation the observed particle acceleration can be accounted for (Gribov *et al.*, 1986). However, there are two questions still not solved in this respect. The first: this is resonant acceleration by MHD waves but the resonance condition $\omega_{\mathrm{Hi}} = +kv\cos\theta$ for energetic particles with $v > u$ can be met only for long wavelength oscillations $k < \omega_{\mathrm{Hi}}/u$. For short wavelength the acceleration is possible only in a small pitch angle region $\cos\theta < u < v$. The second question concerns the fact that the magnetic field irregularities on which the particles are stochastically accelerated should not move together with the same velocity. One possibility is that oppositely-propagating Alfvén waves are excited (e.g., by firehose instability) as it was assumed by Gribov *et al.* (1963), but this situation was not observed experimentally. The other possibility is that the acceleration happens via oblique magnetosonic waves when their phase velocity depends on the propagation angle. If both mechanisms are present, there is a break in the observed ion spectra. This is not in contradiction with the observations, e.g., in TUNDE experiment (Kecskemety *et al.*, 1986) but this issiue is far from being closed.

Upstream the ions move with supersonic velocity, downstream they are slowed down to subsonic values. Around planets such change is not possible without shock formation. But at the ICE–Giacobini–Zinner encounter the clear signature of the bow shock crossing was missing, and even the question arose whether either a continuous transition is also possible or the shock is extremely weak. The observations were termed as bow wave.

The Halley missions made this interpretation unlikely. However, it has been acknowledged that the already mentioned MHD turbulence is not just a by-product of

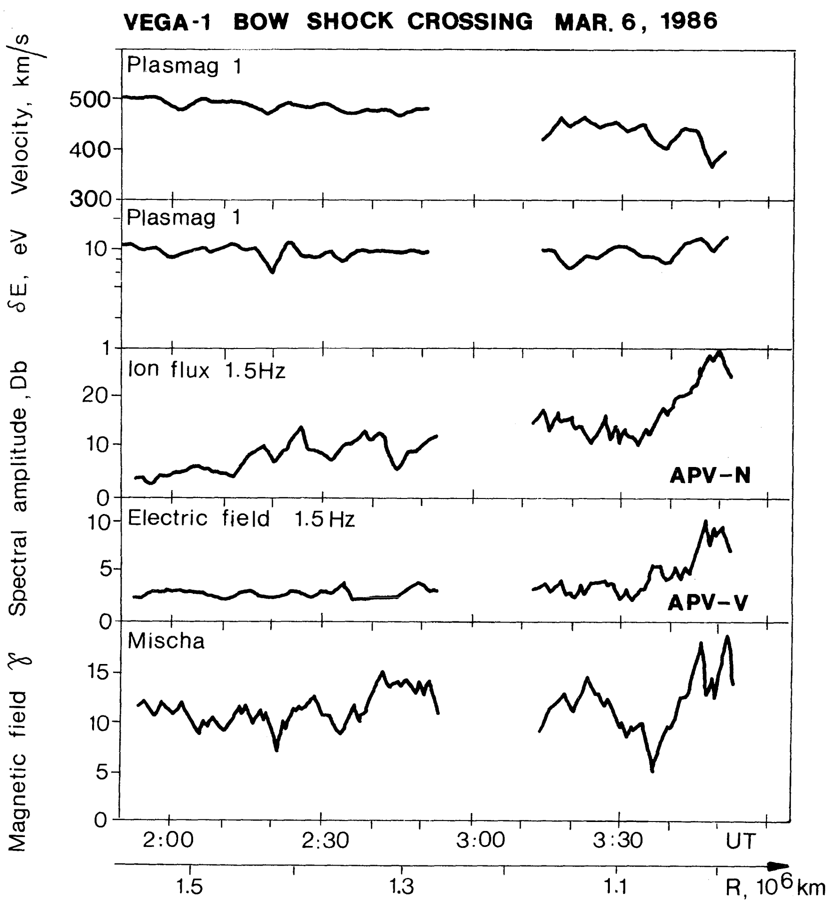


Fig. 2. The behaviour of solar wind parameters during the inbound bow shock crossing by the 'Vega-1' spacecraft is shown. *From top to bottom*: solar wind velocity and effective temperature measured in solar direction; spectral amplitude of ion flux fluctuations and electric field oscillations with the frequency f = 1.5 Hz; magnetic field strength.

the mass loading but that this defines the structure and position of the bow shock. This was already envisaged in Sagdeev *et al.* (1986c) and later elaborated in Galeev *et al.* (1986). An important factor in it is that the pitch angle scattering destroys the magnetic moment conservation and changes the degrees of freedom of the plasma flow (from $\gamma = 2$ to $\gamma = \frac{5}{3}$). By solving the one-dimensional equations for the mass-loaded case it was proven that the shock formation is unavoidable. Taking $M = 2$ at the shock from numerical simulations, the subsolar stand-off distance (2.7×10^5 km) and the crossing of the front for the different spacecrafts (about 1.1×10^6 km) were calculated, and were found to be in good agreement with the observations.

During the encounters quasi-parallel and quasi-perpendicular crossings were observed. The identification of the crossing was possible only by jointly analyzing magnetic, field and particle measurements (Figure 2). Due to the magnetic field variations $\Delta B/B_0 \sim 0(1)$ induces by the strong MHD turbulence, the upstream particles were strongly heated; but it was the changes in their behaviour observed together with waves below but close to the lower hybrid frequencies that gave the necessary confidence to define the bow shock crossing. These waves were excited by the picked-up cometary ions reflected from the shock front and accelerated there. The bursts of MHD turbulence are precursors of the shock. The width of the shock also deserves attention. In the quasi-perpendicular case numerical simulations were done by Galeev *et al.* (1985) who concluded that the width is defined by the heavy ion-Larmor radius, which is about 10^4 km.

The quasi-parallel shock is quite different – they are very similar to cosmic-ray diffusive shocks (Sagdeev *et al.*, 1986d). In both cases the high-energy particles moving along the magnetic field excite intensive Alfvén waves. Due to this, the particles escaping from the shock upstream behave as if diffusing due to the scattering on the waves. The shock width in the quasi-parallel case is basically the diffusion length, which is of the order of 10 Larmor radii, i.e., about 10^5 km.

As to downstream region many questions are still open; we shall not discuss these here. The tail phaenomena is also very interesting, but that too is out of the scope of the present work.

In the formulae ω and k are the wave frequency and wave number, respectively; $v_{\|}$ is the velocity component parallel to the magnetic field; ω_{Hi} is the ion gyrofrequency; Q is the gas production rate; V_{A} is the Alfvén velocity; V_g is the gas outflow velocity; τ is the characteristic time for ionization; r is the distance from the comet; T_p is the proton temperature; u_{∞} is the undisturbed solar wind velocity; $p_{\|}$ and $p_{\perp}$ are the pressures parallel and perpendicular to the magnetic field.

Acknowledgements

The author is indebted to Professors R. Z. Sagdeev, V. D. Shapiro, and V. I. Shevchenko for their useful comments.

References

A'Hearn, M., Birch, P. V., and Klinglesmith, D. A.: 1986, *20th Symposium on the Exploration of Halley's Comet*, ESA SP-250, Vol. I, p. 483.

Emerich, C., Lamarre, J. M., Moroz, V. I., Combes, M., Sanko, N. F., Nikolsky, Yu. V., Rocard, F., Gispert, R., Coron, N., Bibring, J. P., Encrenaz, T., and Crovisier, J.: 1986, *20th Symposium on the Exploration of Halley's Comet*, ESA SP-250, Vol. II, p. 381.

Festou, M. C., Drossart, P., Lecacheux, J., Puel, F., Kohlk, J. L., and Encrenaz, T.: 1987, *Proc. of the Symposium on the Diversity and Similarity of Comets*, Brussels (to be published).

Galeev, A. A., Lipatov, A. S., and Sagdeev, R. Z.: 1985, *J. Eksper. Theor. Phys.* **89**, 1495.

Galeev, A. A., Gribov, B. E., Gombosi, T., Gringauz, K. I., Klimov, S. I., Oberz, P., Remizov, A. P., Riedler, W., Sagdeev, R. Z., Savin, S. P., Sokolov, A. Yu., Shapiro, V. D., Shevchenko, V. I., Szegö, K., Verigin, M. I., and Yeroshenko, Ye. G.: 1986, *GRL* **13**, 841.

Gombosi, T. I., Nagy, A. F., and Cravens, T. E.: 1986, *Rev. Geophys. Res.* **24**, 667.

Gombosi, T. I., Szegö, K., Gribov, B. E., Sagdeev, R. Z., Shapiro, V. D., Shevchenko, V. I., Cravens, T. E.: 1982, *Proc. of Int. Conf. Cometary Exploration*, Budapest, Vol. II, p. 99.

Greenberg, J. M. and Grim, R.: 1986, *20th ESLAB Symposium on the Exploration of Halley's Comet*, ESA SP-250, Vol. II, p. 255.

Gribov, B. E., Kecskemety, K., Sagdeev, R. Z., Shapiro, V. I., Shevchenko, V. I., Somogy, A. J., Szegö, K., Erdos, G., Eroshenko, E. G., Gringauz, K. I., Keppler, E., Marsden, R., Remizov, A. P., Richter, A. K., Riedler, W., Schwingenschuh, K., and Wenzel, K. P.: 1986, *20th ESLAB Symposium on the Exploration of Halley's Comet*, ESA-250, Vol. I, p. 271.

Gringauz, K. I., Gombosi, T. I., Remizov, A. P., Apathy, I., Szemery, I., Verigin, M. I., Dechlikova, L. I., Dyachkov, A. V., Keppler, E., Klimenko, I. N., Varga, A., and Vladimirova, G. A.: 1985, *Adv. Space Res.* **5**, 165.

Hanner, M. S.: 1986, *Icarus* **67**, 51.

Horanyi, M., Gombosi, T. I., Cravens, T. E., Korosmezey, A., Kecskemety, K., Nagy, A. F., and Szegö, K.: 1984, *Astrophys. J.* **278**, 449.

Ip, W. H.: 1985, *Adv. Space Res.* **5**, 233.

Ip, W. H. and Axford, W. I.: 1986, *Planetary Space Sci.* **34**, 1061.

Ip, W. H. and Rickman, H.: 1986, *Proc. of the Nucleus Sample Return Mission Workshop*, Canterbury, ESA SP-249, p. 181.

Johnson, R. E., Cooper, J. F., and Lanzerotti, L. J.: 1987, *20th ESLAB Symposium on the Exploration of Halley's Comet*, ESA SP-250, Vol. II, p. 269.

Kaneda, E., Takagi, M., Hirao, K., Shimizu, M., and Ashibara, O.: 1986, *20th ESLAB Symposium on the Exploration of Halley's Comet*, ESA SP-250, Vol. I, p. 397.

Kecskemety, K., Cravens, T. E., Afonin, V. V., Erdos, G., Eroshenko, E. G., Gan, L., Gombosi, T. I., Gringauz, K. I., Keppler, E., Klimenko, I. N., Marsden, R. G., Nagy, A. F., Remizov, A. P., Richter, A. K., Riedler, W., Schwingenschuh, K., Somogyi, A. J., Szegö, K., Tatrallyay, M., Varga, A., Verigin, M. I., and Wenzel, K. P.: 1986, *20th ESLAB Symposium on the Exploration of Halley's Comet*, ESA SP-250, Vol. I, p. 109.

Keller, H. U., Delamere, W. A., Huebner, W. F., Reitsema, H. J., Schmidt, H. U., Whipple, F. L., Wilhelm, K., Curdt, W., Kramm, R., Thomas, N., Arpigny, C., Barbieri, C., Bonnet, R. M., Cazes, S., Coradini, M., Cosmovici, C. B., Hughes, D. W., Jamar, C., Malaise, D., Schmidt, K., Schmidt, W. K. H., and Seige, P.: 1987, *Proc. of the Symposium on the Diversity and Similarity of Comets*, Brussels (to be published).

Mendis, D. A., Houpis, H. L. F., and Marconi, M. L.: 1985, *Fundamentals Cosmic Phys.* **10**, 1.

Millis, R. L. and Schleier, D. G.: 1986, *Nature* **324**, 646.

Moore, M. H., Donn, B., Khanna, R., and A'Hearn, M. F.: 1983, *Icarus* **54**, 388.

Mukai, T., Kitayama, M., Terasawa, T., and Hirao, H.: 1986, *20th ESLAB Symposium on the Exploration of Halley's Comet*, ESA SP-250, Vol. I, p. 71.

Rickman, H.: 1986, *Proc. of the Nucleus Sample Return Mission Workshop*, Canterbury, ESA SP-249, p. 149.

Sagdeev, R. Z., Avanesov, G. A., Barinov, I. V., Dabdov, A. I., Kvasikov, V. A., Moroz, V. I., Shamis, V. A., Tarnopolski, V. I., Usikov, D. A., Ziman, Ya. L., Zhukov, V. S., Szabo, F., Szegö, K., Smith, B. A., Kondor, A., Larson, S., Merenyi, E., Szabo, L., Toth, I., Varhalmi, L., Cruvellier, P., Abergel, A.,

Bertaux, J. L., Blamont, J., Danz, M., Mohlman, D., Stiller, H., and Zapfe, H.: 1985, *Adv. Space Res.* **5**, 12, 95.
Sagdeev, R. Z., Szabo, F., Avanesov, G. A., Cruvellier, P., Szabo, L., Szego, K., Abergel, A., Balazs, A., Barinov, I. V., Bertaux, J. L., Blamont, J., Detaille, M., Demarelis, E., Dul'nev, G. N., Endroczy, G., Gardos, M., Kanyo, M., Kostenko, V. I., Krasikov, V. A., Nguyen-Trong, T., Nyitrai, Z., Renyi, I., Rusznyak, P., Shamis, V. A., Smith, B., Sukhanov, K. G., Szalai, S., Tarnopolsky, V. I., Toth, I., Tsukanova, G., Valnicek, B. I., Varhalmi, L., Zaiko, Yu. K., Zatsepin, S. I., Ziman, Ya. L., Zsenei, M., and Zhukov, B. S.: 1986a, *Nature* **321**, 262.
Sagdeev, R. Z., Ziman, Ya. L., Avanesov, G. A., Tarnopolsky, V. I., Krasikov, V. A., Usikov, D. A., Dabdov, A. S., Sukhanov, K. G., Kogan, A. Yu., Kheifets, V. N., and Szegö, K.: 1986b, *20th ESLAB Symposium on the Exploration of Halley's Comet*, ESA SP-250, Vol. II, p. 307.
Sagdeev, R. Z., Shapiro, V. D., Shevchenko, V. I., and Szegö, K.: 1986, *Geophys. Res. Letters* **13**, No. 1, 85.
Sagdeev, R. Z. *et al.*: 1987a, to be published.
Sagdeev, R. Z. *et al.*: 1987b, preprint, KFKI-75/c.
Scarf, F. L., Coroniti, F. V., Kennel, Ch. F., Gurnett, D. A., Ip, W. H., and Smith, E. J.: 1986, *Science* **232**, 377.
Schmidt, H. U. and Wegman, R.: 1982, in L. L. Wilkening (ed.), *Comets*, Univ. of Arizona Press, Arizona.
Smith, B., Szegö, K., Larson, S., Merenyi, E., Toth, I., Sagdeev, R. Z., Avanesov, G. A., Krasikov, V. A., Shamis, V. A., and Tarnopolski, V. I.: 1986, *20th ESLAB Symposium on the Exploration of Halley's Comet*, ESA SP-250, Vol. II, p. 327.
Szegö, K., Toth, I., and Kondor, A.: 1987, to be published.
Tsurutani, B. T., Brinca, A. L., Smith, E. J., Thorne, R. M., Scarf, F. L., Gosling, J. T., and Ipavich, F. M.: 1986, *20th ESLAB Symposium on the Exploration of Halley's Comet*, ESA SP-250, Vol. III, p. 451.
Vaisberg, O., Smirnov, V., and Omelchenko, A.: 1986, *20th ESLAB Symposium on the Exploration of Halley's Comet*, ESA SP-250, Vol. II, p. 17.
Whipple, F. L.: 1987, *Sky and Telescope* **73**.
Wu, C. S., Winske, D., and Gaffey, J. D., Jr.: 1986, *Geophys. Res. Letters* **13**, 865.

FILAMENTATION OF VOLCANIC PLUMES ON THE JOVIAN SATELLITE IO

ANTHONY L. PERATT
Los Alamos National Laboratory, Los Alamos, New Mexico, U.S.A.

and

A. J. DESSLER
Department of Space Physics and Astronomy, Rice University, Houston, Texas, U.S.A.

(Received 17 July, 1987)

Abstract. Volcanic plumes on the Jovian satellite Io may be a visible manifestation of a plasma-arc discharge phenomenon. The amount of power in the plasma arc ($\sim 10^{11}$ W) is not enough to account for all the energy dissipated by the volcanoes. However, once a volcano is initiated by tidal and geologic processes, the dynamics of the volcanic plumes can be influenced by the plasma arcs. As initially pointed out by Gold (1979), plasma arcs are expected because of $\sim 10^6$ A currents and 400 kV potentials generated by the flow past Io of a torus of relatively dense magnetospheric plasma. We utilize our experience with laboratory plasma arcs to investigate the plume dynamics. The filamentation in the plume of the volcano Prometheus and its cross-sectional shape is quantitatively consistent with theories developed from laboratory observation.

1. Introduction

In the late 1950's and early 1960's, H. Alfvén directed a program of research on the physics of the plasma gun (Lindberg and Jacobsen, 1961, 1964; Dattner and Eninger, 1964; Wilcox *et al.*, 1964). As discussed by Wilcox *et al.*, one of the reasons motivating this research involved the origin of the planets and satellites (Alfvén, 1960; Alfvén and Wilcox, 1962). The apparent filamentary penumbra on Io may be the first direct verification of the plasma gun mechanism at work in the solar system.

The satellite Io was observed by the spacecraft Voyager 1 and Voyager 2 to be covered with volcanoes (Morabite *et al.*, 1979; Smith *et al.*, 1979; Strom *et al.*, 1979; Strom and Schneider, 1982). Nine active volcanoes were observed during the Voyager 1 encounter, eight of which were still active during the Voyager 2 flyby four months later. Detailed pictures of the plumes from one of these volcanoes were rather striking in that the plume material was ejected in a well-defined cone whose geometry showed converging (rather than diverging) matter at large lateral distances from the vent, and the plume material was concentrated into striations. Thus, we have a volcanic vent with exit velocities of about 0.5 km s^{-1}, but with the volcanic effluent concentrated into a cone with a half angle initially less than about 25° to the vent axis, and the material in the cone further

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 451–461.

tending to concentrate into filaments that terminate on a narrow, well-defined, concentric annulus (Strom and Schneider, 1982).

We wish to raise the possibility that details of the volcanic discharge are a manifestation of a plasma arc at the volcanic vent as initially suggested by Gold (1979).

2. Plasma Gun Arc Discharge Mechanism

The plasma gun or plasma focus (Alfvén, 1958; Lindberg *et al.*, 1960; Lindberg and Jacobsen, 1961; Filippov *et al.*, 1962; Alfvén and Fälthammar, 1963; Mather, 1971; Bostick *et al.*, 1975; Alfvén *et al.*, 1981) is a plasma discharge consisting of a short but finite z-pinch which forms near or at the end of a coaxial plasma discharge. In the laboratory, the coaxial discharge is formed by switching a charged capacitor bank between a center electrode and an outer conducting tube electrode (Figure 1). The discharge is manifested by a current sheath, called a penumbra, that forms between the inner and outer electrodes when the inner electrode is at a positive potential (i.e., an anode) and the outer electrode is at ground (i.e., a cathode). The current sheath is equally well-defined whether the inner electrode is at a positive or negative potential but ion and neutron production at the focus are observed only when the inner electrode is an anode. Part of the stored magnetic energy in the tube and external circuit is rapidly converted to plasma energy during the current sheath's collapse toward axis. The current flow convergence is largely due to the self-consisting nature of the current filament, whereas

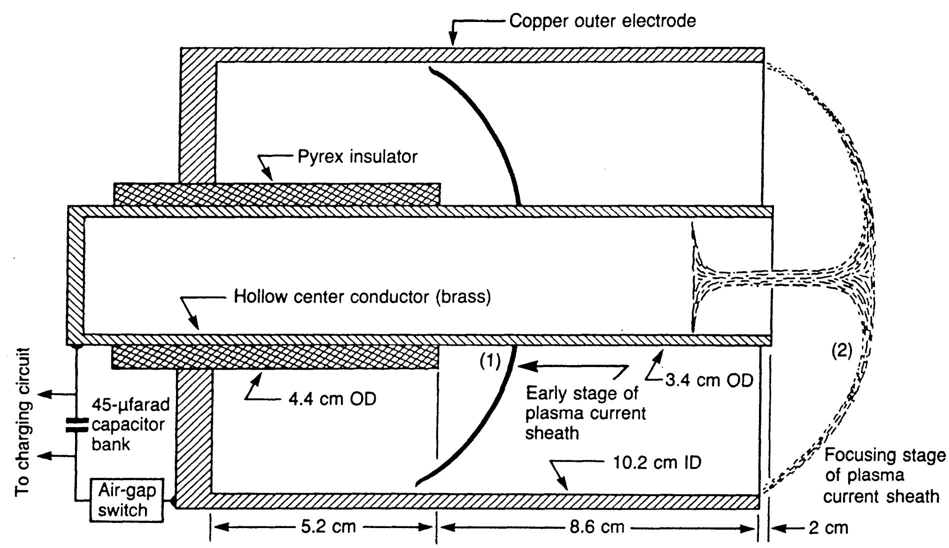


Fig. 1. Typical plasma gun apparatus. A capacitor bank is discharged through two coaxial electrodes forming a plasma current sheath between inner and outer electrodes. The $\mathbf{j} \times \mathbf{B}$ force accelerates the sheath outward to the ends of the electrodes where the inner sheath radius is forced inwards forming a columnar pinch or 'focus' on-axis. Also depicted on the diagram is the 'penumbra', the long-lived state of the current sheath at the muzzle of the plasma gun.

the heating and compression from the r, z implosion on the axis are due both to the magnetic forces of the current-carrying plasma filament and to inertial forces. Partial conversion of the kinetic energy of the imploding axisymmetric current to internal heat energy may occur during the implosive phase owing to self-collision.

Develoment of the plasma current leading to the formation of a plasma focus at the center electrode terminus can be conveniently subdivided into three main phases. First is the initial gas breakdown and the formation of a parabolic current front. Second is the hydromagnetic acceleration of a current sheath toward the open end. The third part

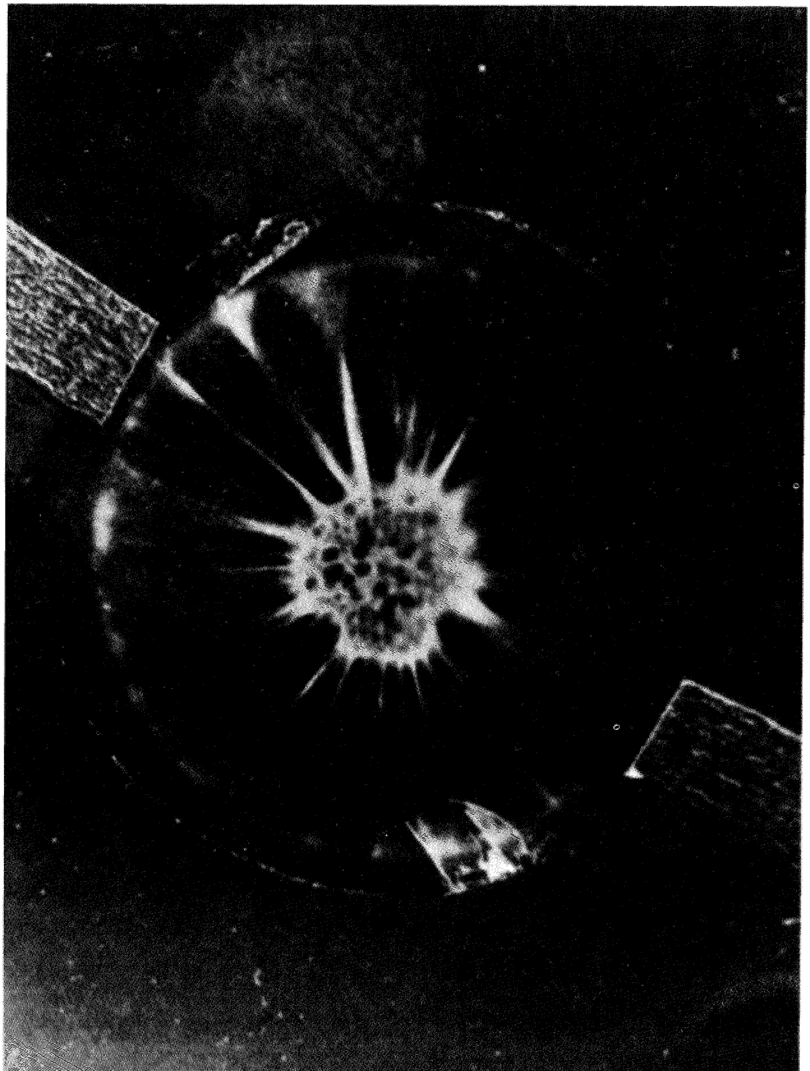


Fig. 2. End-on recording of the plasma gun penumbra.

of the discharge is the rapid collapse of the azimuthally-symmetric current sheath toward the axis to form the plasma focus.

2.1. Breakdown phase

The breakdown has a radial, striated light pattern with a definite multifilamentary structure (Figure 2). This structure, except for its obvious radial striation, appears cylindrically symmetric (Figures 1 and 3). As the current increases, the terminus of this visible pattern moves radially outward until it reaches the outer electrode. The current front motion is best described as an unpinch or inverse process; i.e., the $(\mathbf{j} \times \mathbf{B})_r$ body force is exerted outward between the center electrode surface and the plasma current sheath. During this inverse phase, the sheath remains stable because of the stability of the inverse pinch process (the B_θ lines are convex).

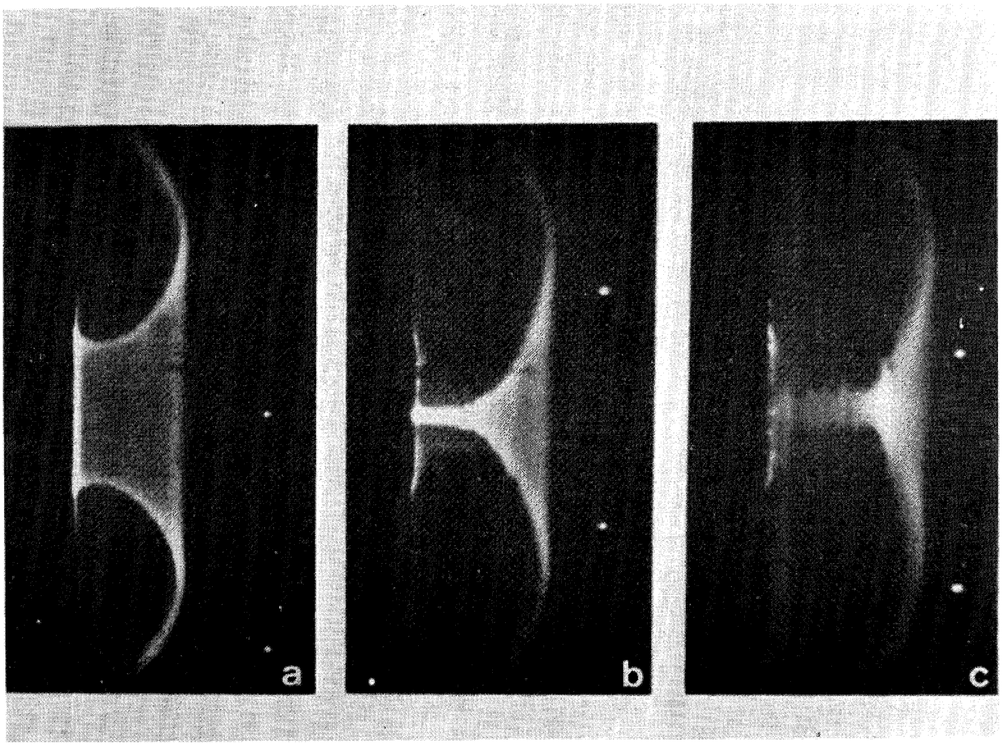


Fig. 3. Side-on image converter photographs of the plasma penumbra. (a) Before columnar collapse or pinch. (b) During pinch or focus.

The filamentary structures within the focus, rather than blending together, form a finite number of intense radial spokes ('spider legs'). These spokes appear to retain their identity throughout the acceleration phase and finally coalesce or focus on axis beyond the end of the center electrode, forming a thin circular annulus (Figure 2).

2.2. Acceleration phase

The current sheath across the annulus Δr is not planar, but canted backward from the anode to the cathode owing to the radial dependence of the magnetic pressure gradient. The total acceleration force $\mathbf{j} \times \mathbf{B}$ acting perpendicular to the current boundary leads to radial and axial motion. The $(\mathbf{j} \times \mathbf{B})_r$ radial component is outward and forces the current sheath to be annular at the outer electrode. The axial force $(\mathbf{j} \times \mathbf{B})_z$ varies as $1/r^2$ across the annulus and leads to higher sheath velocities near the surface of the center electrode. In the laboratory, fast image converter photographs distinctly show a parabolic current front. Owing to the parabolic current boundary, plasma flows centrifugally outward from anode to cathode along the current bounday as the acceleration continues. Plasma is seen to progress radially outward and beyond the outer electrode diameter as the current front accelerates downstream.

The overall time for plasma sheath acceleration to the end of the center electrode is related to the discharge potential across the annulus and the mass density of the background gas. For the case in which the current sheath pushes the gas ahead of the sheath, the sheath velocity is

$$v_s = \left[\frac{c^2 E^2}{4\pi\rho_0}\right]^{1/4}, \tag{1}$$

where E and ρ_0 are the discharge field strength and initial mass density, respectively (c.g.s. units).

2.3. Collapse phase

The third phase encompasses the rapid convergence of the axisymmetric current sheath to the axis and the conversion of stored magnetic energy to plasma energy in the focus. The r, z convergence is due to the $\mathbf{j} \times \mathbf{B}$ pinch force. With this configuration, there is no equilibrium along the axis; hence, the plasma may readily escape axially in either direction. By the very nature of the convergence, much of the gas that the sheath encounters during collapse is ejected downstream and lost. The gas trapped in the focus is estimated as $\sim 10\%$ of that originally present.

The pinch effect is perhaps the most efficient way of heating and compressing a plasma. As the pinch induced implosion velocity of the current boundary imparts the same velocity to both ions and electrons, and because of the ion-electron mass difference, most of the energy appears as kinetic energy of the ions. In pinch devices, the ions are preferentially heated.

2.4. Dynamic behaviour of the current sheath

The dynamic behaviour of the plasma current sheath can be analyzed using the measured time-dependent values of the voltage $V(t)$ across the electrodes, the tube current $I(t)$, and the sheath resistance $R(t)$.

The circuit equation representing the voltage across the electrodes ab in the equivalent

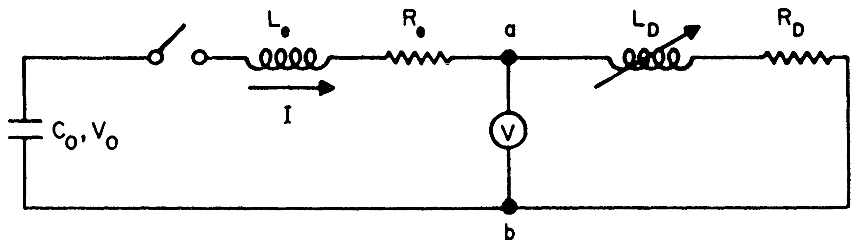


Fig. 4. Equivalent electrical circuit of the plasma gun discharge. The subscripts e and D denote external and discharge circuit elements, respectively.

electrical circuit (Figure 4) is

$$V(t) = (\mathrm{d}/\mathrm{d}t)\,[L_D(t)I(t)] + I(t)R_D(t)\,. \tag{2}$$

The part of the circuit to the left of ab in Figure 4 represents the electrical discharge and external circuit resistance R_e and external parasitic inductance L_e. The circuit to the left of ab represents the discharge inductance $L_D(t)$ and resistance $R_D(t)$.

By use of Equation (2), the following quantities are calculated.

(a) Discharge inductance:

$$L_D(t) = \int_0^{t'} [V(t) - I(t)R(t)]\,\mathrm{d}t/I(t)\,. \tag{3}$$

(b) Magnetic energy storage:

$$W_B(t) = \tfrac{1}{2}(L_e + L_D)I^2\,. \tag{4}$$

(c) Mechanical energy of the sheath:

$$W_s(t) = \frac{1}{2}\int_0^{t'} \dot{L}_D I^2\,\mathrm{d}t\,, \tag{5}$$

where $\mathrm{d}/\mathrm{d}t(L_D)$ is obtained from Equation (2).

(d) Pinch voltage during collapse:

$$V_p(t) = \dot{L}_D I + IR_D\,. \tag{6}$$

The instantaneous mass of the plasma sheath can be estimated using the momentum equation $\mathrm{d}(\rho\mathbf{v})/\mathrm{d}t = \mathbf{j} \times \mathbf{B}$,

$$m_0(t) = \left[10^{-7}\int_0^{t_c} I^2 \ln(r_o/r_i)\,\mathrm{d}t\right]v^{-1}\,, \tag{7}$$

where $v = v_s$ is the sheath velocity and r_o/r_i the ratio of the outer to inner electrode radii.

The upper integration limit t_c is the time to inner penumbra plasma collapse or pinch on axis.

3. The Plasma Gun Mechanism on Io

Plasma in Jupiter's magnetosphere injected from Io (the Io plasma torus) flows pas Io with a speed of about 57 km s^{-1}. The magnetic field from Jupiter at Io is 1900 nT. The $\mathbf{v} \times \mathbf{B}$ voltage induced across Io (3630 km) is, therefore, 400 kV, and approximately 10^6 A was observed to be flowing out of the satellite (Acuna *et al.*, 1981; Southwood *et al.*, 1980). It would seem plausible that the current would tend to concentrate in the volcanic plumes, which would give the current easy access to the highly conducting molten interior of Io. We would suppose that the crust, consisting of sulfer and frozen sulfer dioxide, would be a relatively poor conductor, thus directing the current to the volcanic vents. If we assume the available power (~ 0.4 TW) is equally divided between the four largest volcanic plumes, we have $\sim 10^{11}$ W of continuous power available for each volcanic plasma arc. This is roughly equal to the kinetic energy flux of material issuing from volcanic vent. A small fraction of this power can account for the faint auroral glow reported by Cook *et al.* (1981).

Figure 5 illustrates the basic geometry at hand. Viewed from the north Jovian and Ionian poles, Jupiter's dipole magnetic field is into the plane of the figure while the plasma flow within the torus is counter-clockwise toward Io. With this orientation, the

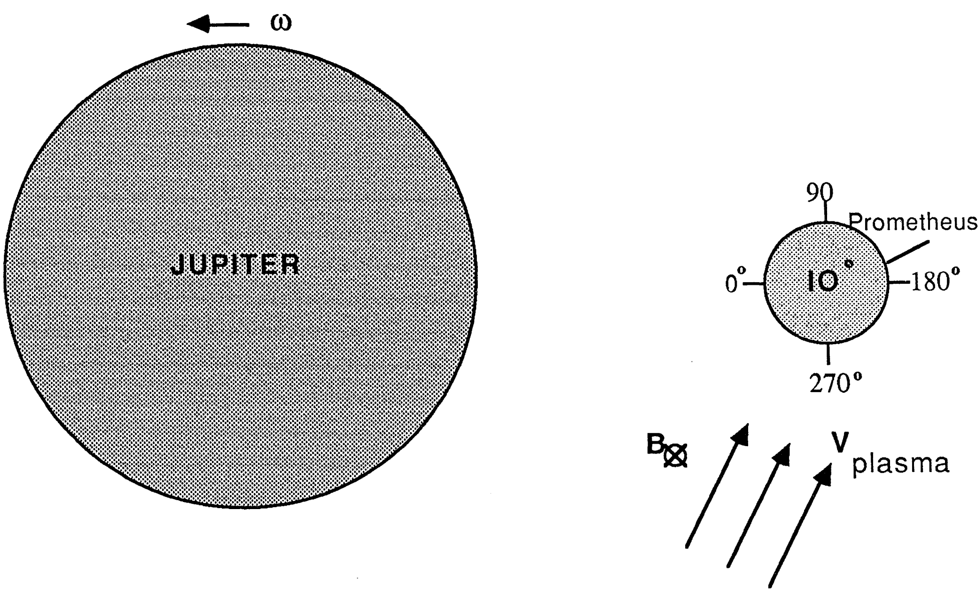


Fig. 5. The Jupiter-Io system (north pole view). The rotational period of Jupiter is approximately 9^h50^m. The orbital (synchronous) period of Io is 1.770 days. The relative velocity of the plasma torus flowing past Io is 57 km s^{-1}. (Diagram is not to scale.)

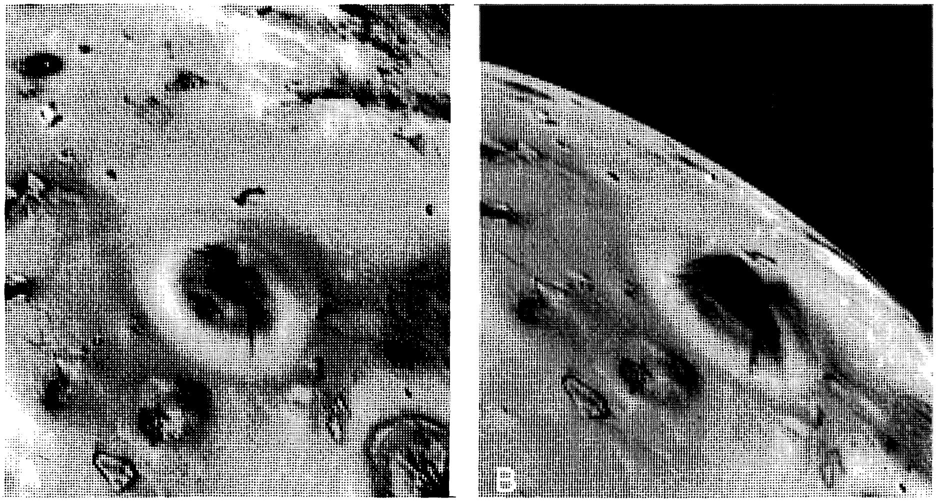


Fig. 6. Voyager 1 oblique views of Prometheus's plume. The left-hand image (A) was taken 2.3 hours before the right-hand image (B).

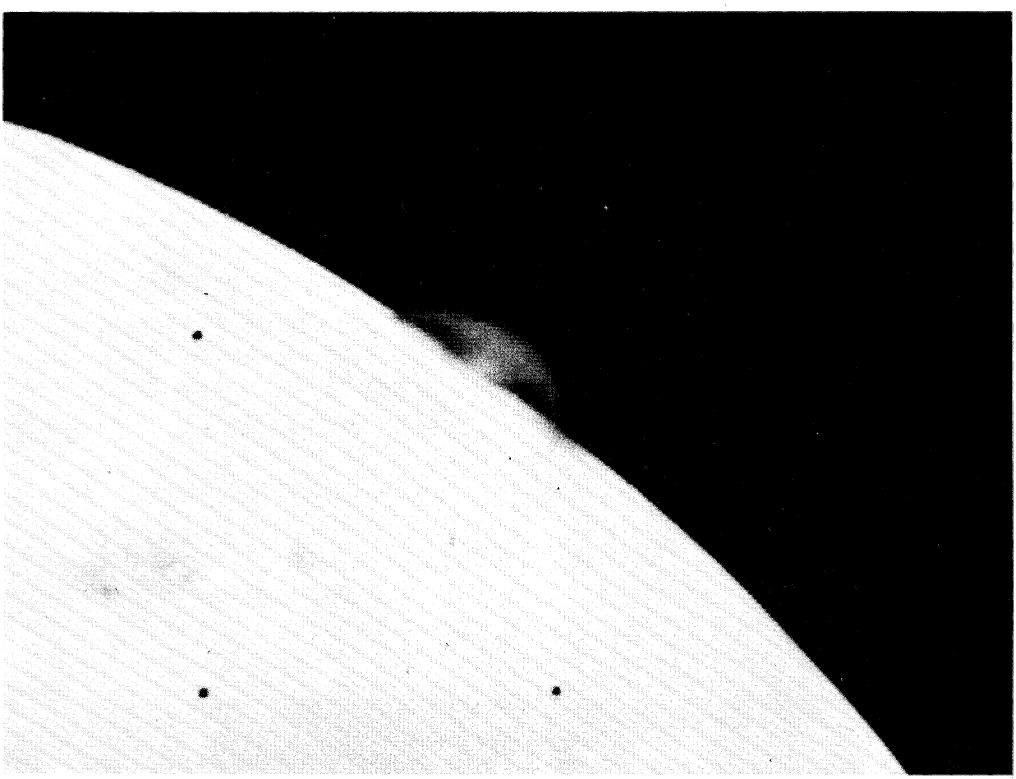


Fig. 7. Horizon view of Prometheus's plume.

electric field is directed from Jupiter to Io. Volcanic activity on Io generally occurs within an equatorial band of $\pm 30°$ latitude.

Figures 6 and 7 are photographs of the particularly well-developed volcanic plume Prometheus (latitude, $-2.9°$; longitude, $153°$) taken at two different aspect angles by Voyager 1. The height and width of this plume is 77 km and 272 km, respectively (Strom and Schneider, 1982) while the vent velocity of the gaseous material ejected is 0.49 km s^{-1}. The current flow is outward from Prometheus, i.e., the vent of Prometheus is an anode. To relate these pictures to the plasma-arc process described in Section 2, we must explicitly assume that the fine particulate matter that makes the volcanic plumes visible is entrained by the plasma. Thus, as the plasma moves to form filaments, we assume that the plasma carries with it the small solid particles.

While an exact calculation of the breakdown field associated with a volcanic arc discharge requires precise knowledge about the region where the breakdown occurs*, an estimate can be made in the following way. For air, nitrogen, freon, and sulfur hexafluoride, for a sufficiently large separation between anode and cathode electrodes, the breakdown field can be expressed (Miller, 1982) in the form

$$E_{br}(\text{kV cm}^{-1}) = 24.6pF^{-1}\,, \tag{8}$$

where p is the pressure of the ambient gas in atmospheres and F represents a field enhancement factor of order unity that depends on the shapes of the anode and cathode. Applying Equation (8) to SO_2, the most common gas on Io, while setting $F = 1$ and taking an ambient pressure of the gas near the vapour-liquid transition (the triple point for SO_2 is 0.0163 atmospheres), yields $E_{br} = 0.4$ kV cm^{-1}. This value is to be compared to the breakdown field strength for lightning on Earth; 4.4 kV cm^{-1}.

Substituting $E = 0.4$ kV cm^{-1} and $\rho_0 = 2$ g cm^{-3} into Equation (1), we obtain a parabolic sheath velocity of $v_s = 0.893$ km s^{-1}. It is at this velocity that gas and plasma are pushed into Io's upper atmosphere by means of the arc discharge mechanism. The volcanoe effluent is concentrated into a penumbra whose morphology differs from those of ballastic or shock models in two ways. The first is the radial striations resulting from the electromagnetic pinch and accretion of matter into filaments (Figures 2 and 6). The second is the termination of the penumbra on a narrow cathode annulus (Figures 3 and 7).

4. Summary

Volcanic plumes on the Jovian satellite Io may be a visible manifestation of a plasma-arc discharge phenomenon. As pointed out by Gold (1979), plasma arcs are expected because of $\sim 10^6$ A currents and 400 kV potentials generated by the flow past Io of a torus of relatively dense (2–3×10^3 cm^{-3}) magnetospheric plasma.

In this paper, the penumbra of current sheath produced in an arc discharge (Figures 2

* Comparison of the side-on penumbra morphology (Figures 1 and 3) to the side-on plume morphology (Figure 7) suggests that the location of the electrical discharge may be well below the surface of Io.

and 3) has been compared to photographs of the plume Prometheus (Figures 6 and 7). Both penumbra and plume are found to share a common morphology, i.e., a parabolic sheath profile, filamentation of matter within the sheath, and the termination of the sheath onto a rather thin annular ring. Moreover, the effluent ejection velocity as calculated from an expression for the sheath velocity in a plasma gun (0.893 km s^{-1}) is close to that observed for Prometheus, 0.49 km s^{-1}.

Motivated by a theory of the origin of the planets and satellites, H. Alfvén, in the late 1950's and early 1960's, directed a program of research on the physics of the plasma gun: the progenitor of the plasma focus discharge (Filippov, 1962; Mather, 1971). The apparent filamentary penumbra on Io may be the first direct verification of the plasma gun arc discharge mechanism at work in the solar system.

Acknowledgements

We thank Torrence Johnson for several encouraging and helpful comments. This work was supported by the U.S. Department of Energy and the National Aeronautics and Space Administration.

References

Acuna, M. H., Neubauer, F. M., and Ness, N. F.: 1981, *J. Geophys. Res.* **86**, 8513.
Alfvén, H.: 1958, in *Proceedings of the 2nd Int. Conf. on Peaceful Uses of Atomic Energy*, United Nations, Geneva, Vol. 31, p. 3.
Alfvén, H.: 1960, *On the Origin of the Solar System*, Oxford University Press, New York (see also *Rev. Mod. Phys.* **32**, 710).
Alfvén, H.: 1981, *Cosmic Plasma*, D. Reidel Publ. Co., Dordrecht, Holland, Ch. IV.
Alfvén, H. and Fälthammar, C.-G.: 1963, *Cosmic Electrodynamics*, Oxford University Press, London.
Alfvén, H. and Wilcox, J. M.: 1962, *Astrophys. J.* **136**, 1016.
Bostick, W. H., Nardi, V., and Prior, W.: 1975, *Ann. N.Y. Acad. Sci.* **251**, 2.
Cook, A. F., Shoemaker, E. M., Smith, B. A., Danielson, G. E., Johnson, T. V., and Synnott, S. P.: 1981, *Science* **211**, 1419.
Dattner, A. and Eninger, J.: 1964, *Phys. Fluids Suppl.* **7**, S41.
Filippov, N. V., Filippova, T. I., and Vinogradov, V. P.: 1962, *Nucl. Fusion Suppl.*, Pt. 2.
Fowler, T. K. and Coengens, F. H.: 1979, *9th Eur. Conf. Controlled Fusion and Plasma Physics*, Oxford, September 1979.
3rd Symp. on Physics and Technology of Compact Toroids in the Magnetic Fusion Energy Program, Los Alamos, 1980, Los Alamos National Laboratory Report LA-8700-C (1981).
Gerdin, G., Stygar, W., and Venneri, F.: 1981, *J. Appl. Phys.* **52**, 3269.
Gold, T.: 1979, *Science* **206**, 1071.
Goldenbaum, G. C., Irby, I. H., Chong, Y. P., and Hart, G.: 1979, *9th Eur. Conf. Controlled Fusion and Plasma Physics*, Oxford, September 1979.
Koloc, P. and Ogden, J.: 1979, *Symp. on Compact Toruses and Energetic Particle Injection*, Princeton Univ., December 1979.
Latham, R. V.: 1982, *Xth Int. Symp. on Discharge and Electrical Insulation in Vacuum*, Columbia, South Carolani, IEEE Cat. No. 82CH1826–7, p. 3.
Lindberg, L. and Jacobsen, C. T.: 1961, *Astrophys. J.* **133**, 1043.
Lindberg, L. and Jacobsen, C. T.: 1964, *Phys. Fluids Suppl.* **7**, S44.
Lindberg, L., Witalis, E., and Jacobsen, C. T.: 1960, *Nature* **185**, 452.
Mather, J. W.: 1971, in R. H. Lovberg and H. R. Griem (eds.), *Plasma Physics*, Academic Press, New York, Ch. 15.

Miller, R. B.: 1982, *Intense Charged Particle Beams*, Plenum, New York, Ch. 1.
Morabite, L. A., Synnott, S. P., Kupferman, P. N., and Collins, S. A.: 1979, *Science* **204**, 972.
Ness, N. F. *et al.*: 1979, *Science* **204**, 982.
Smith, B. A. *et al.*: 1979, *Science* **204**, 945.
Smith, B. A., Shoemaker, E. M., Kieffer, S. W., and Cook, A. F.: 1979, *Nature* **280**, 738.
Southwood, D. J., Kivelson, M. G., Walker, R. J., and Slavin, J. A.: 1980, *J. Geophys. Res.* **85**, 5959.
Strom, R. G. and Schneider, N. M.: 1982, in D. Morrison (ed.), *Satellites of Jupiter*, Univ. Arizona Press, Tucson, Arizona, p. 598.
Strom, R. G., Terrile, R. J., Masursky, H., and Hansen, C.: 1979, *Nature* **280**, 733.
Wilcox, J. M., Pugh, E., Dattner, A., and Eninger, J.: 1964, *Phys. Fluids Suppl.* **7**, S51.

IN SITU ACCELERATION AND GRADIENTS OF CHARGED PARTICLES IN THE OUTER SOLAR SYSTEM OBSERVED BY THE VOYAGER SPACECRAFT*

S. M. KRIMIGIS

Applied Physics Laboratory, The Johns Hopkins University, Laurel, Maryland, U.S.A.

and

D. VENKATESAN

Department of Physics, University of Calgary, Alberta, Canada

(Received 2 September, 1987)

Abstract. The probable connection between cosmic rays and the electromagnetic state of the interplanetary medium was recognized by Hannes Alfvén as early as 1949 (Alfvén, 1949, 1950); he pointed out that the properties of cosmic rays necessitate a mechanism, external to Earth but within the solar system, capable of accelerating particles to extremely high energies. In advocating the view of local origin for part of the cosmic-ray spectrum, Alfvén and his colleagues developed a very general type of acceleration mechanism called 'magnetic pumping'. The unique data set of the two Voyagers extends over an entire decade (1977–1987) and is most suitable to explore the problem of acceleration of charged particles in the heliosphere. The energy coverage of the Low Energy Charged Particle (LECP) experiment covers the range ~ 30 keV to several hundred MeV for ions and ~ 22 keV to several MeV for electrons. Selected observations of interplanetary acceleration events from ~ 1 to ~ 25 AU are presented and reviewed. These show frequent acceleration of ions to several tens of MeV in association with shocks; highest energies ($\gtrsim 220$ MeV oxygen) were measured in the near-perpendicular ($\theta_{Bn} \simeq 87.5°$) shock of January 5, 1978 at ~ 1.9 AU, where electron acceleration was also observed. Examples of ion acceleration in association with corotating interaction regions are presented and discussed. It is shown that shock structures have profound effects on high-energy (≥ 70 MeV) cosmic rays, especially during solar minimum, when a negative latitudinal gradient was observed after early 1985 at all energies from ~ 70 MeV down to ~ 30 keV. By early 1987, most shock acceleration activity in the outer heliosphere (~ 25 to 30 AU) had ceased both in the ecliptic (Voyager-2) and at higher ($\sim 30°$) ecliptic latitudes (Voyager-1). The totality of observations demonstrate that local acceleration to a few hundred MeV, and as high as a few GeV is continually present throughout the heliosphere. It should be noted that in 1954 when Alfvén suggested local acceleration and containment of cosmic rays within the solar system, no one treated his suggestion seriously, at any energy. The observations reviewed in this paper illustrate once more Alfvén's remarkable prescience and demonstrate how unwise it is to dismiss his ideas.

1. Introduction

Since the discovery of ionizing radiation just under eight decades ago, the subject of cosmic rays has developed into a distinct discipline. About forty years of study of cosmic-ray intensity variations by ground-based detectors has revealed the existence of such features as the solar flare increase, the Forbush decrease, and the daily variation of cosmic-ray intensity, all of which have demonstrated the solar control of the interplanetary electro-magnetic state through the flowing solar wind plasma with the

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 463–486.

imbedded solar magnetic field. There is general agreement at present that there are two components of cosmic rays, the galactic and the solar.

The use of artificial satellites and space probes since the IGY (1957) has transformed what was an exploration of the interplanetary medium by ground-based detectors into an *in situ* exploration of space. Conjectures about the interplanetary medium can now be tested with realistic measurements. In contrast to the so-called 'ground level' events which could only measure high-energy particles ($\gtrsim 1$ GeV), it has become clear that the near-geophysical environment, for example, is flooded with low-energy (a few MeV) particles, which typically persist for several days after a solar event. At still lower (~ 300 keV) energies, measurable ($\gtrsim 10^{-2}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}\,\mathrm{keV}^{-1}$) intensities are always present in the interplanetary medium near 1 AU.

The connection between cosmic rays and magnetic fields in the solar system had been recognized early by Hannes Alfvén (e.g., Alfvén, 1954a) who pointed out that the spectral properties of cosmic rays require that there be a mechanism outside the Earth which is able to accelerate particles to extremely high energies. Furthermore, Alfvén (1954b) suggested the possibility of local origin of cosmic rays. The question 'Is cosmic radiation a galactic or local phenomenon?' was again addressed by Hannes Alfvén in more detail (1959). In dealing with this question he suggested an acceleration mechanism of a very general type and which has come to be known as 'magnetic pumping'. It is a combination of two simple effects – the betatron acceleration and the scattering of particles from small inhomogeneities of the magnetic field. In essence, if charged particles are injected at a certain point, they are accelerated but at the same time diffuse away from the point of injection. The process conserves the first adiabatic invariant ($\mu = E_\perp/B$ where $E_\perp$ is perpendicular energy and B is the magnetic field) but violates the third (more details may be found in Fälthammar, 1963). It was concluded by Alfvén at the time that the observed power-law momentum spectrum of cosmic rays is compatible with the proposed acceleration mechanism, and that cosmic radiation below $\sim 10^{14}$ eV is generated near the Sun or its environment. He made the comment that the highest portion of the spectrum – namely $> 10^{14}$ eV – may derive from other sources such as supernovae, magnetically variable stars, double stars, colliding magnetized clouds, etc.

Observations by ground-based detectors have stimulated ideas about processes in the interplanetary medium, including the possibility of *in situ* acceleration. The availability of space-based measurements have allowed several hypotheses to be tested directly near 1 AU. During the past decade, the Pioneer and Voyager spacecraft have initiated an exploration of the three-dimensional heliosphere, whose boundary is currently believed to be at ~ 100 to 150 AU. The study of time variations of cosmic rays has thus transformed into one of space and time. It is the purpose of this manuscript to study the acceleration of charged particles with specific reference to the outer regions of the solar system. We would also like to explore how much these processes can contribute to the observed cosmic-ray intensity.

A number of different acceleration mechanisms have been proposed in order to account for particle enhancements observed in the heliosphere. Most of these involve

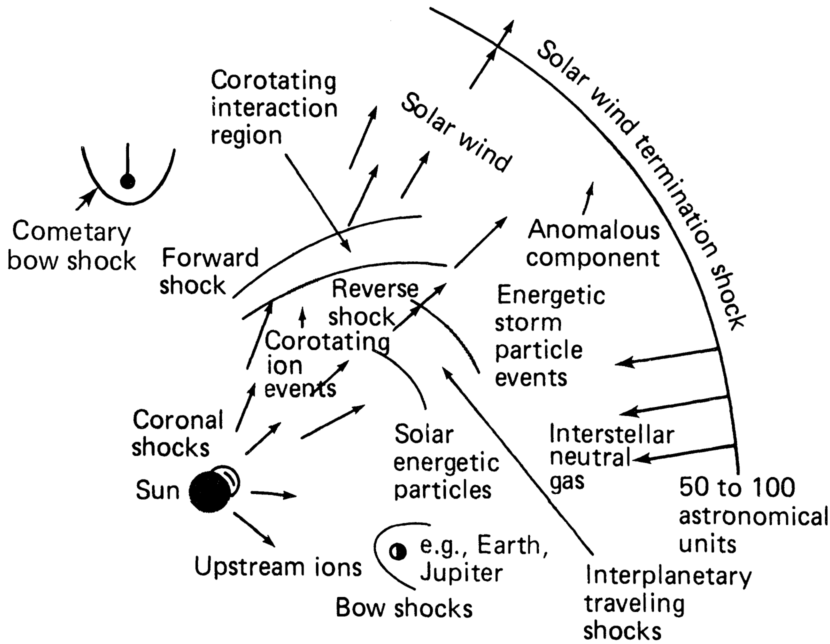


Fig. 1. A schematic view of shock acceleration regions both observed and hypothesized at the Sun and in the interplanetary medium (adapted from Rosner *et al.*, 1984).

some form of shock acceleration as most likely to account for the observations. Figure 1 (not drawn to scale) (Rosner *et al.*, 1984) provides a schematic representation of diverse acceleration processes observed in the heliosphere and the most likely sites of acceleration. Shocks associated with each region of acceleration are bow shocks associated with comets and planetary magnetospheres, solar flare-associated shocks traveling outwards in the heliosphere, coronal shocks, forward-reverse shocks in Corotating Interaction Regions (CIR), and the solar wind termination shock. The scale size, speed, strength, and configuration of a particular shock are believed to determine the maximum energy of the accelerated ions, while the duration of the intensity increases observed at any one point depends on the interplanetary propagation processes and the configuration of the magnetic field between the observation point and the acceleration region.

The data to be reviewed in this paper are principally those obtained by the Voyager-1 and -2 spacecraft, which were launched in 1977 towards encounters with the planets Jupiter, Saturn, Uranus, and Neptune. The trajectories of the two spacecraft are shown in Figure 2. It should be noted that Voyager-1 after encounter with Saturn in early November 1900 proceeded in an orbit taking it to higher heliolatitudes; currently (July 1987) it has reached a heliolatitude of $\sim 29°$ N and a radial distance of ~ 31 AU. Thus this pair of spacecraft provides a unique data set corresponding to in-ecliptic and out-of-ecliptic radial and temporal variations. Measurements over a broad energy range with adequate time and angular coverage are necessary to address the questions outlined in the preceding paragraphs. The Low Energy Charged Particle (LECP) instrument

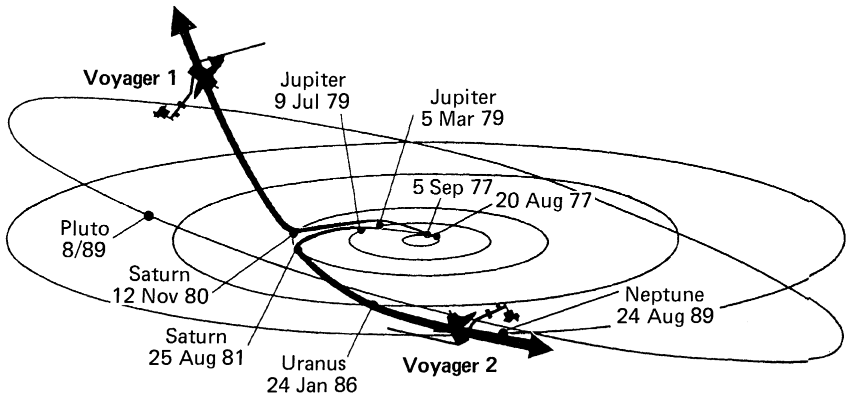


Fig. 2. The trajectories of the Voyager-1 and -2 spacecraft in the solar system. The orbits of the planets are shown to scale. After the Saturn encounter, Voyager-1 was directed above the ecliptic plane and is escaping the solar system at an ecliptic latitude of $\sim 35°$. Voyager-2 is expected to encounter Neptune on 24 August, 1989 and will be deflected south of the ecliptic at a latitude of $\sim 47°$.

onboard the two Voyager spacecraft covers the energy range from ~ 30 keV for ions and ~ 22 keV for electrons to several tens of MeV, and is capable of both high time-resolution (~ 1.2 s) and broad energy and compositional coverage (for details see Krimigis *et al.*, 1977). The instrument has performed well over the past decade and has yielded a large, well characterized and valuable set of observations, some of which have been analyzed in detail. Several examples of these data will be presented in this paper.

2. Heliospheric Shocks

2.1. The shock mechanism

The many shocks depicted in Figure 1 are varying manifestations of a single physical phenomenon described by relatively few parameters. The basic geometry of a shock (Figure 3(a)) is shown in the 'shock frame' where the shock is stationary and the upstream plasma flows towards it, parallel to the shock normal (along the x-axis) at a speed $\mathbf{V}$. The most important parameters in the context of particle acceleration are the shock strength $|B_2/B_1|$, the angle θ_1 between the upstream magnetic field and the shock normal, generally designated as θ_{Bn}, and the shock speed relative to the upstream plasma rest frame $\mathbf{V}_1$, here.

The physical basis for particle motion through a collisionless shock is the Lorentz force on the particle and Newton's second law – i.e.,

$$\mathbf{F} = Ze(\mathbf{E} + \mathbf{v} \times \mathbf{B}/c) = \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} ,$$

where $\mathbf{v}$ is the particle velocity, $\mathbf{p}$ its momentum, and $\mathbf{E}$ and $\mathbf{B}$ are the local electric and magnetic fields with $\mathbf{E} = -\mathbf{V} \times \mathbf{B}/c$. Integration of this equation in specified fields

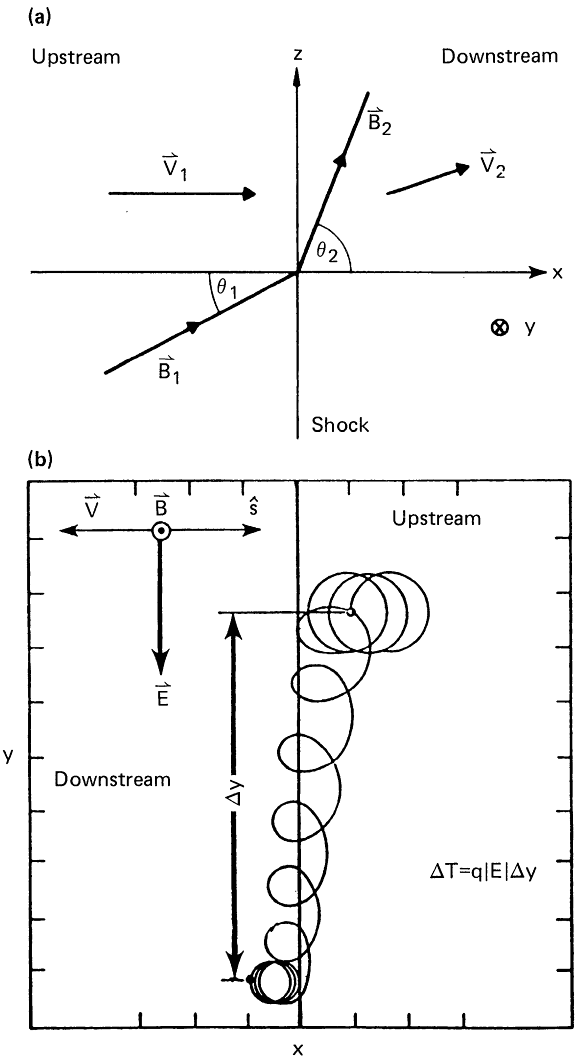


Fig. 3. (a) Geometrical relationships of velocity and magnetic field in a fast magnetosonic shock; the y-axis is pointing into the paper. (b) Sample trajectory in the $x-y$ plane of an ion injected into a perpendicular shock some 3 gyroradii upstream, drifting along the shock in the direction of the electric field and thereby gaining energy by the indicated amount, and exiting into the downstream region (adapted from Pesses, 1981).

determines the velocity and position changes of a particle interacting with the shock. A particularly simple case is shown in Figure 3(b) where $\theta_{Bn} = 90°$ and a particle is injected a few gyroradii upstream of the shock and undergoes drift motion along the shock surface in the y-direction. The energy gain while the particle is interacting with the shock, and before it is transmitted downstream, is given by $\Delta T = Ze\,|\mathbf{E}|\,\Delta y$.

Shocks occurring in the heliosphere are generally classified into two broad categories:

quasi-parallel shocks, i.e., those where the angle θ_{Bn} between the shock normal and the upstream magnetic field is $\lesssim 45°$ and quasi-perpendicular shocks where $\theta_{Bn} \gtrsim 45°$. In general, quasi-parallel shocks tend to display more magnetic and electric field fluctuations than quasi-perpendicular shocks (Bavassano *et al.*, 1987). Quasi-parallel shocks mostly occur in the inner heliosphere, generally at distances $\lesssim 1.5$ AU, whereas quasi-perpendicular shocks occur throughout the entire heliosphere, but most often at distances > 2 AU, where on the average the angle between the interplanetary magnetic field and the shock normal is closer to 90°. It is very important, however, to realize that the instantaneous geometry between the upstream magnetic field and the shock normal can vary over large ranges in angle as the shock propagates from the inner to the outer heliosphere; in fact a shock can be instantaneously quasi-parallel in some parts of the outer heliosphere even though, on the average, such shocks are generally perpendicular (Sarris *et al.*, 1987b). Thus it is essential to measure the shock parameters *in situ* before a clear determination can be made as to the nature of the shock.

2.2. Shock of January 6, 1978

The heliospheric shocks depicted schematically in Figure 1 are of both varieties but, in line with the classification given above, are on the average quasi-perpendicular. Such shocks have been observed to accelerate particles to the highest observed energies so far and do so in very short (a few seconds to tens of seconds) time intervals. An example of a propagating interplanetary shock with a near-perpendicular geometry is shown in Figure 4 (Krimigis and Sarris, 1985). This shock wave which passed Voyager-2 on January 6, 1987 at $\sim$00 : 01 : 30 UT had its origin in a solar flare of importance $2N$ (21° S, 06° E), on January 1, 1978 with Hα maximum at 21 : 45 UT (*Solar Geophysical Data*). The following are the characteristics of the quasi-perpendicular shock in question: $\theta_{Bn} \sim 87.5°$; local shock speed $V = 588$ km s^{-1}; Alfvén mach number $M_A = 3.4$; the ratio of downstream to upstream field $B_2/B_1 = 2.4$; upstream solar wind speed with respect to the shock front $V_1 = 200$ km s^{-1} (for further details see Sarris and Krimigis, 1985).

Figure 4(a) shows the hourly averages of the responses of selected channels from the LECP instrument during January 1–7, 1978; the dashed line marks the time of shock passage. It is seen that the shock-associated energetic particle enhancement over the ambient solar particle intensities is a factor of $\sim$20 at the lowest energies (top curve), in excess of a factor $\sim 10^2$ for protons in the range of 4.4–11.4 MeV (second curve), and even larger factors for He- and Fe-group nuclei. Note that the higher energy He nuclei have total energies of 88–120 MeV while those in the Fe-group have total energies of 112–672 MeV (lowest curve). In addition, there is a clear signature of acceleration for energetic electrons in the 61–112 keV energy range which is mostly confined to the shock. It is important to note that full pulse-height analysis spectra obtained by the LECP instrument shows all species to be populated from protons through iron including an oxygen track that extends in total energy to $\sim$220 MeV (Sarris and Krimigis, 1985). This is the highest energy, shock-accelerated ion ever observed on a spacecraft.

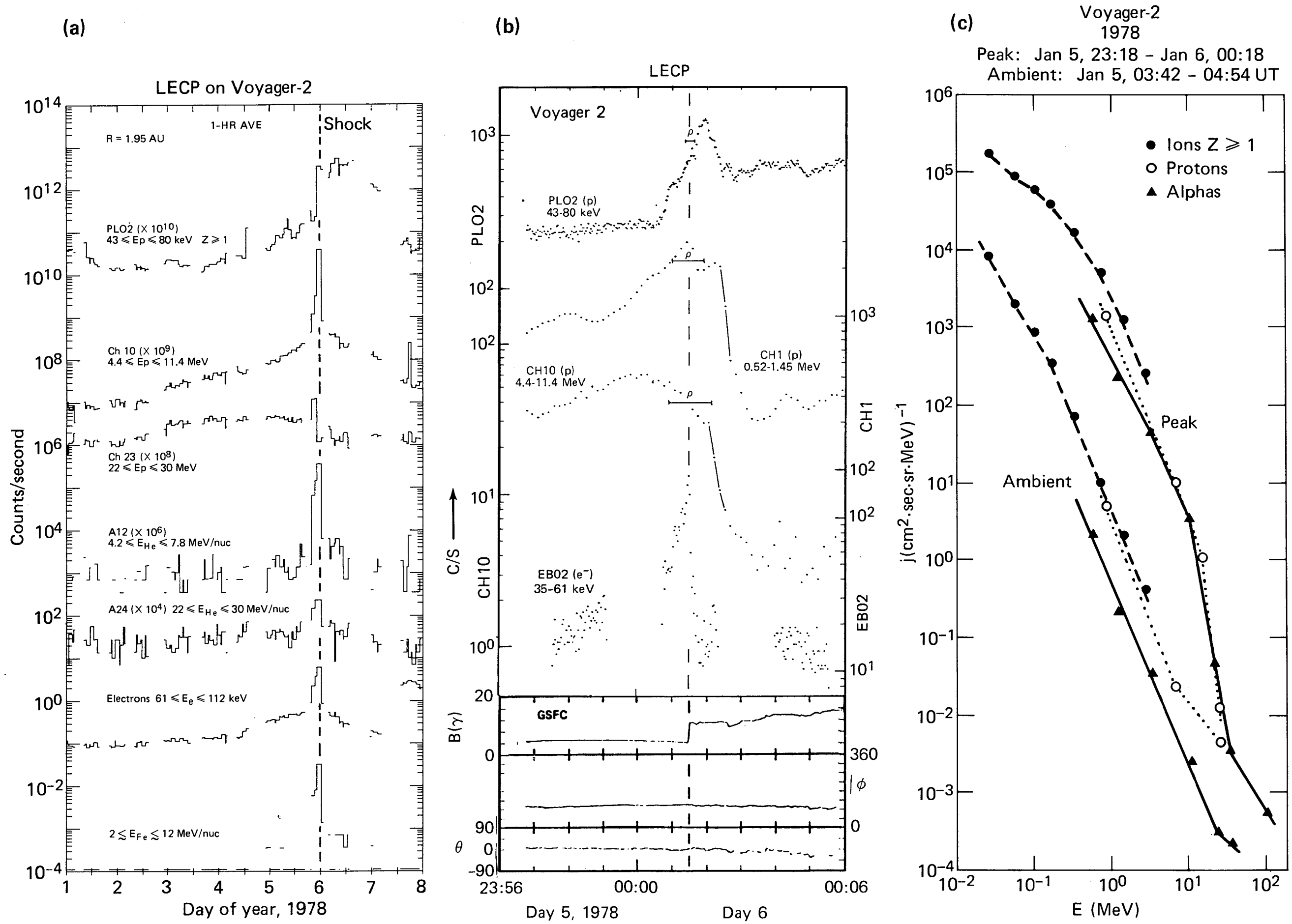


Fig. 4. (a) Time profile of selected channels from the solar event of 1 January, 1978 as observed by the Voyager spacecraft ~1.9 AU. The time of the shock is indicated by the vertical line. Note the appearance of Fe-group ions in association with the shock just before shock passage on 5 January. (b) Detail of the time profile of intensity and magnetic field surrounding shock passage. The horizontal bar labeled ρ shows the particle gyroradius for the particular energy channel. Time resolution for measurements at the lowest energy channel were 1.2 s. (c) The energy spectrum of energetic ions for both the ambient population (lower two curves) and the population around the time of shock passage. Enhancements occurred at all energies shown in the spectrum.

It is important to examine the details of the acceleration process and its evolution close to the shock, together with the magnetic field data. This is shown in detail in Figure 4(b) which includes in the bottom panel the magnetic field parameters for a period of ~ 10 min centered on the shock passage. The magnetic field measurements (time resolution of 1.92 s) show a laminar shock signature, i.e., absence of turbulence at any significant level in the frequency range from 0.2 to 10^{-3} Hz. The top panel shows the evolution of the intensity profiles, with the lowest energy peaking immediately downstream of the shock within ~ 2 gyroradii, as indicated by the horizontal bar labeled ρ. The next curve, at a factor approximately ten higher in energy, peaks exactly during shock passage, with a sharp (factor of ~ 10) drop in the intensity within a gyroradius after shock passage. The third curve, at a factor of ten higher in energy still, shows the peak intensity occurring ~ 1 gyroradius upstream of the shock, with a sharp intensity drop immediately downstream; the fourth curve shows that energetic electrons are essentially accelerated at the shock. Note that the electron gyroradius is small (equivalent to ~ 0.8 s) even on this time-scale (2.4 s), comparable to the size of the point. The overall picture here seems to indicate that most of the acceleration is actually taking place within a gyroradius of the shock, and that the evolution of the intensity profile is exactly as predicted by the shock drift acceleration mechanism (Armstrong *et al.*, 1977) where the most energetic ions should peak upstream of the shock (reflected) while the less energetic ones peak downstream (Sarris *et al.*, 1976).

A more quantitative assessment of the intensity enhancement as a function of energy is shown in the energy spectra of Figure 4(c). Here the pre-shock spectrum of protons, He, and $Z \geq 1$ ions is plotted together with the spectrum at the time surrounding shock passage. It is evident that the enhancement occurs at all energies, but is most pronounced in the range of a few hundred keV to a few MeV, as expected from theory, where these particles ride with the shock for a longer time. Note the increase in intensity of He at energies $E \gtrsim 100$ MeV. By using the ambient and the shock spectra it is possible to estimate the average energy gain per particle (efficiency). This numbers turns out to be $\sim 10^2$ to 10^3 higher than that in a strong, supercritical, quasi-parallel shock (e.g., Kennel *et al.*, 1986).

An important aspect of any propagating shock is the amount of energy transfer from the shock to accelerated ions. For the event shown in Figure 4 the particle energy density peaks in the range from ~ 200 keV to ~ 2 MeV, with an integrated density of $\sim 3.7 \times 10^{-10}$ erg cm^{-3} (Sarris and Krimigis, 1985). The energy density of the upstream magnetic field was $\sim 10^{-10}$ erg cm^{-3}, so that the plasma β (i.e., the particle energy density divided by $B^2/8\pi$) was > 4, since the contribution to particle energy density at < 17 keV could easily increase this number. The kinetic energy density of the upstream plasma in the shock frame of reference was $\sim 9.2 \times 10^{-10}$ erg cm^{-3}, i.e., the accelerated ions constituted a significant fraction ($\sim 40\%$) of the shock energy. Of course, it is recognized that the shock probably has a very large spatial extent and it is unlikely that such efficient acceleration is taking place over the extended shock front. Such energy loss to accelerated ions would have profound effects on the overall shock evolution and structure, if it obtained over the entire shock surface.

In exploring the high-energy extent of shock-accelerated ions, it is important to note that there have been at least two ground-level events (GLE's) observed with neutron monitors which have been interpreted as arising from shock acceleration processes in the interplanetary medium (Pomerantz and Duggal, 1974). One of the events occurred following the series of flares in early August, 1972 which produced successive shocks with velocities in the range ~ 1200 to ~ 2800 km s^{-1}. The GLE event was observed on August 4 and was characterized by a rather steep rigidity spectrum ($\gamma = 6.6 \pm 0.5$ in a power law), suggesting a different origin for the protons than a flare-associated enhancement (Pomerantz and Duggal, 1974). Maximum energies observed were $\gtrsim 1.3 \times 10^9$ eV (~ 2 GV) at the Swarthmore neutron monitor. Levy *et al.* (1976) proposed a detailed model to explain the event in terms of shock acceleration between two converging shocks, although other interpretations cannot be positively excluded (for a review see Dugal, 1979). Thus it appears that maximum energies attained in interplanetary shock-associated events may well exceed ~ 1 GeV.

2.3. Corotating interaction region (cir) shocks

As the radial distance from the Sun increases, shocks observed in the heliosphere are commonly associated with Corotating Interaction Regions (Smith and Wolfe, 1979). These regions are formed when a fast solar wind stream interacts with a pre-existing slow stream, thus creating a forward shock which is typically followed by a reverse shock, observed when a quiet region of the solar wind returns. The signatures of such forward-reverse shock pairs are readily recognizable as increases in the magnetic field, solar wind velocity, density, and temperature at the forward shock, and a decrease in the magnetic field, increase in the solar wind velocity, and decrease in density and temperature at the reverse shock.

An example of a CIR observed by Voyager at ~ 3.9 AU is shown in Figure 5. The top panel shows the solar wind velocity and interplanetary magnetic field (IMF) magnitude while the other three panels show several energetic particle parameters. On day 263 the magnetic field increased abruptly at the forward shock together with a sharp rise in the solar wind velocity, followed two days later by a decrease in the magnetic field and a further increase in the velocity. As indicated by the fluctuations in the magnitude of the magnetic field, the region between the two shock was quite turbulent. The intensity profiles of energetic ions in the third panel (Decker *et al.*, 1981) show a general enhancement of intensities prior to the forward shock, most evident at the higher energies. The enhancement of low-energy (~ 43 keV) ions occurred mostly at the time of shock passage and the decay at these energies was substantially slower than for higher energy protons and α-particles. The intensity peak on day 265 is completely detached from the reverse shock. This is likely due to an abrupt change in the IMF direction (not shown here) which pointed toward ecliptic south; this change, coupled with the strong anisotropies in the particle distribution throughout this period (Decker *et al.*, 1981), may well have taken the highest intensities out of the detector field of view, which points generally in the ecliptic plane (Krimigis *et al.*, 1977).

The spectral indices of 80 to 215 keV particles, second panel from the top, are

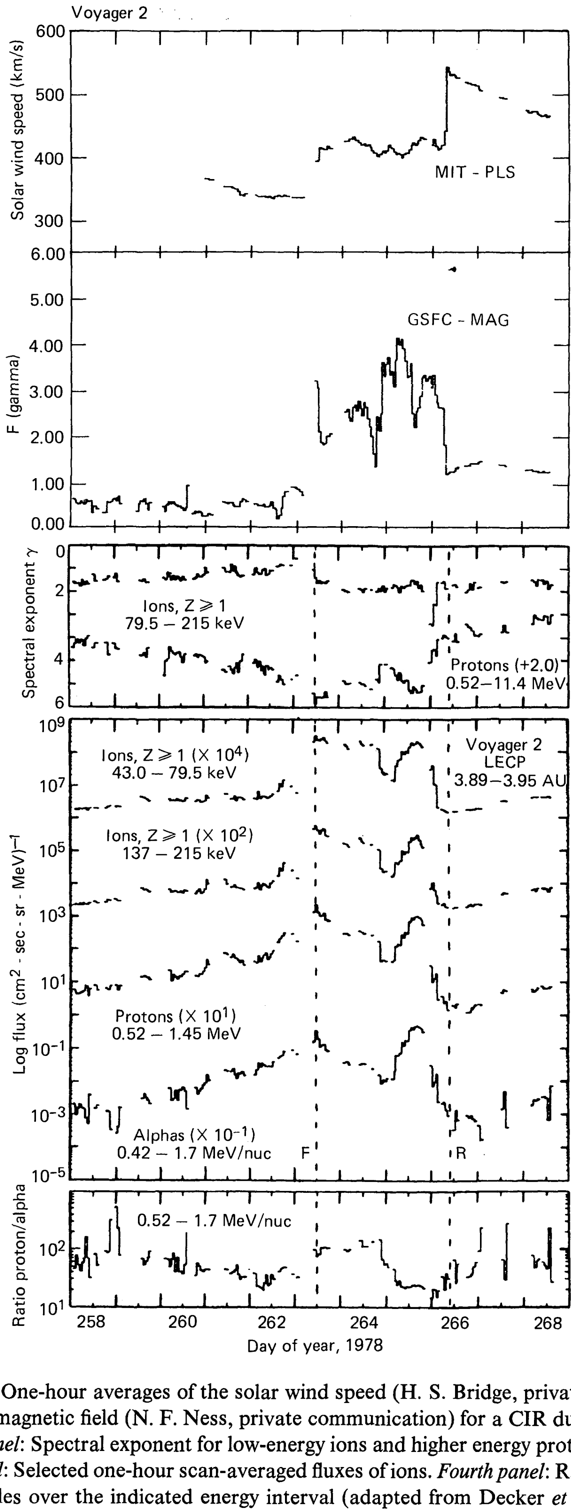


Fig. 5. *Top panel*: One-hour averages of the solar wind speed (H. S. Bridge, private communication) and the interplanetary magnetic field (N. F. Ness, private communication) for a CIR during the indicated time interval. *Second panel*: Spectral exponent for low-energy ions and higher energy protons over the same time interval. *Third panel*: Selected one-hour scan-averaged fluxes of ions. *Fourth panel*: Ratio of protons to alpha particles over the indicated energy interval (adapted from Decker *et al.*, 1981).

generally softer inside the forward shock and they gradually harden as the reverse shock is approached. The bottom panel shows that the proton-to-alpha ratio increased substantially across the forward shock but then decreased to the preshock level well before the reverse shock was approached. Another effect of CIR regions in the outer heliosphere, not explicitly shown in this figure, is the exclusion of relativistic electrons (most probably of Jovian origin) from these interactions regions (Smith and Wolfe, 1979).

The properties of interplanetary shocks relating to energetic particle acceleration are summarized in Table I. As has been indicated in our discussion so far, the maximum proton energy attained in quasi-parallel shock is $\lesssim 300$ keV (e.g., Kennel

TABLE I

Particle acceleration at interplanetary shocks

Property	Quasi-perpendicular	Quasi-parallel
Percent occurrence	$\sim 80\%$ at 1 AU $\sim 100\%$ in outer heliosphere	$\sim 20\%$ at 1 AU $\sim 0\%$ in outer heliosphere
Mach number/turbulence	$M_A \sim 1$ to ~ 4/generally absent	$M_A \sim 1$ to ~ 4/always present
Ion acceleration/ energy range	Observed for all M_A, $\sim$ few tens of keV to > 100 MeV	Observed for $M_A \gtrsim 3$, $\sim$ few tens of keV to $\lesssim 300$ keV
Electron acceleration	Observed for all M_A, ~ 1 keV to ~ 1 MeV	None reported
Relative efficiency $\left(\frac{\text{energy gain}}{\text{particle}}\right)$	$\sim 10^2$–10^3	1
Acceleration time	Tens of seconds	Tens of hours
Anisotropy	Large ($> 3:1$) anisotropies for ions, electrons	Significant ($> 2:1$) anisotropies for ions
Dominant acceleration mechanism	Shock-drift process, predicts spectra and large anisotropies of ion and electron events	First-order Fermi process, predicts spectra and weak anisotrpies of ions events; no electron acceleration

et al., 1986) even for supercritical shocks. By contrast, the maximum energy attained in quasi-perpendicular shocks is a factor $\sim 10^3$ larger, and the acceleration time is faster by a similar factor. Also, there are observations of electrons accelerated at near-perpendicular shocks, but none have been reported for quasi-parallel shocks. Even for quasi-perpendicular shocks, however, electrons are rarely accelerated to energies larger than a few keV (Potter, 1981). This is principally due to the large velocity of such electrons, relative to typical shocks speeds of $\sim 10^3$ km s^{-1}. Even a 2 keV electron, at a speed of 2.6×10^4 km s^{-1} is substantially faster than the shock, unless θ_{Bn} is close to 90° so that the field line intersection speed with the shock ($V_t = V_1 \sec \theta_{Bn}$) is a significant fraction of the particle speed (e.g., Decker, 1983). The large θ_{Bn} for the January 6, 1978 shock is the principal reason for the observation of electrons to such high ($\gtrsim 100$ keV) energies. In the quasi-parallel shock case high intersection speeds are clearly not attainable, hence, the absence of observations of accelerated electrons.

Finally it is important to point out in this context that most 'cosmic-ray' electrons are not cosmic but rather are accelerated within Jupiter's magnetosphere and escape upstream and downstream from the planet. Such electrons dominate the interplanetary electron spectrum up to ~ 40 MeV and even retain the 10-hr periodicity of Jupiter's rotation (e.g., Pyle and Simpson, 1977). Thus local acceleration of electrons in the interplanetary medium does not contribute significantly to the ambient interplanetary population at energies $\gtrsim$ a few keV.

3. Cosmic-Ray Effects of Shocks and Gradients

Since the early observations by spacecraft of energetic ions in the interplanetary medium, there has been a phenomenological association between the arrival of the so-called energetic storm particles (ESP) (e.g., Bryant *et al.*, 1962) and a decrease in the cosmic-ray intensity which can last from a few hours to several days, called a Forbush decrease. It was later realized that ESP enhancements were typically associated with propagating interplanetary shocks and were observed to occur following major solar flare events. Since that time there has been significant progress in both the quantitative description and physical understanding of such decreases in cosmic-ray intensity both within the inner solar system as well as the outer heliosphere (Lockwood *et al.*, 1986; Webber *et al.*, 1986). More recently, it has been recognized that large Forbush decreases are associated with flare sites east of the Sun-spacecraft line (Sarris *et al.*, 1987a), which in turn produce quasi-perpendicular shocks (Sarris *et al.*, 1984).

An example of Forbush decreases and their association with shocks is given in Figure 6 (Burlaga *et al.*, 1984) and illustrates a number of points. The figure shows the measurement of high ($\gtrsim 75$ MeV nucl^{-1}) energy cosmic rays at 3 radial ranges in heliospheric distance at ~ 6.8 AU, ~ 8 AU, and ~ 22 AU obtained by the spacecraft Voyager-2, Voyager-1, and Pioneer-10, respectively, roughly over the same time period. The first panel in the figure shows successive decreases and subsequent recoveries in cosmic-ray intensity (labeled *A*, *B*, *C*, and *D*) which were observed at the distance indicated. Note that, in contrast to period *A* and *B*, no recovery in the cosmic-ray intensity is observed following decrease *C* on June 24, 1980 at Voyager-2. The count rates from similar detectors on Voyager-1, farther out in radial distance, follow a similar pattern except that they are displaced by a few days consistent with a propagation of the decrease from the inner to the outer heliosphere at a velocity similar to that of the solar wind (~ 500 km s^{-1}). The bottom panel shows a similar channel from Pioneer-10 at the distance of ~ 22 AU and it is again seen that the decreases observed at ~ 7 AU continue to propagate, with an appropriate delay, into the outer heliosphere. Thus the evidence from the simultaneous observations shows that the Forbush decrease first develops in the inner heliosphere and propagates outward with a velocity similar to that of the solar wind.

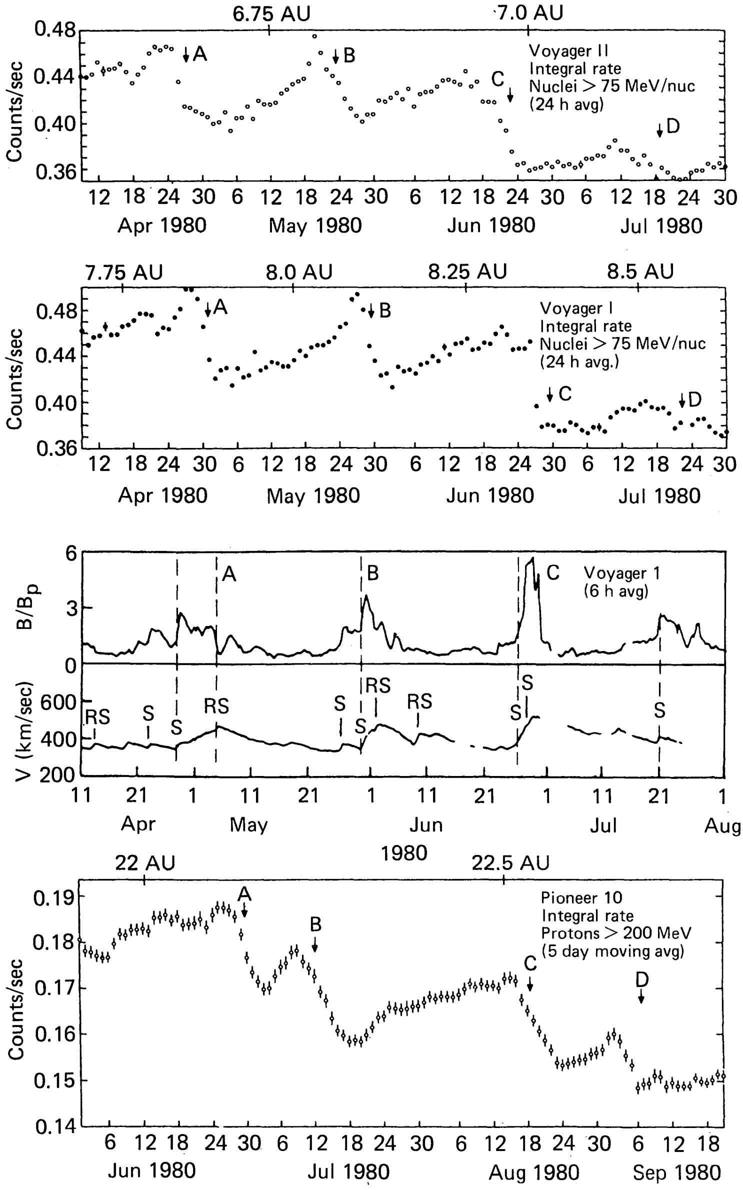


Fig. 6. A striking demonstration of the propagation of Forbush decreases from the inner to the outer heliosphere as recorded by the Voyager-2, Voyager-1, and Pioneer 10 spacecraft located at the progressively greater heliocentric radial distances. *Top panel*: The integral cosmic-ray count rate at Voyager-2. *Second panel*: Similar rate from Voyager-1. *Third panel*: The interplanetary magnetic field divided by the ideal Parker spiral field, together with the solar wind velocity over the same time interval as the second panel. *Bottom panel*: Integral count rate of cosmic rays for Pioneer-10 at the indicated energy in the outer heliosphere (adapted from Burlaga *et al.*, 1984).

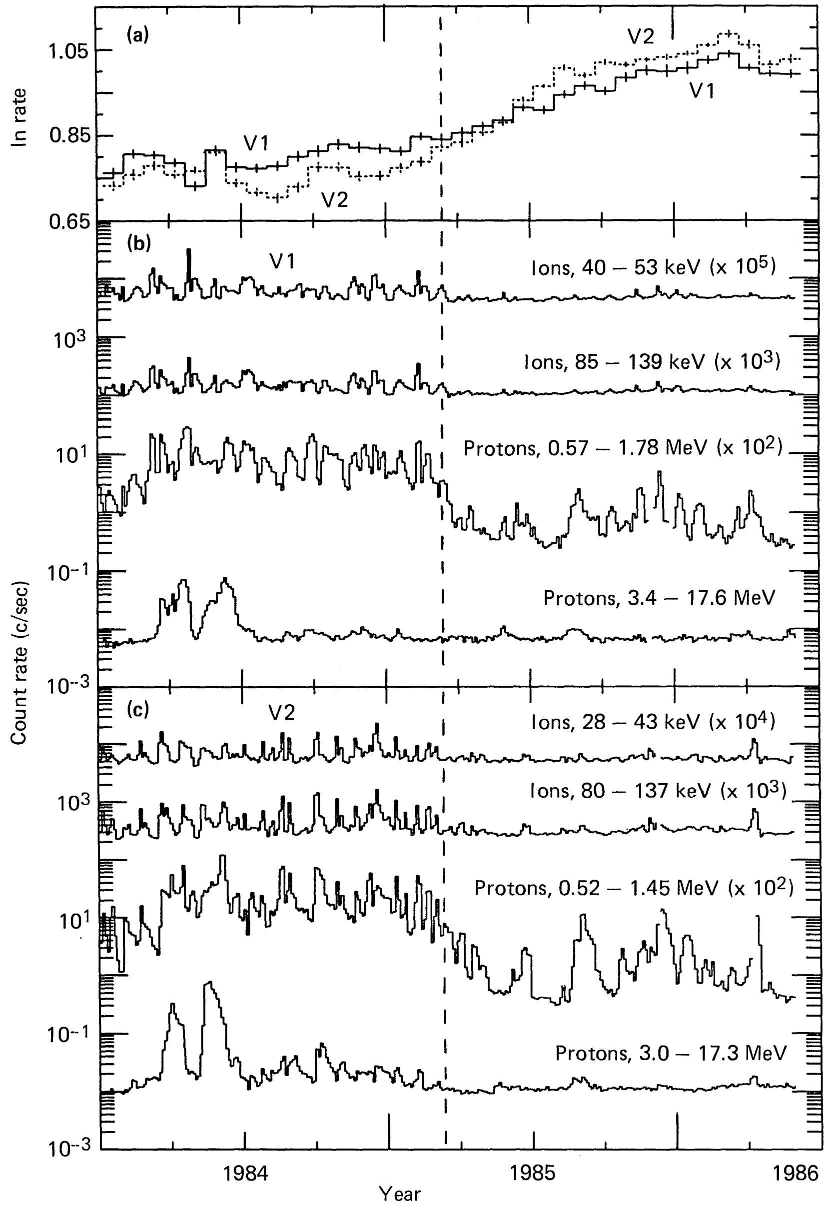


Fig. 7. (a) Natural logarithm of normalized cosmic-ray intensities at Voyager-1 and -2. (b) Selected low-energy ion and proton channels for Voyager-1. (c) Selected low-energy ion and proton channels for Voyager-2. Dashed vertical line denotes first rotation where intensity at Voyager-2 begins the noticeable increase in relation to that at Voyager-1 (after Decker *et al.*, 1987).

The third panel in Figure 6 shows the interplanetary magnetic field and solar wind measurements from Voyager-1 over the same time interval. The top curve in that panel shows the observed interplanetary magnetic field divided by that expected from Parker's ideal spiral approximation while the second curve shows the solar wind velocity observed at the spacecraft. The shock events at Voyager-1 are labeled as *S* for forward shock and *R* for reverse shocks, as indicated. One notes that the cosmic-ray decreases observed at Voyager correspond directly to relative enhancements in the interplanetary magnetic field accompanied by concurrent increases in solar wind velocity. In particular, it is evident that the biggest enhancement in the IMF is coincident with a durable decrease in cosmic-ray intensity which continued on for several months (Burlaga *et al.*, 1984).

It is important to note that the cosmic-ray intensity recovered between the passages of forward-reverse shock pairs, while in the decrease on June 24 the shocks were single, transient shocks and were not accompanied by reverse shock configurations. The condition whereby beyond a few AU fast flows overtake slower flows, interplanetary transients, corotating flows, and shocks from different sources, thereby producing a durable change of the structure of the interplanetary medium which may be irreversible has been called 'entrainment' by Burlaga *et al.* (1984). Thus the large increase in B/Bp at the end of June may be the result of an entrainment of several transients. Burlage *et al.* (1984) find that the long-term modulation of cosmic rays is associated with systems of transient magnetic fields and flows, whereas systems of corotating configurations produce only temporary depressions in the cosmic-ray flux which then lead to recovery. The frequency and magnitude of these transient events apparently determine the overall solar cycle modulation of cosmic rays and maintain the cosmic-ray gradient of ~ 2 to $\sim 4\%\ \mathrm{AU}^{-1}$ observed now for several years by both the Pioneer and Voyager spacecraft (for a review see Venkatesan *et al.*, 1985). Perko and Burlaga (1987) have used the observed magnetic field signatures as inputs to the cosmic-ray transport equation and were able to reproduce the recovery phase of the 11-year solar-driven cosmic-ray cycle as observed by Voyager-2.

A related question to the radial gradient is that of the latitudinal gradient. It relates to the fact that some models (e.g., Fisk, 1976) predict a density gradient in cosmic rays with the cosmic-ray intensity increasing at higher heliographic latitudes. It has been possible in the last two years to make measurements of the latitudinal gradient of cosmic rays due to the excursion of the trajectory of the Voyager-1 spacecraft out of the ecliptic plane, as shown in Figure 2. These measurements are shown in Figure 7 for the period 1984 through 1986 (Decker *et al.*, 1987). The top panel shows the 26-day averages of the cosmic-ray count rates from Voyagers-1 and -2 (> 70 MeV). We note that the intensity at Voyager-2 began to increase above that of Voyager-1 ($R \sim 24$ AU, $\lambda \sim 26°$) in mid-1985 and continued to be higher for at least 24 solar rotations. Using the Voyager-2 and IMP-8 data to correct for the radial gradient, it has been determined that there is an average latitudinal gradient during this period of $\sim(-0.53 \pm 0.1)\%\ \mathrm{deg}^{-1}$ (Decker *et al.*, 1987).

The second and third panels in Figure 7 show lower energy channels from Voyager-1

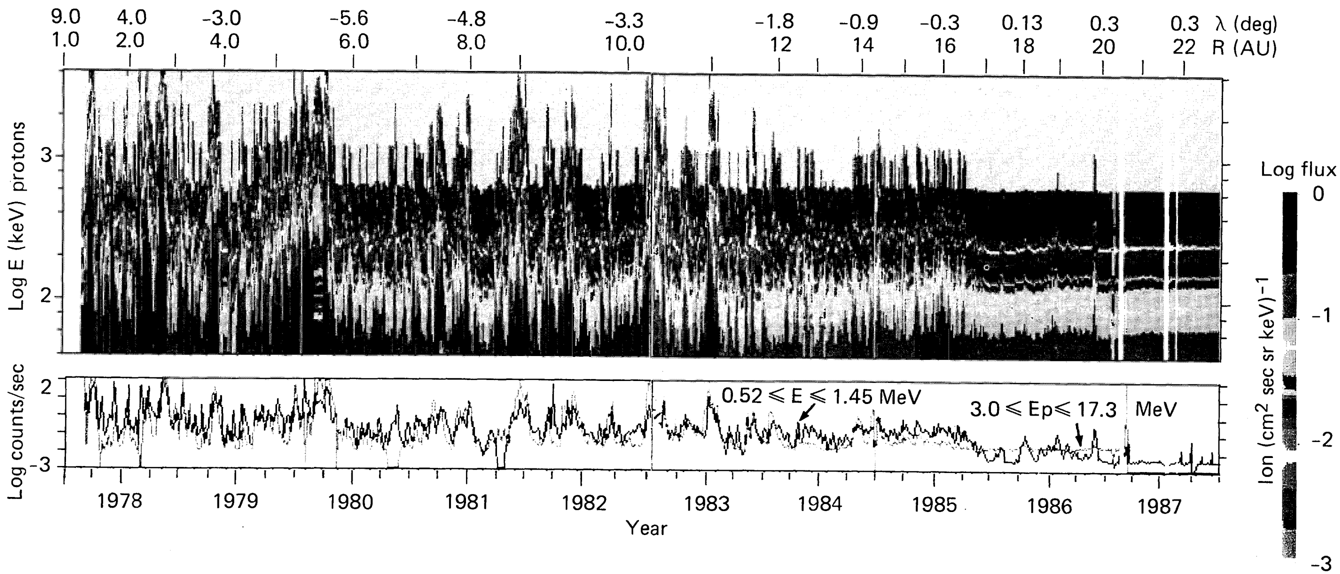


Fig. 8. Colour spectrogram of the ion channels from Voyager-2 over the energy range of 40 keV to ~4 MeV for the 10-year period following the launch of Voyager-2 on 20 August, 1977. The intensity scale interms of protons $cm^{-2} s^{-1} sr^{-1} keV^{-1}$ is shown by the colour bar on the right-hand side. The data around early July 1979 contains significant intensities from the encounter with the planet Jupiter. Note the abrupt dropoff on day ~ 100 of 1985 and the relatively little activity after the spike on day ~ 90, 1986. The line plot below the spectrograms shows the four-day averages of the intensity profile of medium ~ 0.5 MeV and high ($\gtrsim$ 3 MeV) protons from detectors wit a higher geometry factor and lower background than those included in the spectrogram.

and Voyager-2. The bottom curve in each panel ($E_p \gtrsim 3$ MeV) may be used as an indicator of solar flare-associated proton enhancements, while the lower energy ($E_p \sim 0.5$ MeV) channel (third curve) is principally responding to *in situ* shock-accelerated protons, as are the two lowest energy channels. It is evident from the data that few solar-flare proton events were detected throughout the ~ 2.5 year period at these distances (~ 15–25 AU), especially after early 1985. Thus the many individual enhancements were most likely associated with CIR's, i.e., quasi-perpendicular shock pairs. Further support of this view comes from the fact that a power spectrum analysis shows peaks at ~ 13 days and ~ 26 days, indicating strong solar wind stream control (Gold *et al.*, 1987).

It is remarkable that the appearance of the cosmic-ray latitudinal gradient occurred near-simultaneously with a large drop in the intensity of low-energy, shock-accelerated interplanetary ions in early 1985, indicating some change in the level of shock-associated activity in the interplanetary medium both in the eclitic (Voyager-2) and higher latitudes (Voyager-1). The fact that the two Voyager detectors are nearly identifical allows an absolute comparison to be made; we note that the intensities in the ecliptic plane are generally higher than those observed at higher latitudes. It is found that these ions, presumably of local origin, also exhibit a latitudinal gradient ($\sim -3\%$ deg^{-1}) with lower intensities at higher latitudes (Decker *et al.*, 1987; Venkatesan *et al.*, 1987).

4. Low-Energy Profile of Heliosphere

As pointed out in the Introduction, by August, 1987 the Voyager-2 data set consisted of an essentially continuous record of particle intensities from 1 AU through ~ 25 AU, while Voyager-1 extended these measurements to high latitude and to ~ 31 AU. It is thus important to view this data set over this entire time interval, in as broad an energy range as possible, and examine whether some overall patterns can be discerned. This is done in Figure 8 which is a colour spectrogram of intensities from ~ 40 keV to ~ 4 MeV together with line plots of protons in the middle (~ 0.5 MeV) and higher (3–17 MeV) energy ranges.

There are several distinct epochs contained in the data displayed in Figure 8. First, following launch, it is evident that there were several high-intensity solar events with accompanying interplanetary shocks lasting through most of 1978, with substantial fluxes of $\gtrsim 3$ MeV protons as seen in the line plot underneath the spectrogram. Intensities reached a relative minimum by early 1979 but began to increase gradually towards the middle of the year. It is likely that the increases seen in early to mid-79 were associated with upsteam ion intensity enhancements from the planet Jupiter. Note that these enhancements dropped abruptly following the encounter with Jupiter in early July, 1979 (for details see Zwickl *et al.*, 1981; Krimigis *et al.*, 1985). After that time, past the Saturn encounter in August 1981 and through the end of 1982, the intensities were significantly lower with fewer high-energy enhancements than were observed inside ~ 5 AU. This is also evident from the line plot where both the ~ 0.5 MeV and > 3 MeV protons have a decidedly lower baseline throughout this period.

The time interval from 1983 through early 1985 was characterized by few enhance-

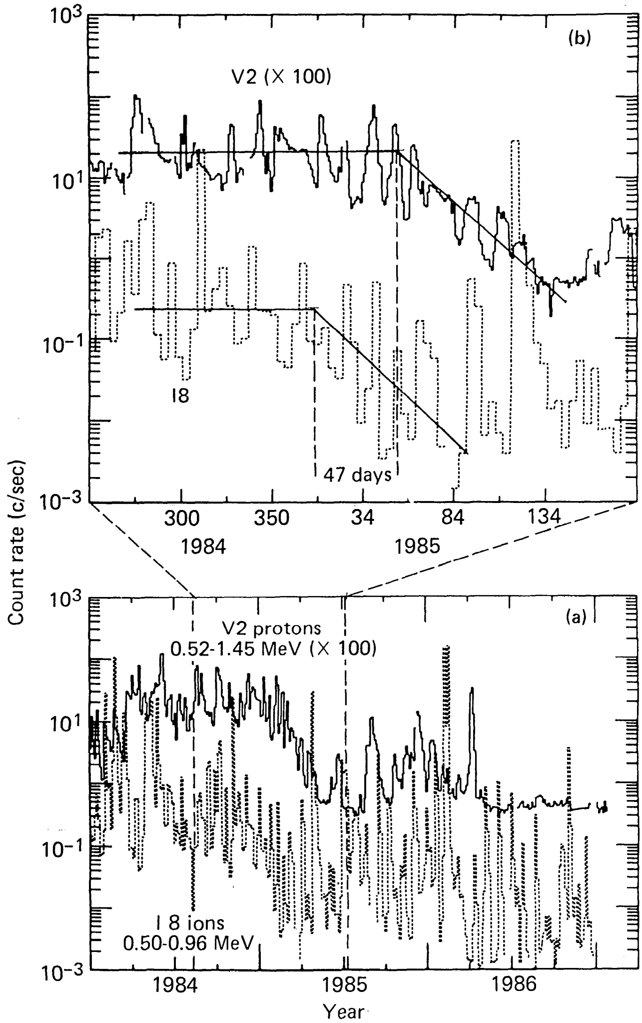


Fig. 9. The flux decrease in early 1985 compared to similar measurements at 1 AU by the IMP-8 spacecraft. (a) Daily average of $0.5 \lesssim E_p \lesssim 1.45$ MeV proton count rates observed by Voyager-2, and 4-day decitile filtered $0.5 \lesssim E_i \lesssim 0.96$ MeV ions observed by IMP-8 at 1 AU. (b) The 300-day interval immediately surrounding the decrease with visual fit to the slope of the decrease and the pre-decrease levels. The time-delay between onset of the decrease at IMP and Voyager-2 is 47 days with an estimated accuracy of $\pm$ 10 days. This time-delay implies a propagation velocity of 567 km s^{-1} between 1 and 16.4 AU (Gold *et al.*, 1988).

ments above 1 MeV and an overall level of activity significantly below what had been observed earlier. This is again reflected in the activity of 0.5 MeV and > 3 MeV protons in the line plot below. From 1985 on, very few enhancements are seen in the colour spectrogram above ~ 600 keV, although the higher sensitivity detector shown in the line plot indicated substantial activity at ~ 0.5 MeV, albeit from a lower baseline, by about an order of magnitude. No enhancements of any kind are observed in the > 3 MeV channel. A new, lower level of intensity was reached after day ~ 100 of 1986 where no discernible activity is seen in the lower energies in the colour spectrogram except an overall increase in background levels associated with the rising cosmic-ray intensity. The lower panel showing the 0.5 MeV proton intensity also shows that at this time interval very little activity is present at that level through the end of the data in early 1987.

To examine whether this overall decrease in activity is a spatial or temporal variation we shown in Figure 9 the intensity of 0.5 MeV protons at Voyager-2 and compare it to the intensity in a similar channel as observed by the IMP-8 spacecraft in the vicinity of Earth (Gold *et al.*, 1988). The IMP-8 data has been filtered using a deci-tile filter because the intensity at the lower energies is greatly affected by magnetospheric activity and it is, therefore, difficult to establish the overall long-term variability at that level. As is seen from the figure, there appears to be a general variation in intensities at 1 AU similar to that observed by the Voyager-2 spacecraft in the outer heliosphere, suggesting that the principal variation in activity observed by Voyager is essentially temporal, and related to solar activity. In fact, the drop in intensities in early 1985 occurred first at 1 AU and then appeared at Voyager ~ 47 days later (Figure 3(a), top), suggesting that the discontinuity in intensity propagated at a speed of ~ 570 km s^{-1}, somewhat faster than the average solar wind velocity.

The most remarkable aspect of the entire set of observations, however, is the absence of large intensities of low-energy protons in the outer heliosphere as the spacecraft approached the ~ 25 AU range at the time of solar minimum when the heliosphere boundary is expected to be close to its minimum distance from the Sun. In fact, modulation theory suggests that the interstellar spectrum at such low energies would be enhanced by several orders of magnitude (Goldstein *et al.*, 1970). Not only do the data not show any increase, but taken together with the data at 1 AU suggest that the radial gradient at these energies is likely to be negative rather than positive at this time of the solar cycle. It will be interesting to continue the measurements at the beginning of the new solar cycle through the last part of the decade as the spacecraft proceeds past the orbit of Neptune and towards the heliospheric boundary.

5. Discussion

The overall structure of the heliosphere varies on several time-scales from a few minutes on a local level (shock transit time) to several hours for ESP events and Forbush decreases, and several days, weeks, and months for CIR regions, culminating in the solar cycle variation of 11 and 22 years. It is possible to attempt to model this structure by taking into account specific eruptions and coronal mass ejections (CME) at the Sun and their passage throughout the inner and outer heliosphere.

Figure 10 shows results from a simulation (Akasofu and Hakamada, 1983) that assumes a pre-existing two-stream state of the solar wind, into which are propagated disturbances caused by six hypothetical flares from the same active region over a period of ~ 10 days. On the left is shown a view of the ecliptic plane with a scale of 5 AU approximately 7 days after several CME's propagated through the inner heliosphere. It

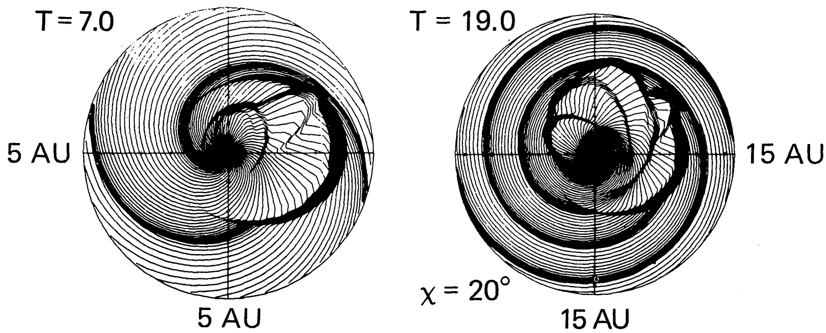


Fig. 10. Example of an empirical model of the propagation of solar flare generated disturbances in the interplanetary medium. *Left*: A view of the ecliptic plane, which prior to the shock contained a basic two-spiral of the IMF, as affected by the shock generated by four successive flares in a ten-day period after ~ 7 days. *Right*: A view of the ecliptic plane to a distance of 15 AU following two additional flares from the same active region. Note that the basic two-sector pattern has essentially been overtaken by the entrainment of several shock wave which appear to have coalesced at the distance of ~ 10 AU (adapted from Akasofu and Hakamada, 1983).

is evident that the interplanetary magnetic field structure is rather complex with an inner boundary in the vicinity of Earth and additional structure extending to ~ 3 AU where the fast solar wind interacts strongly with the pre-existing shock already there in the range from ~ 2.5 to ~ 5 AU. The magnetic field lines in some places are almost in a quasi-parallel configuration with the shock surface, while at other points the geometry is essentially quasi-perpendicular, especially at greater radial distances. The picture on the right follows the development of the interplanetary medium for the same coronal mass ejections, but viewed at about 19 days after the first CME. It is seen from the figure that the inner heliosphere is severely disturbed with several interacting shock fronts containing rather peculiar magnetic field configurations, while the structure beyond ~ 10 AU remains relatively undisturbed.

While this model is *ad hoc*, it is quite useful in illustrating the complexity of interplanetary structures following major solar disturbances. and the variable nature of θ_{Bn} as a particular shock propagates from the inner to the outer heliosphere. In fact Voyager-1 and -2 data have been used by Sarris *et al.* (1987b) to illustrate the fact that energetic particle acceleration signatures changed substantially as a shock propagated from one spacecraft to the next in the course of only a few hours. These authors point out that diffusive acceleration, where the particle must stay confined for several tens of hours in a quasi-parallel geometry for significant energy gain, cannot be effective in the interplanetary medium. This explains the lack of observation of high ($\gtrsim$ 300 keV) energy ions in association with quasi-parallel interplanetary shocks (Table I).

The interpretation of the data presented in the preceding sections, especially those relating to the long-term profile of low-energy ions throughout the heliosphere, can be viewed in the context of these models. The apparent discontinuities in the intensity profile contained in Figures 8 and 9 and the difference in the peak energies attained for several of these events are undoubtedly due to the solar origin of the CME discontinuities and the varying geometry of the IMF with respect to the shock normal. Clearly, in those cases where the shock is close to being quasi-perpendicular the intensity enhancements at higher energy are more substantial. These associations between the intensity enhancements and the soalr wind parameters need to be established on a case-by-case basis as was done for the January 1978 event discussed in detail in this paper. This work is presently under way.

What then, is the most important thrust of the Voyager data? We have seen in this data set a view of the interplanetary medium for the firt time at energies as low as ~ 30 keV over an entire solar cycle and a range in radial distance to some 30 AU. We have observed that energization of ions is a ubiquitous phenomenon in space and time. We have also seen a glimpse of an unusual change in early 1985 in both the cosmic-ray gradient and the cessation of intense shock acceleration activity at the lower energies. An even quieter period is observed after about May 1986. What does this portend for

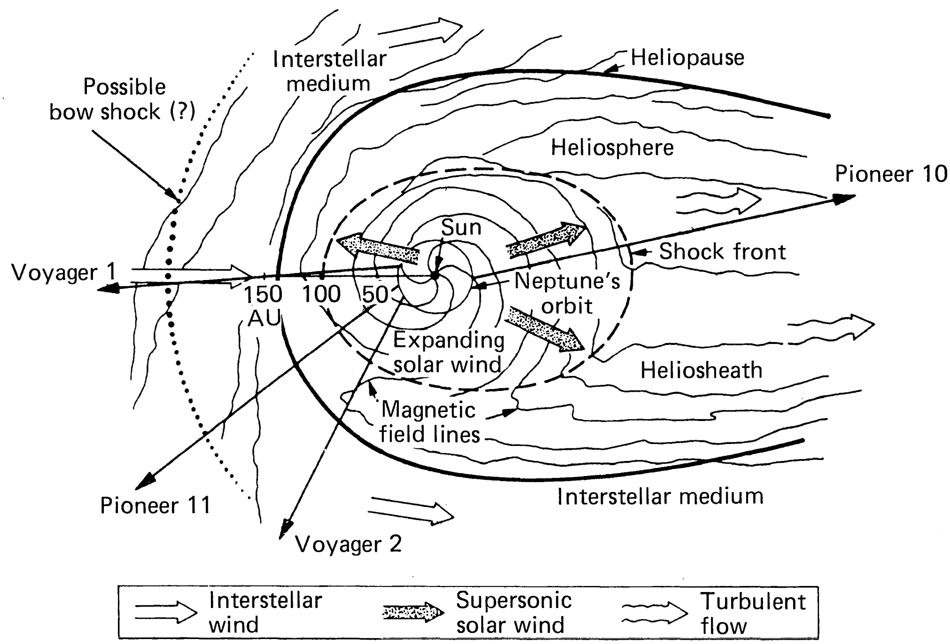


Fig. 11. A schematic view of the heliosphere and its interaction with the local interstellar medium. The expanding solar wind sets up a shock front which is likely to be inside the heliopause since the interaction is subsonic; a potential bow shock may be present well outside the heliopause at distances perhaps as large as 200 AU. The nomenclature follows that established for magnetospheres. The trajectories of the two Voyager and two Pioneer spacecraft are indicated. The Voyager spacecraft will be at distances of approximately 70 AU by the year 2000 and will continue to provide data until the Sun sensor is no longer capable of locking the spacecraft antenna on to the Sun. This is expected to happen at ~ 2010 while the spacecraft is at a distance of ~ 110 AU (adapted from *Solar-Terrestrial Research* for the 1980's, NAP, 1981; also Voyager Document PD 618–150).

the new solar cycle? Will the low-energy fluxes recover? Are we closing in on the heliopause? If so, where are the huge fluxes of low-energy unmodulated interstellar ions expected there?

As is often the case in science, the observational data have generated more questions than they have answered. This, of course is a fitting tribute to the genius of Hannes Alfvén. In developing his ideas, he kept his eyes firmly fixed on the experimental results throughout his career. We believe that that explains why he has been right so many times. What the Voyager results to date have shown is that his early ideas that all cosmic rays are not 'cosmic' were on target, even if all the details did not turn out to be exactly what he might have anticipated.

What, then, is our current picture and what about future data from the Voyagers and Pioneers? Figure 11 shows one possible picture of the heliosphere with the trajectories of Voyagers-1, -2 and Pioneers-10, -11 superimposed. The expanding solar wind produces the typical Archimedian spiral interplanetary magnetic field through the first 50 or perhaps 100 AU, while beyond that point the influence of the interstellar medium begins to be felt, forming a shock front and a heliopause outside the shock front, perhaps extending as far out as 150 AU. Substantially ahead of the heliopause, there may well be a possible bow shock or current layer which would be the true boundary between the influence of the Sun and the local interstellar medium. Cosmic rays at the highest energies are arriving from the interstellar medium but they would seem to be propagating into the heliosphere mostly in the ecliptic plane, if our latitudinal gradient observations are interpreted in this manner. There are suggestions that the latitude gradient will reverse at the end of the current sunspot cycle. Hopefully, the Voyagers will still be there to record the event.

6. Conclusions

We set out with the primary purpose of presenting our *in situ* measurements in the outer heliosphere obtained from instruments on board the Voyagers and to demonstrate the shock-associated acceleration of charged particles. This we have successfully accomplished; the energization to hundreds of MeV and possibly even to a GeV by such a process is clearly a tribute to Hannes Alfvén, who even in the early fifties, envisioned the possibility of acceleration within the solar system. His intuition had no other support than ground-based instrumentation. Particularly over the last 15 years, technological progress has enabled observations in the three-dimensional heliosphere to test out some of his ideas, at least with respect to the lower energies.

But what about the broader picture? What about his contention in the 50's and perhaps still to this day, that the local origin (and thus acceleration) extends to quite high energies (except for particles whose gyroradii exceed the scale of the heliosphere)? He is likely to say 'acceleration to a few hundred MeV's, that is a good beginning, but that is not enough!' While the observations reviewed in this paper do illustrate the potential vision of Alfvén, they do not directly address the larger problem, nor was it intended to be, based on the measurements we have.

It is perhaps appropriate and wise to keep one's mind open and continue to look for further evidence for local origin of cosmic rays of GeV energy, although we have to profess that the prevailing view is in terms of galactic cosmic-rays incident at the heliospheric boundary and traversing in a steady fashion, although modulated by solar activity. It is well known that we have sporadic and short term emissions of solar particles, of even tens of GeV. Perhaps the time is right to explore the acceleration by double layers over the solar poles, in view of the eventual observations by the Ulysses spacecraft.

Acknowledgements

We are grateful to our colleagues Drs R. B. Decker, B. H. Mauk, and R. E. Gold for assistance in the preparation of the data and for many useful discussions on the results. We wish to express our appreciation to Drs T. P. Armstrong and E. T. Sarris for a careful reading of the manuscript, and the referee for their critical comments. This work has been supported by NASA under Task I of Contract N00039–87–C–5301 and at the University of Calgary under subcontract 601566. Additional support to D. Venkatesan from NSGRC (Ottawa) Grant 69–1565 is also acknowledged.

References

Akasofu, S.-I. and Hakamada, K.: 1983, *Geophys. Res. Letters* **7**, 577.

Alfvén, H.: 1949, *Phys. Rev.* **75**, 1732.

Alfvén, H.: 1950, *Phys. Rev.* **77**, 375.

Alfvén, H.: 1954a, *Nuovo Cimento Suppl.* **11**, No. 2.

Alfvén, H.: 1954b, *Tellus* **6**, 232.

Alfvén, H.: 19659, *Tellus* **11**, 106.

Armstrong, T. P., Chen, G., Sarris, E. T., and Krimigis, S. M.: 1977, 'Acceleration and Modulation of Electrons and Ions by Propagating Interplanetary Shocks', in M. A. Shea *et al.* (eds.), *Study of Travelling Interplanetary Phenomena,* D. Reidel Publ. Co., Dordrecht, Holland, pp. 367–389.

Bavassano, B., Smith, E. J., and Tsurutani, B. T.: 1987, *J. Geophys. Res.* **92**, 285.

Bryant, D. A., Cline, T. L., Desai, U. D., and MacDonald, F. B.: 1962, *J. Geophys. Res.* **67**, 4983.

Decker, R. B.: 1983, *J. Geophys. Res.* **88**, 9959.

Decker, R. B., Krimigis, S. M., and Venkatesan, D.: 1987, *J. Geophys. Res.* **92**, 3375.

Decker, R. B., Pesses, M. E., and Krimigis, S. M.: 1981, *J. Geophys. Res.* **86**, 8819.

Duggal, S. P.: 1979, *Rev. Geophys. Space Phys.* **17**, 1021.

Fälthammar, C.-G.: 1963, 'Acceleration of Cosmic-Ray Particles by Magnetic Pumping', in B. Peters (ed.), *Proc. International School of Physics 'Enrico Fermi'*, Course XIX, Academic Press, New York, p. 99.

Fisk, L. A.: 1976, *J. Geophys. Res.* **81**, 4646.

Gold, R. E., Decker, R. B., Krimigis, S. M., Lanzerotti, L. J., and Maclennan, C. G.: 1988, *J. Geophys. Res.* **93** (in press).

Goldstein, M. L., Fisk, L. A., and Ramaty, R.: 1970, *Phys. Rev. Letters* **25**, 832.

Kennel, C. F., Coroniti, F. V., Scarf, F. L., Livesey, W. A., Russell, C. T., Smith, E. J., Wenzel, K. P., and Scholer, M.: 1986, *J. Geophys. Res.* **91**, 11917.

Krimigis, S. M. and Sarris, E. T.: 1985, 'Acceleration of Ions and Electrons to Near Cosmic-Ray Energies in a Perpendicular Shock: The January 6, 1987 Event', *Proceedings 19th International Conference of Cosmic Rays,* pp. 170–173.

Krimigis, S. M., Armstrong, T. P., Axford, W. I., Bostrom, C. O., Fan, C. Y., Gloeckler, G., and Lanzerotti, L. J.: 1977, *Space Sci. Rev.* **21**, 329.

Krimigis, S. M., Zwickl, R. D., and Baker, D. N.: 1985, *J. Geophys. Res.* **89**, 3947.

Levy, E. H., Duggal, S. P., and Pomerantz, M. A.: 1976, *J. Geophys. Res.* **81**, 51.

Lockwood, J. A., Webber, W. R., and Jokipii, J. R.: 1986, *J. Geophys. Res.* **91**, 2851.

Perko, J. S. and Burlaga, L. F.: 1987, *J. Geophys. Res.* **92**, 6127.

Pesses, M. E.: 1981, *J. Geophys. Res.* **86**, 150.

Pomerantz, M. A. and Duggal, S. P.: 1974, *J. Geophys. Res.* **79**, 913.

Potter, D. W.: 1981, *J. Geophys. Res.* **86**, 11111.

Pyle, K. R. and Simpson, J. A.: 1977, *Astrophys. J.* **215**, L89.

Rosner, R., Chupp, E. L., Gloeckler, G., Gorney, D. J., Krimigis, S. M., Mok, Y., Ramaty, R., Swift, D. W., Vlahos, L., and Zweibel, E. G.: 1984, 'Particle Acceleration', in D. Butler and K. Papadopoulos (eds.), *Solar-Terrestrial Physics: Present and Future*, NASA-RP-1120.

Sarris, E. T. and Krimigis, S. M.: 1985, *Astrophys. J.* **298**, 676.

Sarris, E. T., Anagnostopoulos, G. C., and Trochoutsos, P. C.: 1984, *Solar Phys.* **93**, 195.

Sarris, E. T., Dodopoulos, C. A., and Venkatesan, D.: 1987a, 'East-West Asymmetry of Solar Flare Shock Effects on High Energy Particles ($E \gtrsim 30$ MeV)', presented at the *20th International Cosmic Ray Conference*, Moscow, SH7.2–5, August 1987.

Sarris, E. T., Krimigis, S. M., and Decker, R. B.: 1987b, 'Evolution of Shock-Associated Energetic Ion Enhancements between Closely Separated Spacecraft', presented at *Solar Wind 5 Conference*, Boulder (CO), August 24–28.

Sarris, E. T., Krimigis, S. M., and Armstrong, T. P.: 1976, *Geophys. Res. Letters* **3**, 133.

Smith, E. J. and Wolfe, J. H.: 1979, *Space Sci. Rev.* **23**, 217.

Venkatesan, D., Decker, R. B., Krimigis, S. M., and Van Allen, J. A.: 1985, *J. Geophys. Res.* **90**, 2905.

Venkatesan, D., Decker, R. B., and Krimigis, S. M.: 1987, 'Cosmic-Ray Intensity Gradients During 1984–'86', SH6.3–5, *Proc. 20th Int. Cosmic Ray Conf. Moscow, U.S.S.R.*, August 2–13, pp. 385–388.

Webber, W. R., Lockwood, J. A., and Jokipii, J. R.: 1986, *J. Geophys. Res.* **91**, 4103.

Zwickl, R. D., Krimigis, S. M., Carbary, J. F., Keath, E. P., Armstrong, T. P., Hamilton, D. C., and Gloeckler, G.: 1981, *J. Geophys. Res.* **86**, 8125.

INTERACTION OF INTERSTELLAR PICK-UP IONS WITH THE SOLAR WIND*

E. MÖBIUS, B. KLECKER, D. HOVESTADT, and M. SCHOLER
Max-Planck-Institut für Physik und Astrophysik, Institut für extraterrestrische Physik, Garching, F.R.G.

(Received 21 September, 1987)

Abstract. The interaction of interstellar pick-up ions with the solar wind is studied by comparing a model for the velocity distribution function of pick-up ions with actual measurements of He^+ ions in the solar wind. The model includes the effects of pitch-angle diffusion due to interplanetary Alfvén waves, adiabatic deceleration in the expanding solar wind and the radial variation of the source function. It is demonstrated that the scattering mean free path is in the range ≤ 0.1 AU and that energy diffusion can be neglected as compared with adiabatic deceleration. The effects of adiabatic focusing, of the radial variation of the neutral density and of a variation of the solar wind velocity with distance from the Sun are investigated. With the correct choice of these parameters we can model the measured energy spectra of the pick-up ions reasonably well. It is shown that the measured differential energy density of the pick-up ions does not vary with the solar wind velocity and the direction of the interplanetary magnetic field for a given local neutral gas density and ionization rate. Therefore, the comparison of the model distributions with the measurements leads to a quantitative determination of the local interstellar gas density.

1. Introduction

The interaction between the interstellar neutral gas and the interplanetary medium has been a topic of considerable interest for many years in theory as well as experimental investigation. It was not before the early 1970's that by means of spaceborn UV scattering measurements the first observational evidence of interstellar hydrogen in the solar system was presented (Bertaux and Blamont, 1971; Thomas and Krassa, 1971). The first conclusive results on interstellar helium were presented by Weller and Meier (1974). These observations initiated a rapid development of the understanding of the interaction processes. For an overview the reader may be referred to the reviews of Axford (1972), Fahr (1974), Holzer (1977), and Thomas (1978 and references quoted therein).

The solar system is moving with respect to the local interstellar medium (LISM) with a relative velocity of about 20 km s^{-1}. Thereby the interstellar gas is streaming through the heliosphere like an interstellar wind, which – when approaching the Sun – is subject to the forces of solar gravitation and radiation pressure. On the other hand the interstellar neutrals are ionized by solar UV radiation, by charge exchange with the solar wind ions and by electron collisions. These newly created ions are then picked up by the solar wind via interaction with the interplanetary magnetic field and finally swept out of the heliosphere. Therefore, the measurement of the absolute flux of the pick-up ions and of its spatial variation in interplanetary space can be used to determine the

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 487–505.

density, temperature, and relative velocity of the local interstellar medium.

A similar interaction with the solar wind is seen for neutral gas originating from other sources in interplanetary space. Planetary atmospheres and comets are only of local importance. The desorption of gas from interplanetary dust constitutes an extended source of neutrals in the inner solar system, i.e., at distances <0.5 AU (e.g., Fahr *et al.*, 1981).

After their initial pick-up by the interplanetary magnetic field the newly generated ions are subjected to efficient scattering processes by means of intrinsic and self-generated MHD waves (e.g., Wu and Davidson, 1972: Wu *et al.*, 1973: Winske *et al.*, 1985). Recently, Lee and Ip (1987) have shown that the waves generated by the hydrogen pick-up ions beyond 5 AU (i.e., the distance where the maximum ionization occurs) dominate the wave spectrum, while in the case of interstellar helium their contribution to the wave power is negligible compared with that of the intrinsic Alfvén waves in the solar wind at 1 AU. In early studies of the pick-up of interstellar gas by the solar wind it was assumed that the newly created ions are quickly thermalized and assimilated into the solar wind (e.g., Semar, 1970; Holzer, 1972; Fahr, 1974). These assumptions led to a contribution of He^+ to the solar wind, which should be detectable by solar wind plasma experiments for low solar wind temperatures. However, Feldman *et al.* (1974) reported an upper limit for He^+ which ruled out a thermalized He^+ ion distribution in the solar wind. Vasyliunas and Siscoe (1976) have presented a model of the pick-up ion distribution in which they assumed instantaneous isotropization of the ions due to pitch-angle scattering and subsequent adiabatic deceleration in the expanding solar wind. The resulting distribution with a sharp cut-off at the solar wind energy (in the rest frame of the solar wind) was recently found for helium by Möbius *et al.* (1985a). Generalizing the model by Vasyliunas and Siscoe (1976), Isenberg (1987) included the effect of energy diffusion. Comparison with the results by Möbius *et al.* (1985a) showed that the effect of energy diffusion is negligible.

In this paper we will derive an analytical model of the velocity distribution of pick-up ions which, based on the model of Vasyliunas and Siscoe (1976), includes explicitly the effects of pitch-angle diffusion, adiabatic deceleration and the variation of the neutral density with distance from the Sun. By comparing the model distributions with the measured pick-up ion distributions of interstellar helium it is demonstrated that the mean free path for pitch-angle scattering is smaller than 0.1 AU and that the shape of the spectrum puts an upper limit on the ionization rate for the interstellar helium. With the model as presented in this paper the measured pick-up ion distributions of interstellar helium are reproduced in a consistent way. This enables us to use the measurement of pick-up ions as a quantitative diagnostic tool for determining the local neutral gas density in the solar wind. Due to the annual journey of the Earth around the Sun the spatial distribution of the neutrals can be constructed, from which we can derive the temperature and relative velocity of the LISM.

In Section 2 we derive the distribution function of pick-up ions and the basic results of the model calculation. In Sections 3 and 4 we describe the instrumentation and the simulation of the energy spectra as seen by the instrument. The results as seen by the instrument in interplanetary space are compared with the simulated spectra in Section 5. The results are discussed in Section 6 in terms of the processes which determine the observed energy spectra, followed by a summary.

2. The Model of the Pick-up Ion Distribution

2.1. Initial pick-up

In contrast to genuine solar wind ions freshly created ions in interplanetary space are initially at rest. Immediately after ionization they are subject to the combined forces of the interplanetary $v_{sw} \times B$ electric field and the magnetic field B. In the inertial system the ions initially perform a cycloidal motion perpendicular to the local magnetic field. Their velocity varies between basically zero (the relative velocity of the neutral gas, ≈ 20 km s^{-1} for the interstellar gas, is neglected here compared with the solar wind velocity) and a maximum value

$$v_{\perp \max} = 2v_{sw} \sin \Theta \,, \tag{1}$$

which is determined by the solar wind velocity and the angle Θ between its flow direction and the local magnetic field. The maximum energy of pick-up ions then is

$$E_{\perp \max} = 4(m/2)v_{sw}^2 \sin^2 \Theta \,. \tag{2}$$

For constant solar wind conditions the resulting velocity distribution in the solar wind frame is a ring in velocity space with the pitch angle Θ. The ions gyrate with a velocity $v_\perp = v_{sw} \sin \Theta$ and move along the field with $v_\| = v_{sw} \cos \Theta$. The signatures of such undisturbed ring distributions have been observed after the artificial injections of lithium clouds into the solar wind (e.g., Möbius *et al.*, 1986).

2.2. Scattering processes

The energetic ring distribution with ion velocities $v_{sw} \gg v_A$ in the solar wind frame of reference is unstable to the generation of low-frequency MHD waves as has been discussed by, e.g., Wu and Davidson (1972), Wu *et al.* (1973), and Winske *et al.* (1985). On the other hand there exists already a background distribution of Alfvén waves in the solar wind. The resonant interaction with the ambient and the self-generated waves leads to an isotropization of the original ion distribution. Here scattering of the ions in pitch-angle μ is much more efficient than diffusion in energy E, since according to quasi-linear theory the corresponding scattering efficiencies scale like $(v_A/v_{sw})^2$ for pick-up ions at the solar wind velocity, where v_A is the local Alfvén velocity.

These effects also determine the basic behaviour of the propagation of cosmic rays in interplanetary space. For a review of cosmic-ray propagation (see, e.g., Jokipii, 1971). From an extrapolation of compiled cosmic-ray propagation parameters (Mason *et al.*, 1983) the mean-free scattering length λ of interstellar He^+ pick-up ions (with a typical magnetic rigidity of 5–10 MV in the rest frame of the solar wind) can be estimated to be $\lambda \approx 0.05$ AU. Hence, pitch-angle scattering seems to be rather efficient for pick-up ions from an extended source like the interstellar gas. Within a fraction of 1 AU away from the ionization site the ions fill a spherical shell in velocity space which is essentially comoving with the solar wind. Since the ions are injected into the solar wind with an initial energy of $E_0 = (m/2)v_{sw}^2$ the radius of the shell is equal to the solar wind velocity v_{sw}.

The scattering in pitch-angle $\mu = \cos\Theta$ on a sphere in velocity space in the expanding solar wind (with a diverging magnetic field), may be described by (e.g., Roelof, 1969)

$$\frac{\partial f}{\partial t} = \frac{1}{\tau}\frac{\partial}{\partial \mu}\left[\{1-\mu^2\}\frac{\partial f}{\partial \mu}\right] - \frac{v_{sw}}{2L}\{1-\mu^2\}\frac{\partial f}{\partial \mu}, \tag{3}$$

where $L \approx R/2$ (R = distance from the Sun) is the typical scale-length for the variation of the magnetic field strength, $1/L = -1/B\,(\partial B/\partial z)$ (z taken along the magnetic field line). The assumptions underlying (3) are the following: the pitch-angle diffusion coefficient $D_{\mu\mu}$ varies like $\sin^2\Theta = 1-\mu^2$ over the sphere (isotropic scattering). The pick-up ions are convected with the solar wind velocity v_{sw}, which now is the macroscopic transport velocity for the ions. This is different from cosmic-ray propagation theory, where the ion velocity v generally is large compared to the solar wind velocity v_{sw}. Therefore, the ion-velocity v has already been replaced by v_{sw} in the second term of Equation (3). Let $\lambda = v_{sw}\tau$ be a scattering mean-free path during the convection of the distribution with the solar wind. Then (3) transforms in the co-moving frame into

$$\frac{\partial f}{\partial R} = \frac{1}{\lambda}\frac{\partial}{\partial \mu}\left[\{1-\mu^2\}\frac{\partial f}{\partial \mu}\right] - \frac{1}{2L}\{1-\mu^2\}\frac{\partial f}{\partial \mu}. \tag{4}$$

The asymptotic solution, i.e., $\partial f/\partial R$, or $\partial f/\partial t = 0$, of relation (4) is given (Roelof, 1969) by

$$f(\mu, R) = 1/4\pi\, e^{\lambda\mu/R}. \tag{5}$$

This leads to a deviation from an isotropic distribution in the asymptotic case ($t \to \infty$) of the order of λ/R. At distances > 0.5 AU from the Sun this deviation, i.e., the second term on the right-hand side of (4), is small, and the relation can be solved for variations of f with R including only the first term. For λ independent of R the solution then becomes

$$f(\mu, R) = 1/4\pi \sum_{l=0}^{\infty} (2l+1)\, e^{-l(l+1)R/\lambda}\, P_l(\mu)\, P_l(\mu_0), \tag{6}$$

where $P_l(\mu)$ are Legendre polynomials of order l.

2.3. Adiabatic Deceleration

Also known from cosmic-ray propagation is the fact that ion distributions, which are convected with the solar wind, are subject to adiabatic cooling due to the radial expansion of the solar wind. Treating the pick-up ion distribution as an ideal gas (adiabatic index $\gamma = \frac{5}{3}$, the radius v of the spherical shell in velocity space, as seen at the distance R_0 from the Sun, varies as

$$v/v_{sw} = (R_0/R)^{-2/3}\,, \tag{7}$$

where R corresponds to the location of ionization of the ion sample. The ideal gas approximation is valid for immediate isotropization due to pitch-angle scattering, i.e., the scattering mean free path λ is small compared with the typical scale-length for the typical value of $\lambda\,(\approx 0.05)$ as given above. This relation is equivalent to a unique mapping of the distance R from the Sun into velocity space and, hence, the velocity distribution of pick-up ions can be derived from the radial source distribution, as is demonstrated in Figure 1. Since the ions are quickly distributed over a sphere in velocity space, the source function may be written as

$$S(R)\,\mathrm{d}R = v_{sw}(4\pi)v^2 f(v)\,\mathrm{d}v \tag{8}$$

i.e.,

$$f(v) = \frac{S(R)}{v_{sw}(4\pi)v^2\,(\mathrm{d}v/\mathrm{d}R)}\,, \tag{9}$$

where the right-hand side of Equation (8) describes the ion flux radially outward with the solar wind velocity v_{sw}. This relation holds for any functional dependence $v(R)$. For

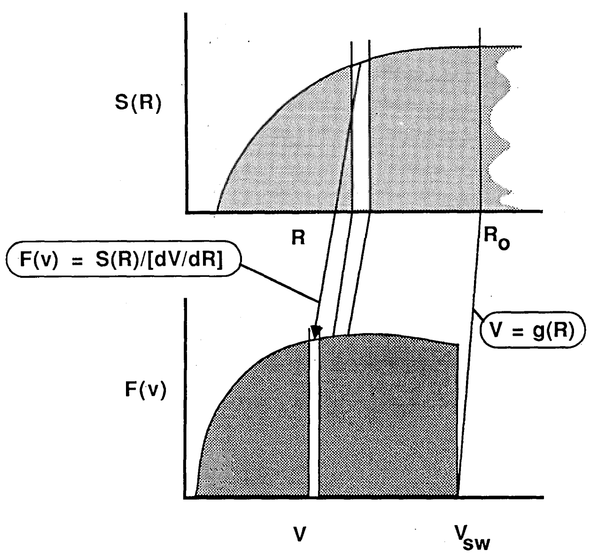


Fig. 1. Mapping of a source distribution over radial distance from the Sun into a velocity distribution, assuming adiabatic deceleration of the ion velocity distribution.

adiabatic deceleration (Equation (7)) this results in a velocity distribution

$$f(v) = \frac{3S(R)R_0}{8\pi v_{sw}^4} \left\{\frac{v}{v_{sw}}\right\}^{-3/2}, \tag{10}$$

where R is taken as a function of v and, therefore, leads to a simple relationship for the distribution function.

2.4. Ionization of the Interstellar Gas

The source distribution $S(R) = N(R)v_{ion}(R)$ depends on the local neutral gas density $N(R)$ and ionization rate $v_{ion}(R)$. Interstellar neutrals are ionized by solar UV radiation, charge exchange with solar wind protons, and electron collisions. The solar UV flux varies like $1/R^2$ with distance from the Sun and depends substantially on the solar activity (see, e.g., Hinteregger, 1976). For the purpose of this paper we will assume constant UV flux with time. However, for a study of the temporal variations of the ion distribution the variability has to be taken into account. The solar wind flux also varies like $1/R^2$ and depends on the conditions in the solar corona (Hundhausen, 1972). The electron collisional ionization rate depends on the distribution function of the solar wind electrons. In the outer heliosphere, i.e., beyond 5 AU, the electron ionization rate may be enhanced by Alfvén's critical ionization velocity effect (Alfvén, 1954), as has been discussed by Petelski *et al.* (1980). In the inner heliosphere the ionization of interstellar hydrogen is dominated by charge exchange with the solar wind, while UV ionization prevails for interstellar helium (Holzer, 1977), both varying as $1/R^2$ as mentioned above. The newly generated ions are convected with the expanding solar wind, so that (after ionization) the density varies as $1/R^2$ with distance from the Sun. Therefore, the density of ions injected into the solar wind does not vary with distance R.

Thus the velocity distribution $f(v)$ solely reflects the radial variation of the neutral gas density $N(R)$. The distribution of interstellar neutral particles in the vicinity of the Sun is basically determined by the solar gravitation, radiation pressure, and removal by ionization. According to e.g., Fahr (1968), Blum and Fahr (1969), Holzer (1972), and Axford (1972) for a cold interstellar gas the spatial distribution $N(R, \Theta)$ of the neutrals is given by

$$N(R, \theta) = \frac{N_0}{\sin\theta} \left\{\frac{\partial b_1}{\partial R} e^{\Lambda\theta/b_1} + \frac{\partial b_2}{\partial R} e^{-\Lambda(2\pi-\theta)/b_2}\right\}, \tag{11}$$

with

$$\begin{Bmatrix} b_1 \\ b_2 \end{Bmatrix} = \left[\left(\frac{1}{2}\sin\theta\right)^2 + (1-\sigma)\frac{GM}{V_0^2} R(1-\cos\theta)\right]^{1/2} + \frac{1}{2} R\sin\theta; \tag{11a}$$

N_0 being the density at infinity, θ the angle between the direction of the Sun's motion (with velocity V_0) with respect to the interstellar medium and the line connecting the observer with the Sun, GM/R^2 is the gravitational force of the Sun and σ is the relative contribution of the radiation pressure (which is negligible for helium). $\Lambda = R_0^2 \nu_{\mathrm{ion}}/V_0$ is a characteristic penetration depth of the interstellar gas, which depends on the ionization rate ν_{ion} at the reference distance R_0 and the velocity of the interstellar wind V_0. Equation (11) neglects the thermal spread in the velocities of the neutral particles, which leads to significant modifications in the close vicinity of the downwind axis. Equation (11) may still be used for angles θ larger than the halfwidth of the gravitational focusing cone on the downwind axis, which is typically 15° for helium (e.g., Dalaudier *et al.*, 1984).

2.5. Resulting distribution function

By the combination of Equation (6) with (10) the velocity and pitch-angle distribution function of pick-up ions in the solar wind can be written as

$$f(\mu, w) = \frac{3S(R)R_0}{8\pi v_{sw}^4} \{w\}^{-2/3} \sum_{l=0}^{\infty} (2l+1)\, e^{-l(l+1)R/\lambda} P_l(\mu) P_l(\mu_0)\,, \qquad (12)$$

where $w = v/v_{sw}$ is the ion velocity normalized to the solar wind velocity. This solution is exact in the limits of negligible energy diffusion and ideal isotropic adiabatic deceleration, i.e., when separation of the v and μ variables can be applied. Isenberg (1987) modelled the velocity distribution including explicitly the effects of energy diffusion. A comparison with the measurements of pick-up He^+ ions by Möbius *et al.* (1985a) showed that energy diffusion can be indeed neglected compared with adiabatic deceleration for observations at 1 AU. This can be confirmed by taking the velocity distribution function derived by Isenberg (1987) instead of the distribution (12) when comparing with the measurement according to the procedure outlined below. Isotropic adiabatic deceleration in the solar wind requires rapid pitch-angle diffusion compared with the expansion of the volume element in the solar wind. With λ of the order of

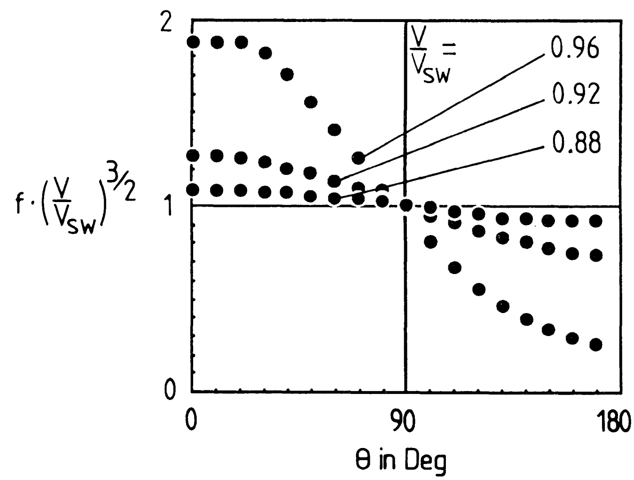


Fig. 2. Normalized velocity distribution function $f(v)\,(v/v_{sw})^{3/2}$ for various values of $w = v/v_{sw}$ as a function of pitch-angle Θ after injection of pick-up ions with an initial pitch-angle $\Theta = 0$.

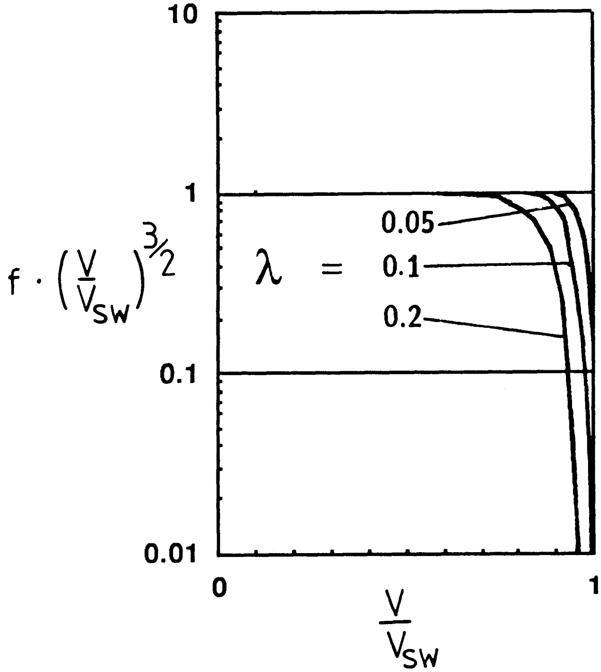


Fig. 3. Normalized velocity distribution function in the solar wind direction (Sun-sector) as a function of V/V_{sw} for different values of the scattering mean free path ($\lambda = 0.05, 0.1, 0.2$).

0.05 AU and a typical scale-length for deceleration $1/(\mathrm{d}B/\mathrm{d}R)$ of 1 AU this requirement holds for distances ≥ 0.5 AU. Near the Sun modifications are necessary, which will be outlined qualitatively in Section 6.

Equation (12) presents a description in terms of nested spherical shells with the pitch-angle distribution evolving from a ring at pitch-angle μ_0 to a homogeneous sphere towards lower velocities w, as is illustrated in Figure 2 for an injection pitch-angle $\mu_0 = 1$ and a scattering mean free path $\lambda = 0.1$ AU. Significant deviation from a homogeneous sphere are found only for $v/v_{sw} \geq 0.9$. Figure 3 shows a cut through the distribution at $\mu = -1$ for different values of λ. It is demonstrated that the cut-off is smoothed considerably for $\lambda \geq 0.1$ AU. For smaller values the change is hardly noticeable and will be of no importance in the comparison with measurements.

3. Measurements of the Pick-up Ions

The measurements presented in this paper were obtained with the SULEICA (Supra-thermal Energy Ionic Charge Analyzer) instrument of the Max-Planck-Institut and the University of Maryland on board the AMPTE/IRM spacecraft. The IRM was launched on August 16, 1984, into a highly elliptical orbit with an apogee of $\approx 18.7\, R_E$. The satellite and the instrumentation worked until the spacecraft battery failed on August 12, 1986. Between July and December the satellite spent a large fraction of each orbit in the solar wind upstream of the Earth's bow shock. Therefore, a comprehensive data set on interstellar pick-up ions over the two years mission could be obtained.

The SULEICA instrument combines the selection of incoming ions according to their energy per charge by a spherical section electrostatic analyzer with a subsequent time-of-flight analysis and the final measurement of the residual ion energy in a silicon surface barrier detector. With these techniques the ions are differentiated according to their mass, ionic charge, and energy. The energy range of the instrument from 5 to 270 keV e^{-1} is covered by stepping the analyzer voltage in up to 24 logarithmically spaced voltage steps synchronously with the spacecraft spin. In the low-energy range (i.e., below 40 keV total) the signal of the solid state detector does not exceed the noise level. However, for a specific energy step, ions of different mass per charge ratios are clearly separated by their time-of-flight. The major ion species are accumulated on board according to a matrix rate system using look-up tables for time-of-flight and energy pulse heights. In our investigation we rely on the corresponding rate information for the He^+ ions.

The fan-like aperture of the SULEICA electrostatic analyzer covers a solid angle of 10° in azimuth and 40° in polar angle symmetric to the plane perpendicular to the spin axis. The directional information in azimuth is provided by a sectoring scheme with 16 sectors for H^+ and He^{2+} and eight sectors for all other ions, including He^+ which is of interest here. The center viewing direction of sector number 4 (out of sectors 0 to 7) is pointing towards the Sun (GSE x direction) and will be further called the Sun-sector. A more detailed description of the instrument may be found elsewhere (Möbius *et al.*, 1985b).

4. Simulation of the Ion Measurements

In order to allow a quantitative comparison of the model distributions with the results of the ion spectrometer measurements the quantities, as measured by the instrument, have to be simulated by integration over the instrumental sectors and energy steps, which have been described in the previous section, in the spacecraft frame of reference. Figure 4 shows a cut through a pick-up ion velocity distribution in the plane perpendicular to the spacecraft spin axis together with the sectoring scheme (radial fans) and the energy steps (concentric ring system) of the SULEICA instrument. For the simulation an equidistant grid with 10 points in each direction – velocity, azimuthal angle, and elevation angle (not shown in Figure 4) – is used within each individual channel.

Each individual data point $((v', \Theta', \Phi')$ in the spacecraft frame is first transformed by a rotation into a system (v'', Θ'', Φ'') with the z-axis pointing into the solar wind direction, followed by a transformation into the solar wind frame of reference $(v''', \Theta''', \Phi''')$ according to

$$\begin{aligned} v''' &= \sqrt{[v''^2 + v_{sw}^2 - v_{sw} v'' \cos(\Theta'')]}\,, \\ \Theta''' &= \arctan[\sin(\Theta'')/(\cos(\Theta'') + v_{sw}/v'')]\,. \end{aligned} \tag{13}$$

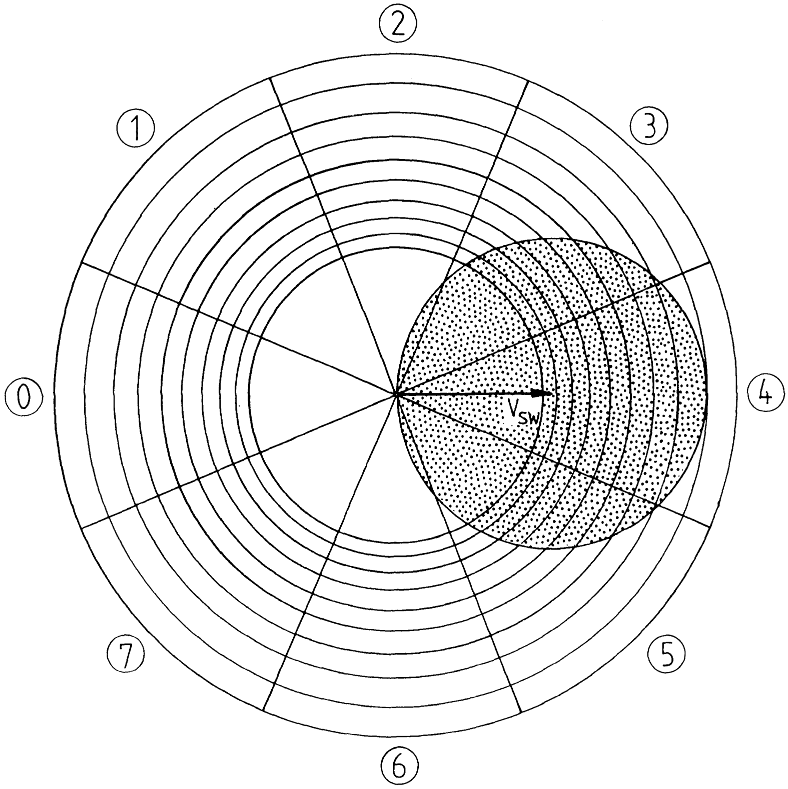


Fig. 4. Cut through the pick-up ion distribution function in the plane perpendicular to the spacecraft spin together with the sectoring scheme (radial fans) and the energy stepping (concentric rings) of the SULEICA instrument.

Finally, a second rotation into the natural frame of reference of the pick-up ion distribution function (v, Θ, Φ), i.e., with the z-axis parallel to the magnetic field, is applied. Due to the gyrotropic distribution ($\partial f/\partial\Phi = 0$) the calculation is significantly simplified in this way. From the ion distribution, as derived in Section 2, the differential energy flux density $E\,\Delta J/\Delta E$ is computed according to the relation

$$\frac{E\,\Delta J}{\Delta E} = \frac{v_{sw}^4}{\Delta\Omega\,\Delta E}\int f'(\mu', w')\,(m/2)\,w'^2\,w'2\pi\,\mathrm{d}\mu'\,\mathrm{d}w' \,. \tag{14}$$

The integration is performed over individual instrument channels. The simulation is performed using the actual parameters in the solar wind (solar wind velocity v_{sw} and the azimuth Φ_B and elevation λ_B of the magnetic field), as obtained by the 3D-plasma instrument (Paschmann *et al.*, 1985) and the flux-gate magnetometer (Lühr *et al.*, 1985) on the AMPTE/IRM satellite. It should be noted here that the appropriate quantity to determine the local source strength $S(R)$ is, indeed, the differential energy flux. From

Equations (12) and (14) it can be seen that the relation between the source strength and the differential energy flux density is independent of the solar wind velocity.

The computed quantities can now be compared with the corresponding values, as measured by the SULEICA instrument in the individual sectors and energy steps. Energy and angular response, as well as detection efficiencies of the instrument have been applied already for the determination of the measured differential energy flux densities. For a comparison with the model calculation the data of one pass of the satellite (≈ 8–10 hours of data) are rearranged in terms of the same solar wind velocity and the angle between the solar wind flow direction and the magnetic field vector. In this way pick-up ion distributions for a specific well-defined set of solar wind parameters with reasonably good counting statistics can be compared with the model.

5. Comparison with Measured Energy Spectra

In order to allow a comprehensive comparison of a measured energy spectrum with the stimulated energy spectrum, the data have been selected for:

– Periods with high solar wind velocities (>600 km s^{-1}): since the energy cut-off scales with the solar wind velocity and the SULEICA instrument is limited to energies above 5 keV e^{-1}, the most complete energy spectra can be obtained during periods with high solar wind velocities.

– Periods during November and December: during the early December days the Earth is moving through the gravitational focusing cone of the interstellar gas, where the neutral density is significantly enhanced. This leads to a substantial increase of the pick-up flux (Möbius *et al.*, 1985a) and, therefore, to increased ion count rates.

– Periods, when almost no energetic ions were observed upstream of the Earth's bow shock: although the pick-up ion distribution can be clearly identified in the presence of upstream ions, signatures like the energy cut-off and the angular distribution can be studied best during their absence.

In Figure 5 the measured energy flux density spectra in sectors 3, 4, and 5 are shown for samples with two different interplanetary magnetic field orientations on November 15, 1985, during the time period 3 : 00 UT to 11 : 00 UT. During this day the angle between the solar wind flow and the magnetic field varied from about 70° tc 150°, and the solar wind velocity covered about 630 to 690 km s^{-1}. The two samples were accumulated for a solar wind velocity range of 670 to 690 km s^{-1} and angular ranges of 67.5° to 112.5° and 112.5° to 157.5, denoted as 90° in the upper three panels and as 135° in the lower three panels of Figure 5, respectively. The data are shown above 8 keV e^{-1}. At lower energies the He^+ channel, as derived in the on board data handling of the instrument, does not match exactly with the mass/charge range of these ions and, therefore, is also sensitive to other ions.

As expected from the schematic of the distribution shown in Figure 4, significant fluxes of the pick-up ions are indeed only found in the Sun-sector and the two adjacent sectors (cf. Möbius *et al.*, 1985a, Möbius, 1986). In the Sun-sector (sector 4) the energy flux density shows a cut-off at 33 keV e^{-1} and a plateau between 30 and 16 keV e^{-1},

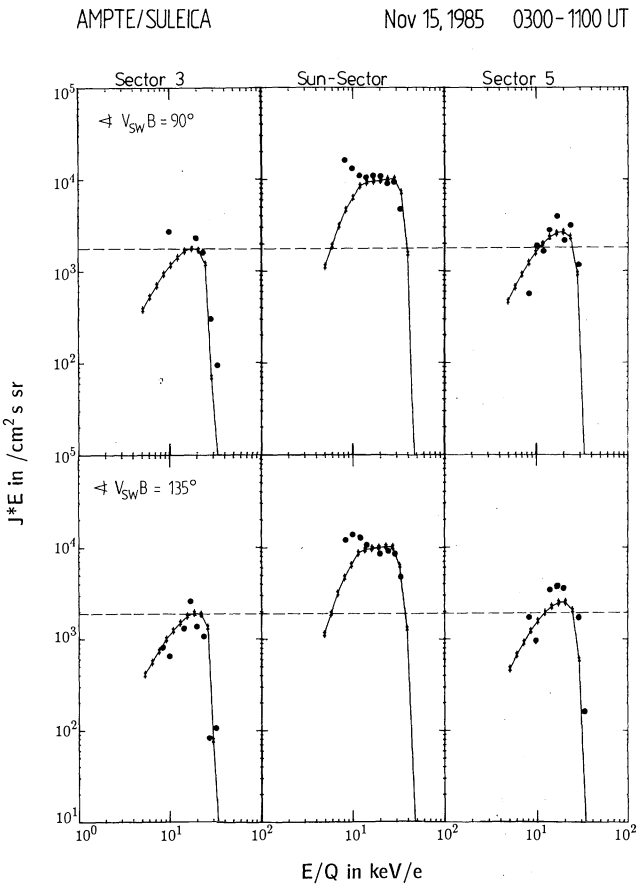


Fig. 5. Comparison of simulated energy flux density spectra (stars) for the actual solar wind parameters (v_{sw} = 680 km s^{-1}, angle ϕ between solar wind and magnetic field direction 90° in the upper panels and 135° in the lower panels) with spectra as measured by the SULEICA instrument for the Sun-sector and the two adjacent sectors.

which are both more pronounced for the 90° orientation. Near 10 keV e^{-1} a clear increase of the ion flux by about 50% is seen. However, at least part of this increase has to be attributed to solar wind ions: due to the limited resolution of the instrument at low energie, also heavy ions like, e.g., O^{5+} and O^{6+} contribute to the M/Q = 4 count rate at $E \approx$ 10 keV e^{-1} for solar wind velocities of $\approx$680 km s^{-1}. In the two adjacent sectors the pick-up ions form spectra, which peak between about 15 and 20 keV e^{-1} with a cut-off at $\approx$25 and $\approx$28 keV e^{-1} in sectors 3 and 5, respectively. The energy flux density in these sectors is generally lower by a factor of 6 in sector 3 and a factor of 3 in sector 5 compared to the Sun-sector, i.e., the pick-up ion distribution shows a significant anisotropy with respect to the Sun direction.

The model spectra in Figure 5 have been calculated with a neutral density variation according to Equation (11) with a typical penetration distance Λ = 0.5 AU, which corresponds to an ionization rate $\nu_{\mathrm{ion}} = 6.5 \times 10^8\ \mathrm{s}^{-1}$ at 1 AU and a velocity of the interstellar wind of 20 km s^{-1}. The spectra were normalized to the measured ion fluxes such that the plateau in the Sun-sector matches that of the measured distributions. From this procedure the product $N_0\ \nu_{\mathrm{ion}} = 4.2 \times 10^{-10}\ \mathrm{cm}^{-3}\ \mathrm{s}^{-1}$ of the helium density at infinity N_0 times the ionization rate ν_{ion} at 1 AU was derived. With ν_{ion} as given above this is equivalent to a neutral He density $N_0 \approx 0.0065\ \mathrm{cm}^{-3}$. In addition, the direction of the solar wind, as taken in the model calculation, has been corrected for the aberration due to the motion of the Earth ($\approx 3°$ for the actual solar wind velocity). Both, the measured and the model spectra are in reasonably good agreement. All the main features of the measured pick-up ion distribution, as the cut-off, the plateau and the flux radio between the Sun-sector and the two adjacent sectors, are well represented in the model spectra. The anisotropy between sectors 3 and 5 seems to be somewhat larger in the measured spectra. Furthermore, it should be noted here that according to Equation (8) the spectral shape depends crucially on the assumptions made for the source distribution, which will be discussed in detail below.

6. Discussion

There are several processes which may add to the observed anisotropy in sectors 3 and 5. They are illustrated in terms of the present model in Figure 6. The upper three panels show the distribution, as derived from Equation (12) for the solar wind blowing into a purely radial direction and a magnetic field orientation of 135° with respect to the solar wind. The injection of pick-up ions into the solar wind with a pitch-angle of 45° leads to higher fluxes in sector 3, which is scanning the distribution function perpendicular to the magnetic field, while sector 5 scans the part of the distribution mainly parallel to the field. A finite mean free scattering length λ produces higher phase space densities perpendicular to the magnetic field in the outer shells of the velocity distribution (cf. Equation (12) and Figure 2). The resulting anisotropy is of the order of 15%. The three center panels of Figure 6 show a distribution, which includes the effects of the adiabatic focusing in the limit of a stationary distribution. In this limit the solution for the distribution on a velocity shell as given in Equation (5) applies. As a result the phase space density is increased in the direction of the diverging magnetic field, i.e., in the anti-solar direction. Therefore, the flux in sector 5, which covers the distribution function parallel to B is higher than the flux in sector 3. The resulting anisotropy is of the order of 20%. These two effects act on the distribution function at the same time, but have only been described separately for illustration. In a realistic approach they are expected to almost cancel each other.

It should be noted here that the effect of adiabatic focusing gives also a handle on the scattering mean free path λ, since the lowest ion flux in the Sun-sector would be expected for a magnetic field orientation of 90° keeping all other parameters, like solar wind velocity and ionization rate, constant. For orientations of 135° or 180° the flux would be increased, more and more pronounced with increasing values of λ. According to (5) and (14) a 20% difference between the two orientations in Figure 5 would be

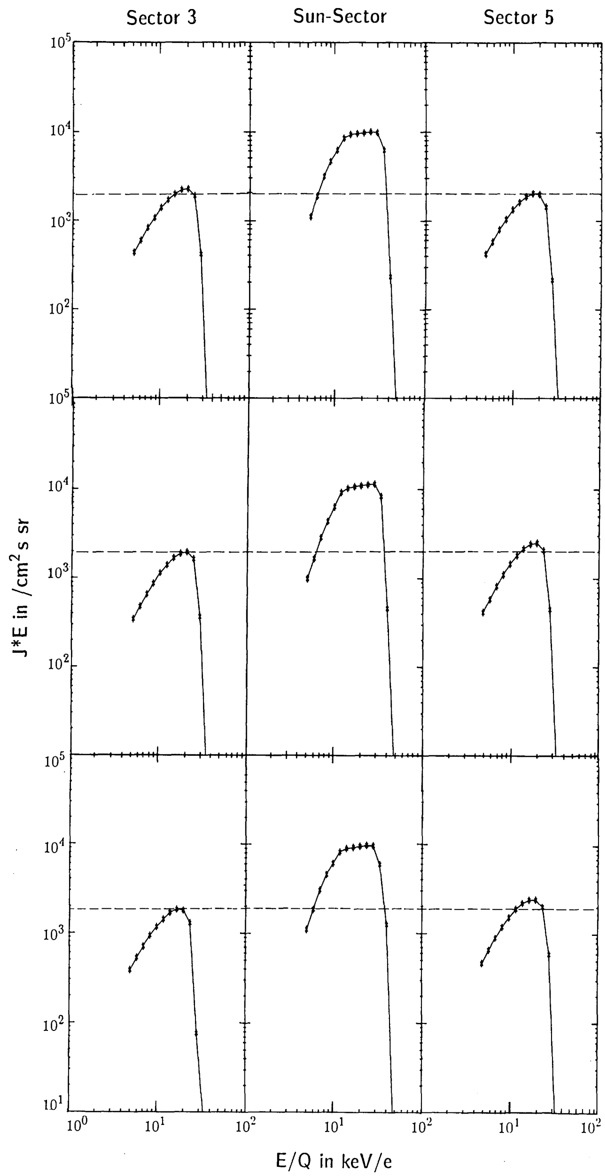


Fig. 6. Simulated energy flux density spectra for the Sun-sector and the two adjacent sectors with $\phi = 135°$ (left) and $\phi = 225°$ (right) under various assumptions. From top to bottom: (a) Time-dependent pitch-angle scattering without adiabatic focusing, solar wind direction 180°. (b) Stationary distribution after pitch-angle scattering with adiabatic focusing, solar wind direction 180°. (c) Time dependent pitch-angle scattering without adiabatic focusing, solar wind direction 177° (mean aberration of the solar wind due to the motion of the Earth).

expected for $\lambda = 0.2$ AU. The experimental results are compatible with values $\lambda \leq 0.1$ AU.

In the lower three panels of Figure 6 an aberration by 3° of the solar wind from the radial direction to 177° in the GSE coordinate system has been applied. This is equivalent to the aberration introduced by the motion of the Earth around the Sun for the given solar wind velocity. The resulting anisotropy between sectors 3 and 5 is a factor of 1.5. It is evident that this is the dominant effect for the anisotropy of the ion distribution. In addition, the direction of the solar wind velocity is variable within a few degrees (e.g., Pizzo *et al.*, 1983) which may lead to larger anisotropies.

As stated already above the spectral shape of the pick-up ions depends crucially on the actual source distribution of the neutrals. The influence of different variations of the neutral density with distance from the Sun is illustrated in Figure 7 for four values of the penetration distance Λ for interstellar helium (0.5, 0.7, 1.0, and 1.5 AU). The measured energy spectra seem to be best represented by the lowest value of Λ. This value is equivalent to an ionization rate $\nu_{ion} = 6.5 \times 10^{-8}$ s^{-1} at 1 AU, which has been given as a typical value under solar minimum conditions by Rucinski (1985), and a velocity of 20 km s^{-1} for the interstellar wind. The latter value is consistent with the relative velocity of the Sun and the interstellar medium, as derived from the Lα background radiation (e.g., Bertaux and Blamont, 1971; Chassefiere *et al.*, 1986). It should be noted here that the present measurements have indeed been obtained during the solar minimum. The larger values of Λ correspond to increasing values of the ionization rate up to 2×10^{-7} s^{-1}, which is, according to Ruczinski (1986), the value for the solar maximum. Thus a significantly different energy spectrum of interstellar He ions can be

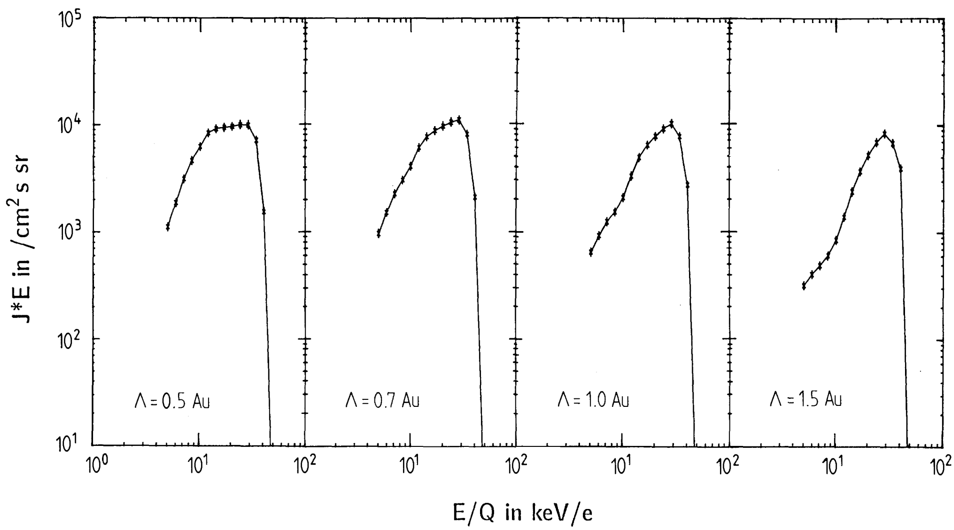


Fig. 7. Simulated energy flux density spectra in the Sun-sector for different radial distributions (see text) of the interstellar He density in interplanetary space from left to right: typical penetration distance Λ = 0.5, 0.7, 1.0, and 1.5 AU.

expected for the solar maximum. Although we may put constraints on the ionization rate from the comparison with measured spectra, the method is not sufficiently accurate as a quantitative measurement of $\nu_{\rm ion}$, since the velocity distribution of pick-up ions also depends on its evolution in velocity space with distance from the Sun (cf. Equation (8)).

In the present model calculation we have assumed that the solar wind velocity does not vary with distance from the Sun. However, a radial variation may substantially change the mapping relation for the radial source distribution. In the following we will discuss these effects.

The main acceleration of the solar wind occurs at a distance of a few solar radii from the Sun. However, there is still a moderate increase of the solar wind velocity at distances between 0.3 to 1 AU, which are of interest here. Hundhausen (1972) discussed a model, which shows a velocity increase of $\approx 20\%$ from 0.3 AU to the Earth's orbit. Statistical studies of the HELIOS solar wind data, which were obtained between 0.3 and 1 AU, by Schwenn *et al.* (1981) showed that this is an upper limit for the observed variations. These authors find a velocity increase between 4 and 17% for high solar wind speeds and almost constant velocities with distance from the Sun for low solar wind speeds. In order to demonstrate the effect of the solar wind velocity variation on the pick-up ion distribution we have plotted in Figure 8 the energy spectra in the Sun-sector for $\Lambda = 1.0$ AU for a constant solar wind velocity (left panel) and a solar wind velocity, which varies according to Hundhausen (1972) (right panel). As can be seen from

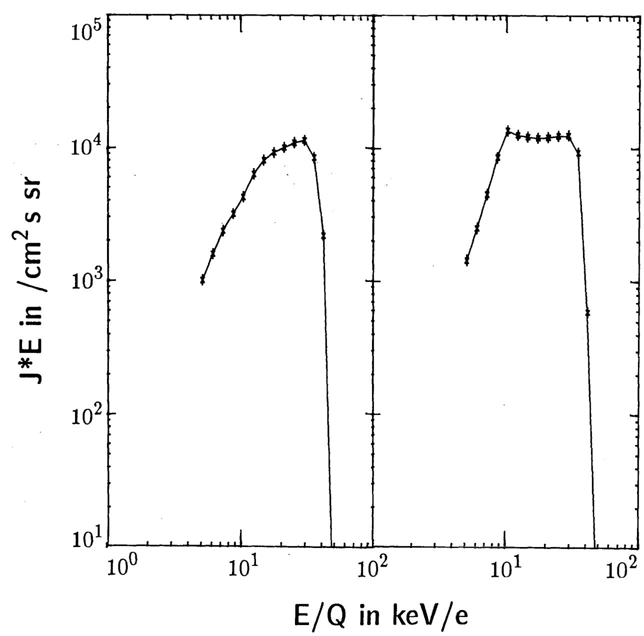


Fig. 8. Simulated energy flux density spectra for a model, which includes the variation of the solar wind velocity with distance from the Sun for $\Lambda = 1.0$ AU. *Right panel*: Variation of the solar wind speed v_{sw} after Hundhausen (1972). *Left panel*: Constant solar wind speed.

Figure 8 the main effect is a flattening of the spectrum towards lower energies. This is due to the fact that the shells in velocity space, which originate from a distance closer to the Sun, are nested closer (dv/dR larger) than the outer shells and, therefore, the phase space density of the inner shells is increased. We have used the maximum possible value for the increase of the solar wind velocity with radial distance from the Sun. In general, the effect may be smaller according to the experimental results by Schwenn *et al.* (1981).

With a UV ionization rate $\nu_{\text{ion}} = 6.5 \times 10^{-8}\ \text{s}^{-1}$ for solar minimum conditions in 1985 the measured value of the ion flux is compatible with an interstellar density $N_0 = 0.0065\ \text{cm}^{-3}$. The data were obtained during a period when the Earth was approaching the gravitational focusing cone of the interstellar gas. Therefore, the local density is enhanced compared to the He density outside the heliosphere. It should be noted here that the simple estimate as given by Möbius *et al.* (1985a) is in good accordance with this model. Furthermore, the energy flux density of pick-up ions is independent of the interplanetary parameters, such as magnetic field and solar wind velocity, as has been demonstrated in Section 4. Thus the model calculation, as presented in this paper, forms a diagnostic tool for the quantitative determination of the interstellar gas density and the interstellar helium temperature from the evaluation of the cone structure by means of He^+ pick-up ions. Such an investigation is currently in preparation.

Finalizing our discussion we would like to repeat the limitations of the present model: the model was derived under the strict assumption of a separability of the variables v (ion velocity) and μ (pitch-angle). In particular, the adiabatic deceleration has been treated under the assumption of instantaneous pitch-angle scattering to isotropy. In a more realistic approach one would have to admit that pitch-angle scattering takes a finite time and adiabatic deceleration of magnetized ions works solely perpendicular to the magnetic field. The information of the velocity decrease is transmitted parallel to B via pitch-angle scattering. Qualitatively the combination of pitch-angle scattering and adiabatic deceleration would, therefore, form an elipsoidal shell which is squeezed perpendicular to B rather than an ideal spherical shell distribution. The deviation from the spherical symmetry is still small at 1 AU (of the order of $2\lambda/R = 0.1$), but would become excessively large much closer to the Sun. The adiabatic deceleration becomes more and more inefficient. In this regime the separation of variables, as used here, is not allowed any more. Since interstellar He atoms even reach distances as close as 0.3 AU to the Sun, a rigorous treatment of the distribution function including simultaneously, becomes necessary for the description of the innermost portion of the velocity distribution function. The treatment of this paper can only illustrate trends in this regime. However, for the main portion of the distribution, which originates from distances > 0.3 AU, the model is fully adequate.

7. Summary

We have simulated the energy spectra of pick-up ions in the solar wind with an analytical model and compared those with measured spectra. We have shown by modeling of pitch-angle scattering and adiabatic deceleration as well as adiabatic focusing that the mean free path is less than 0.1 AU.

The energy spectra of the pick-up ions can be modelled by incorporating the radial dependence of the solar wind velocity and of the source strength. The shape of the energy spectra puts constraints on the ionization rate of interstellar helium and is compatible with values as low as $6.5 \times 10^{-8}\,s^{-1}$ as has been discussed for solar minimum conditions. The observed anisotropy can be understood by the motion of the spacecraft relative to the solar wind and a possible longitudinal component of the solar wind velocity. The energy flux density turns out to be independent of the interplanetary parameters as the magnetic field and the solar wind velocity and can thus be used to determine the absolute value of the source strength at 1 AU. Together with a measurement of the EUV flux the local interstallar gas density can be derived.

Acknowledgements

The authors are grateful to the many individuals at the Max-Planck Institut and the University of Maryland who contributed to the success of the SULEICA instrument and the AMPTE/IRM satellite with hardwave and software. We thank H. Lühr for the magnetic field data and G. Paschmann for the plasma data, and M. A. Lee for helpful comments on the manuscript.

References

Alfvén, H.: 1954, *On the Origin of the Solar System*, Oxford University Press, New York.
Axford, W. I.: 1972, in C. P. Sonnett, P. J. Coleman, Jr., J. M. Wilcox (eds.), *The Solar Wind*, NASA-SP308, 609.
Bertaux, J. L. and Blamont, J. E.: 1971, *Astron. Astrophys.* **11**, 200.
Blum, P. W. and Fahr, H. J.: 1969, *Nature* **223**, 936.
Chassefiere, E., Bertaux, J. L., Lallement, R., and Kurt, V. G.: 1986, *Astron. Astrophys.* **160**, 229.
Dalaudier, F., Bertaux, J. L., Kurt, V. G., and Mironova, E. N.: 1984, *Astron. Astrophys.* **134**, 171.
Fahr, H. J.: 1968, *Astrophys. Space Sci.* **2**,474.
Fahr, H. J.: 1974, *Space Sci. Rev.* **15**, 483.
Fahr, H. J., Ripken, H. W., and Lay, G. J.: 1981, *Astron. Astrophys.* **102**, 359.
Feldman, W. C., Asbridge, J. R., Bame, S. J., and Kearney, P. D.: 1974, *J. Geophys. Res.* **79**, 1808.
Hinteregger, H. E.: 1976, *Atm. Terr. Phys.* **38**, 791.
Holzer, T. E.: 1972, *J. Geophys. Res.* **77**, 5407.
Holzer, T. E.: 1977, *Rev. Geophys. Space Phys.* **15**, 467.
Hundhausen, A. J.: 1972, *Coronal Expansion and Solar Wind*, Springer Verlag, Berlin.
Isenberg, P. A.: 1987, *J. Geophys. Res.* **92**, 1067.
Jokipii, J. R.: 1971, *Rev. Geophys. Space Phys.* **9**, 27.
Lee, M. A. and Ip, W.-H.: 1987, *J. Geophys. Res.* **92**, 11041.
Lühr, H., Klöcker, N., Oelschlegel, W., Häusler, B., and Acuna, M.: 1985, *IEEE Trans. Geos. Remote Sensing*, **GE-23**, 259.
Mason, G. M., Gloeckler, G., and Hovestadt, D.: 1983, *Astrophys. J.* **267**, 844.
Möbius, E.: 1986, *Adv. Space Res.* **6**, 199.
Möbius, E., Hovestadt, D., Klecker, B., Scholer, M., Gloeckler, G., and Ipavich, F. M.: 1985a, *Nature* **318**, 426.
Möbius, E., Gloeckler, G., Hovestadt, D., Ipavich, F. M., Klecker, B., Scholer, M., Arbinger, H., Höfner, N., Künneth, E., Laeverenz, P., Luhn, A., Tums, E. O., and Waldleben, H.: 1985b, *IEEE Trans. on Geos. Remote Sensing* **GE-23**, 274.

Möbius, E., Hovestadt, D., Klecker, B., Scholer, M., Gloeckler, G., Ipavich, F. M., and Lühr, H.: 1986, *J. Geophys. Res.* **91**, 1325.

Paschmann, G., Loidl, H., Obermayer, P., Ertl, M., Laborenz, R., Sckopke, N., Baumjohann, W., Carlson, C. W., and Curtis, D. W.: 1985, *IEEE Trans. Geos. and Remote Sensing* **GE-23**, 259.

Petelski, E. F., Fahr, H. J., Ripken, H. W., Brenning, N., and Axnäs, I.: 1980, *Astron. Astrophys.* **87**, 20.

Pizzo, V. J., Schwenn, R., Marsch, E., Rosenbauer, H., Mühlhäuser, K.-H., and Neubauer, F. M.: 1983, *Astrophys. J.* **271**, 335.

Roelof, E. C.: 1969, in H. Ögelman and J. R. Wayland (eds.), *Lectures in High Energy Astrophysics*, NASA-SP199, 111.

Ruczinski, D.: 1986, Ph.D. Thesis, Warszawa.

Schwenn, R., Mühlhäuser, K.-H., Marsch, E., and Rosenbauer, H.: 1981, in H. Rosenbauer (ed.), *Solar Wind IV*, MPAE-W-100-81-31, p. 126.

Semar, C. L.: 1970, *J. Geophys. Res.* **75**, 6892.

Thomas, G. E.: 1978, *Ann. Rev. Earth Planetary Sci.* **6**, 173.

Thomas, G. E., and Krassa, R. F.: 1971, *Astron. Astrophys.* **11**, 218.

Vasyliunas, V. M. and Siscoe, G. L.: 1976, *J. Geophys. Res.* **81**, 1247.

Weller, C. S., and Meier, R. R.: 1974, *Astrophys. J.* **193**, 471.

Winske, D., Wu, C. S., Li, Y. Y., Mou, Z. Z., Guo, S. V.: 1985, *J. Geophys. Res.* **90**, 2713.

Wu, C. S., and Davidson, R. C.: 1972, *J. Geophys. Res.* **77**, 5399.

Wu, C. S., Hartle, R. E., and Ogilvie, K. W.: 1973, *J. Geophys. Res.* **78**, 306.

THE EVEN-ODD SYSTEMATICS IN *R*-PROCESS NUCLIDE ABUNDANCES*

K. MARTI and H. E. SUESS

Chemistry Department, University of California, San Diego, La Jolla, California, U.S.A.

(Received 31 July, 1987)

"... A genesis of the elements such as is sketched out would not be confined to our little Solar System, but would probably follow the same general sequence of events in every center of energy now visible as a Star."
Sir William Crookes (1886)

Abstract. We report and discuss solar system N_R abundances for nuclides $A > 70$, obtained as differences between measured solar system abundances and calculated *S*-process contributions. The abundance peak at $A \approx 163$ in the rare Earth element region reveals properties which are similar to those of the *R*-process peaks corresponding to magic neutron numbers $N = 82$ and $N = 126$. We observe that systematic differences in the N_R abundances of even-A and odd-A nuclides are restricted to specific mass regions. We discuss possible interpretations and conclude that these differences are most probably related to the properties of nuclear species during β^--decay to the stability valley.

1. Introduction

Forty years ago only limited information was available regarding solar, or stellar, elemental abundances. Yet, based on available information, many authors (e.g., Goldschmidt, 1938) suspected that these 'primordial' nuclear abundance values were determined by nuclear properties. Suess (1947) showed that this was indeed the case, as certain regularities in the abundances of odd-A isotopes could be discerned, such as that of the abundances of odd-A isotopes in the mass range $A = 135$ to 180 which exhibited remarkably smooth distributions. This can best be seen for the mass region of the rare Earth elements that are relatively unfractionated by chemical processes in nature. Suess showed that with the assumption of an essentially smooth dependence of odd-A abundances on mass number, it was possible to adjust the Goldschmidt (1938) abundances of odd-A nuclides within limits of error in such a way that a smooth line was obtained. The adjusted values, in turn, allowed additional regularities for the even-A nuclei to be discerned (Suess, 1947, 1987). On the basis of additional trace-element analyses for meteorites, Suess and Urey (1956) used these regularities to derive improved estimates of what used to be called 'cosmic' abundances. This was achieved by fitting together on a logarithmic scale, like pieces of a jigsaw puzzle, the exactly known isotopic compositions of the elements.

When Suess and Urey (1956) published these revised estimates of elemental

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

abundances (now called Solar System abundances), they realized that these abundance values should soon lead to a fundamentally new insight into processes of nuclear synthesis. At that time the essential constancy of the isotopic composition of the elements seemed to indicate that the nuclides were generated by the same sequence of events. After many unsuccessful attempts (see Jensen and Suess, 1947) to understand the empirical abundance distribution in terms of one common genetic process, Burbidge *et al.* (1957, hereafter referred as B^2FH) and Cameron (1957) showed that an interpretation of the empirical solar system abundances was possible, if one assumed that the matter that surrounds us constitutes a mixture of several components with different genetic histories. The possibility had earlier been suggested by von Weizsäcker (1937), but had not been followed up any further. B^2FH suggested that nuclides with a mass $A > 70$ had predominantly formed by two different processes; a slow (S) and a rapid (R) process. They postulated specific astronomical environments for the formation of the medium-heavy and heavy nuclides ($A > 70$), that could have been realized in different types of stars (see, e.g., the review by Fowler, 1984). The above quote (Crookes, 1886) illustrates that, in connection with the question of the origin of the elements, stars and energies had already been contemplated at a time when nothing was known about elemental and isotope abundances. The following discussion of details in the abundance distribution of nuclides should help to further narrow down the assumptions for possible mechanisms of nuclear synthesis.

No doubt the riddle of the origin of the elements was solved in principle by Burbidge *et al.* (1957), who showed that the elements of which our solar system consists represent mixtures of different genetic components. For one of the two main components, the S-component, there exists a quantitative correlation of neutron capture cross-sections (for medium energetic neutrons of 30 keV), as measured in the laboratory, with the relative abundances of certain nuclear species, certain 'nuclides', in our solar system. For the second predominant component, the R-component, however, only qualitative correlations with nuclear structure can be seen. Therefore, data on the abundance distribution and regularities of this R-component are most valuable for the study of the origin of this component. We discuss here distinct features in this component derived by subtraction of other components from empirical solar system abundance values. A feature that so far has not been investigated is an even-odd effect in limited mass ranges, a feature that we believe resulted from processes during termination of a rapid neutron build up. We expect quantitative evaluation of pertinent data will eventually help to determine accurate parameters for 'rapid' nuclear synthesis.

2. Solar System Abundances

The remarkably uniform isotopic compositions of the elements demonstrate that the solar nebula from which the solar system formed constituted a surprisingly homogeneous mixture of its different genetic components. In the following, we use the elemental solar system abundances from compilations by Palme *et al.* (1981) and by Anders and Ebihara (1982), except for more recent values by Jochum *et al.* (1986) for

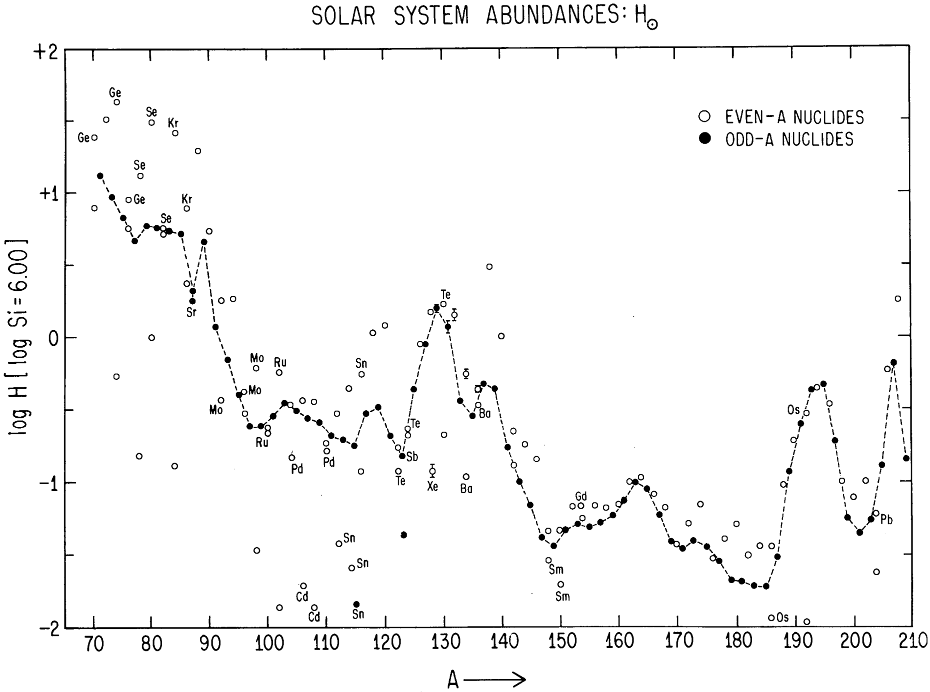


Fig. 1. Logarithms of solar system abundances (normalized to $H(Si) \equiv 10^6$) plotted for even mass-numbered (open circles) and odd mass-numbered (filled circles) nuclear species for $A = 70$ to 210 against their mass numbers. Odd-A values are connected.

Nb and Ta, and by Beer *et al.* (1983, 1984) for Hf and Hg. For Xe, the range indicated in Figure 1 covers the Anders and Ebihara value multiplied by 1.4 (viz., the observed 'solar' Kr/Xe ratio in gas-rich meteorites) and the Beer *et al.* value (calculated by σN_s interpolation). Figure 1 shows the solar system nuclear abundances for $A \geq 70$.

Some important features are easily recognized, especially as different symbols are used for even-A and for odd-A abundances. In the regions of the broad abundance maxima that are correlated with $N = 82$ and 126, no significant even-odd effects can be recognized. These maxima are due to high R-process yields along neutron shell closures in the regions of neutron-rich beta-unstable nuclides. The sharp abundance peaks at $A = 138$ and $A = 208$, however, reflect high S-process yields for $N = 82$ and 126, respectively. These double peaks due to high S- and R-process yields can already be seen in the abundance systematics as given by Suess (1947, 1949). They form the initial basis for the presumption of two different neutron capture components by Burbidge *et al.* (1957) and by Cameron (1957).

The Suess–Urey 'solar system' abundances have been 'revised' and updated by Cameron (1973, 1982), Palme *et al.* (1981), Anders and Ebihara (1982), and others. With a few exceptions, they can be assumed to be accurate within $\pm 20\%$. In addition to the prominent R- and S-process peaks, we can recognize fine structures in the mass

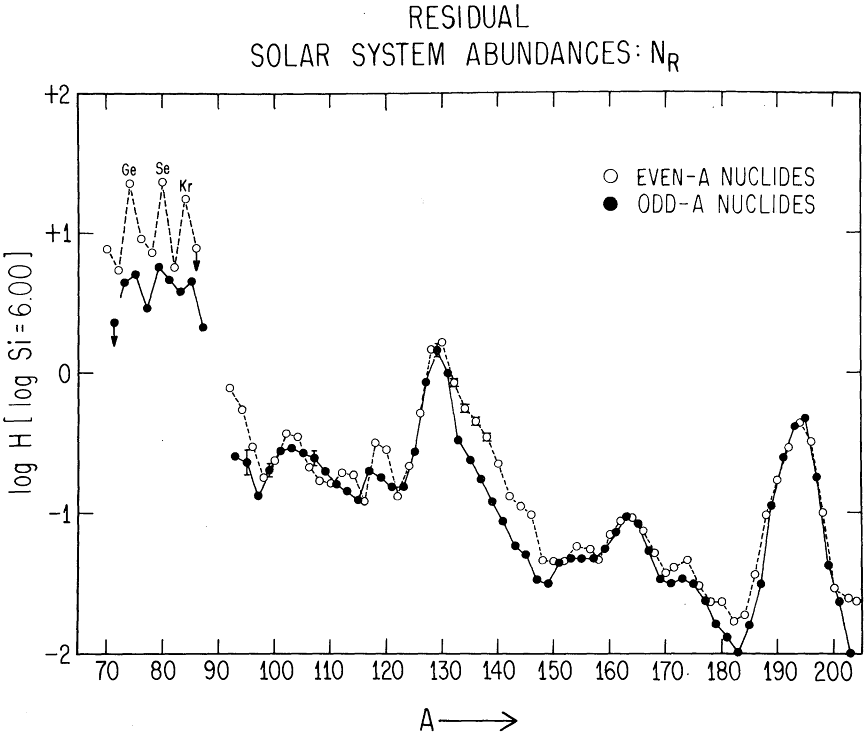


Fig. 2. Residual solar system abundances N_R plotted as in Figure 1. The plotted values correspond closely to R-process yields.

regions $A = 100–120$ and $A = 150–175$. The peak centered around $A = 163$ appears to match closely the characteristics of the R-process peaks, such as shape and vanishing even-odd effect, as discussed by Marti and Zeh (1985). The abundances in this region, which coincide with those of the rare Earth elements (REE), are an order-of-magnitude lower than those of the main R-peaks at $A = 130$ and 195 (see Figure 2).

3. *R*-Process – or 'Residual' Abundances

At least one other process, in addition to the R- and S-processes, is required to account for the synthesis of neutron-poor isotopes that are not the pathway of the S-process build-up (the so-called 'excluded' nuclides). Following established procedures, we divide the solar system abundances plotted in Figure 1 into S-, R- and P-components and obtain the R-component by subtraction. The P-abundances are relatively low ($< 1\%$), and their uncertainties have little effect on the R-abundances.

For careful evaluation of the R-component, reliable estimates of the S-component are important. Detailed calculations of S-abundance were carried out by Conrad (1976), Palme *et al.* (1981), Käppeler *et al.* (1982), and Beer and Macklin (1985). The recent neutron capture cross-sections by Bao and Käppeler (1986) were used to revise the σN_s

data. We interpolate locally for σN_s, assuming a simple seed distribution. In our elementary approach, we do not use a parameterized model with two distinct neutron flux distributions as done by Käppeler *et al.* (1982), nor do we consider specific seed nuclei distributions such as R-process products at the time of S-process synthesis. According to Amiet and Zeh (1968) and Conrad (1976), such models do not yield the empirical S-abundance distribution if a simple neutron flux distribution is assumed.

The calculated empirical R-abundances are listed in Table I and graphically shown in Figure 2. They are obtained as 'residuals' by subtracting the S- (and P-) components from the observed solar system abundances. Different symbols are used for even-A and for odd-A nuclides. They show similar abundance trends with marked differences in certain mass regions. Before we discuss these even-odd differences in detail, however, it seems appropriate to discuss the general features of these abundance 'residuals', which we denote here 'N_R-abundances'. Figure 2 shows that odd-A abundances never exceed the average abundance values of their even-A neighbours, except perhaps at $A = 107$ and 109. Both these numbers represent Ag-isotopes. Their solar system abundances may need to be redetermined carefully. From the σN-data for the two silver isotopes, it is to be expected that R- and S-components should be present in both isotopes, though hardly in the same ratios.

The agreement of the odd-A and the even-A abundance curves is best at and around the three maxima at $A = 130$, 163, and 195. Clearly, the effects from magic neutron numbers 82 and 126 are manifested in an equivalentl way by the abundance maxima at $A = 130$ and 195. A similar effect seems to have caused the maximum at $A = 163$, although its relationship with nuclear properties is less clear. This broad maximum at $A = 163$ in the rare Earth region has frequently been attributed (e.g., Steinberg and Wilkins, 1978) to a fission component. Marti and Zeh (1985), however, state that, for various reasons, a fission origin seems impossible. Most likely the maximum at $A = 163$ reflects enhanced R-process yields. In this case, the R-process path must pass through a region of increased nuclear stability, where neutron capture cross-sections are small. This might be the case in the regions of substantial nuclear deformations and calculated minima in the energy surfaces in the mass regions of helf-filled shells (Hamilton, 1985; Ragnarsson and Sheline, 1984). In any case, it is to be expected that shell effects on R-process yields are fundamentally different from those on fission yields.

Figure 2 also indicates that the effect of magic number $N = 50$ upon R-process yields in this region is obscured by other processes (E-processes) that certainly predominate in the $A < 70$ region. The broad peak or plateau at $A = 74$–84 may be associated with $N = 50$, but the nuclear abundance distribution differs considerably from those in higher mass regions. The curve for even-A nuclides of Ge, Se, and Kr (Figure 2) shows strikingly high abundances of the second heaviest isotopes ^{74}Ge, ^{80}Se, and ^{84}Kr.

4. The Even-Odd Effect in R-Process Yields

Even-A nuclear abundances are, in general, more abundant than their odd-A neighbours. This is because of the smaller neutron capture cross-sections of the even-A

TABLE I

Solar system R-nuclide abundances and even-odd ratios (Si = 10^6)

A	Element	N_R	$\frac{N_R}{N_\odot}$	$\frac{N_R\,(\text{even})}{N_R\,(\text{odd})}$	A	Element	N_R	$\frac{N_R}{N_\odot}$	$\frac{N_R\,(\text{even})}{N_R\,(\text{odd})}$
72	Ge	5.5	0.17		129	Xe	1.31–1.61	0.95	1.00–1.22
73	Ge	4.5	0.49	3.1	130	Te	1.69	1.00	
74	Ge	22.9	0.53		131	Xe	0.978–1.24	0.91	1.14–1.28
75	As	5.0	0.74	3.0	132	Xe	0.806–1.14	0.6–0.7	1.24–1.46
76	Ge	9.2	1.00		133	Cs	0.321	0.86	2.05–2.69
77	Se	3.0	0.64	2.8	134	Xe	0.509–0.589	1.00	
78	Se	7.4	0.51		135	Ba	0.236	0.87	1.97–2.26
79	Br	5.6	0.93	2.8[1]	136	Xe	0.420–0.476	1.00	
80	Se	23.8	0.77	[2]	137	Ba	0.178	0.36	2.08–2.35
81	Br	4.9	0.82	3.1	138	Ba	(0.32–0.36)		
82	Se	5.71	1.00	1.3	139	La	0.118	0.26	2.33
83	Kr	3.8	0.73	3.1	140	Ce	(0.023)		
84	Kr	18.0	0.73	4.3	141	Pr	0.087	0.50	2.06
85	Rb	4.5	0.88	2.0	142	Ce	0.129	1.00	
95	Mo	0.185–0.285	0.57	≥1.50	143	Nd	0.058	0.57	2.08
96	Zr	0.300	1.00		144	Nd	0.112	0.56	
97	Mo	0.129	0.53	≥1.87	145	Nd	0.050	0.72	1.99
98	Mo	0.182	0.30		146	Nd	0.087	0.60	
99	Ru	0.181–0.288	0.181–0.288	0.77–0.97	147	Sm	0.0319	0.79	2.11
100	Mo	0.242	1.00		148	Nd	0.0477	1.00	
101	Ru	0.268	0.85	1.13	149	Sm	0.0306	0.84	1.54
102	Ru	0.366	0.62		150	Nd	0.0468	1.00	
103	Rh		0.85	1.22	151	Eu	0.0446	0.96	1.05
104	Ru	0.348	1.00		152	Sm	0.0464	0.67	
105	Pd	0.269	0.87	1.05	153	Eu	0.0479	0.95	1.10
106	Pd	0.214	0.57		154	Sm	0.0589	1.00	
107	Ag	0.217–0.269	0.79–0.98	0.75–0.93	155	Gd	0.0460	0.94	1.25
108	Pd	≥0.167 (0.189)	0.45		156	Gd	0.0557	0.82	
109	Ag	0.196	0.77	(0.90)	157	Gd	0.0470	0.90	1.09
110	Pd	0.164	1.00		158	Gd	0.0470	0.37	
111	Cd	0.155	0.76	1.16	159	Tb	0.0551	0.93	1.08
112	Cd	0.196	0.51		160	Gd	0.0722	1.00	
113	Cd	0.137–0.145	0.71	(1.31–1.38[3])	161	Dy	0.0724	0.96	1.10
114	Cd	0.183	0.40	(1.44)	162	Dy	0.0877	0.87	
115	In	0.122	0.64	1.24	163	Dy	0.0938	0.95	0.97
116	Cd	0.119	1.00		164	Dy	0.0940	0.84	[6]
117	Sn	0.199	0.67	(1.10)	165	Ho	0.0831	0.95	1.02
118	Sn		0.35	(1.71)	166	Er	0.0748	0.89	
119	Sn	0.176	0.53		167	Er	0.0543	0.94	1.18
120	Sn	0.284	0.23	(1.73)	168	Er	0.0538	0.78	
121	Sb	0.152	0.75	(1.30)	169	Tm	0.0334	0.88	1.35
122	Sn	0.127	0.73	0.84[4]	170	Er	0.0377	1.00	[7]
123	Sb	0.150	0.99	1.14	171	Yb	0.313	0.89	1.24
124	Sn	0.215	1.00	1.03	172	Yb	0.0402	0.76	
125	Te	0.266	0.77	1.35	173	Yb	0.0334	0.85	1.31
126	Te	0.504	0.55	0.90	174	Yb	0.0473	0.62	
127	I	0.849	0.94	1.19	175	Lu	0.0315	0.88	1.24
128	Te	1.51	0.97	1.23–1.40[5]	176	Yb	0.0306	1.00	[8]

Table I (continued)

A	Element	N_R	$\frac{N_R}{N_\odot}$	$\frac{N_R\,(\text{even})}{N_R\,(\text{odd})}$	A	Element	N_R	$\frac{N_R}{N_\odot}$	$\frac{N_R\,(\text{even})}{N_R\,(\text{odd})}$
177	Hf	0.0244	0.87	1.12	191	Ir	0.242	0.98	0.96
178	Hf	0.0242	0.59		192	Os	0.293	1.00	(10)
179	Hf	0.0156	0.75	1.54	193	Ir	0.408	0.98	0.89
180	Hf	0.0239	0.45		194	Pt	0.434	0.96	
181	Ta	0.0129	0.65	1.58	195	Pt	0.458	0.99	0.82
182	W	0.0169	0.47		196	Pt	0.317	0.91	
183	W	0.0095	0.48	1.88	197	Au	0.176	0.95	1.18
184	W	0.0188	0.45		198	Pt	0.0986	1.00	
185	Re	0.0156	0.82	1.79	199	Hg	0.0422	0.74	1.52
186	W	0.0369	0.94	(9)	200	Hg	0.0293	0.37	
187	Re	0.0309	0.90	1.92	201	Hg	0.0237	0.53	1.15
188	Os	0.0820	0.86		202	Hg	0.0251	0.25	
189	Os	0.112	0.97	1.13	203	Tl	0.0090	0.17	(2.7)
190	Os	0.171	0.90		204	Hg	0.0235	1.00	

(1) S-branch = 44%.
(2) S-branch = 60%.
(3) S-branch < 14%.
(4) S-branch = 3%.
(5) S-branch = 6%.
(6) S-branch = 83%.
(7) S-branch = 0.14%.
(8) S-branch = 32%.
(9) S-branch = 7.8%.
(10) S-branch = 4.6%.

nuclides. Their S-process yields are higher than those of the odd-A species. For the R-process yields, resulting from rapid neutron build up, no such differences are expected because of the extreme conditions and decay chains which involve neutron losses. The concept of the R-process explains the most important abundance rule that states that odd-A abundances show a smooth dependence on mass number (Suess, 1947). This is because odd-A nuclides consist predominantly of the R-component.

Figure 3 shows the ratios of interpolated abundances to odd-A abundances as a function of A, calculated from the data listed in Table I and shown in Figure 2. It can be seen that, throughout the mass range of $A = 90$ to 200, these ratios are by no means always close to one, as required by the basic abundance rule. Only in about half the mass range do the R-process yields of even-A nuclides differ by less than 20% from those of the odd-A's. In the other half, they are about twice as high. In spite of many attempts, these deviations cannot be eliminated by assuming different elemental abundance values. It seems that even-odd effects may turn out to be most interesting intrinsic features that may reveal important information on R-process synthesis.

During S-process synthesis, the abundance N_s of a nuclide with mass number A changes with time t according to

$$\frac{\mathrm{d}N_t(A)}{\mathrm{d}t} = \lambda_n(A-1)N_s(A-1) - [\lambda_n(A) + \lambda_\beta(A)]N_s(A)\,,$$

where the neutron capture rate $\lambda_n = \phi\sigma$ is the product of neutron flux ϕ and of the

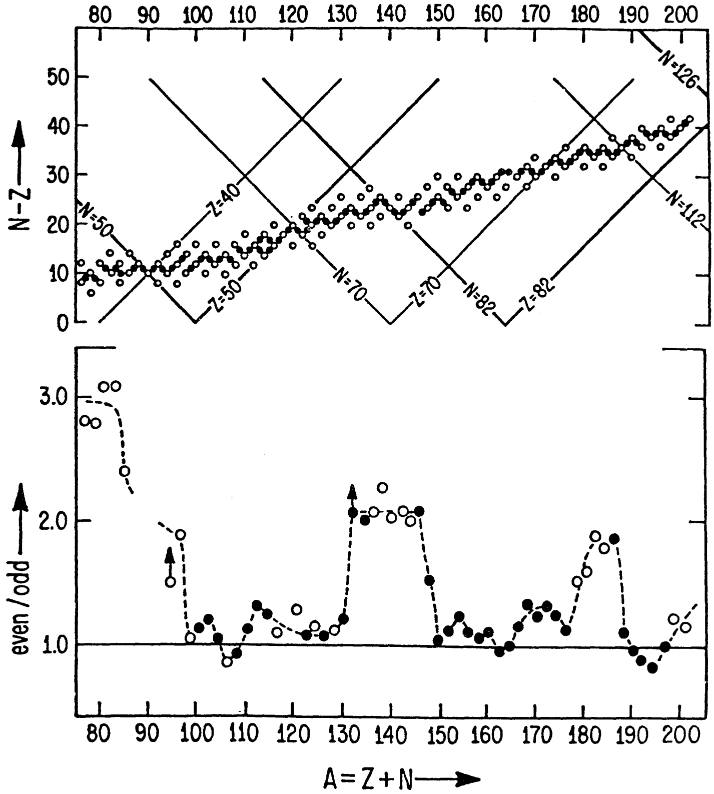


Fig. 3. Ratios of interpolated even-A N_R-abundances to odd-A N_R-abundances, plotted as in Figures 1 and 2.

average effective neutron cross-section σ, and λ_β is the β^--decay constant. The first and second terms in the above equation determine the gain and loss rates, respectively. This equation represents a system of coupled differential equations which cannot be solved in general, because the λ_n are time-dependent and require a knowledge of stellar temperatures and neutron fluxes during S-process synthesis. In order to solve the equations, two assumptions are made: first, it is assumed that $\lambda_n \ll \lambda_\beta$. This means that branching in the radioactive product nuclei in the pathway of S-process synthesis is generally neglected. Only very long-lived nuclides are treated as stable nuclei. In a few special cases, such as for ^{79}Se and ^{85}Kr, the effects of radioactive decay upon S-fields were discussed in detail (e.g., Seeger *et al.*, 1965; Käppeler *et al.*, 1982). Secondly, it is assumed that the temperature is constant during S-process synthesis, which permits the use of energy-averaged neutron cross-sections. With the additional assumption of an experimental distribution for the neutron flux, $\tau = \int \phi \, dt$, the system of coupled differential equations was solved (Clayton and Ward, 1978). By comparison of the calculated $N_s\sigma$ values with the empirical abundances of pure S-process nuclides, it became clear that at least two neutron components of distinct exponential flux distributions are necessary to fit the N_s abundances, even if one excludes the (recycled) lead and bismuth isotopes (e.g., Käppeler *et al*, 1982). In the two-component case, the fitting procedure was carried out in two steps. Käppeler *et al.* first fitted the σN_s curve (neglecting the second component) in the mass range $A > 100$, then repeated the procedure for $A < 97$ to obtain the parameters of the second (weak) component.

Because of the undoubtedly somewhat ambiguous S-process assumption, it is possible that our estimates of S-process yields might be too low so that the presence of some S-component might be responsible for the even-odd abundance differences in the N_R-yields. Indeed, these differences (Figure 3) are large in mass regions where the contributions of the S-component are relatively large. However, no quantitative correlation between the magnitude of the S-component and that of the even-odd effect can be discerned. The even-odd effect of the R-component appears conspicuously constant, as can be seen in Figure 2. There the N_R-abundances (the 'residuals') are plotted for even-A and for odd-A nuclei on a logarithmic scale, using different symbols. Between $A = 133$ and $A = 146$, the two types of nuclides plot on nearly parallel lines, indicating approximately a factor of two difference in N_R-abundance, as can be seen also in Figure 3.

Because of the lack of a correlation of the even-odd effect in R-yields with the magnitude of the S-contribution at the individual mass numbers, and because of the conspicuously regular even-odd effect of the R-process yields in certain mass ranges (Figure 3), we consider beta-delayed neutron emission during beta-decay to stability as the most probable cause of the even-odd effect in the residual N_r-yields. This situation is comparable with that of fission products during their series of beta-decays to ground states. There, neutron emission occurs instantaneously, if the excitation energy of a nucleus is larger than the binding energy of the 'last' neutron. In such cases, there is frequently one more than a 'magic' number of neutrons present in the nucleus. Hence, we may expect that even-odd effects will occur where the beta-decay chains transgress magic shell edges. In such cases, one would expect that a considerable fraction of such neutrons get 'lost' from the decay chains. Indeed, large emission rates were observed just beyond $N = 50$ for ^{83}Ga (43%), ^{85}As (23%), and ^{87}As (44%) (Browne and Firestone, 1986).

The details in the mass ranges $A = 132$–146, $A = 167$–175, and $A = 180$–185 suggest to us that the intrinsic abundance structure was established by one and the same process, R-synthesis. The even-odd-A differences appear to be due to decay-delayed neutron emission at $N \gtrsim 82$. Experimental data indicate that neutron emission is observed in nuclei between $N = 82$ and $N = 90$ (Browne and Firestone, 1986). For example, beta-delayed neutron emission rates in the Cs isotopes increase from 1.7% in ^{143}Cs and 13% in ^{145}Cs to 25% in ^{147}Cs. In this mass region, the nuclei are not spherically symmetric, but become increasingly deformed (Bengtsson *et al.*, 1984). Neutron emission rates in neutron-rich isotopes of $A > 150$ are not measured because such nuclides are not produced in fission reactions. Even-odd-A differences vanish exactly at the $A = 163$ peak, but reappear for masses $A > 165$. Large even-odd effects are observed in the range $A = 180–185$, which correspond to the transition region of deformed to spherical nuclei.

A further correlation with even-A and odd-A abundance differences in R-process yields may perhaps be recognized by considering the differences in the pairing energies of neutrons and protons. These differences are responsible for the nonexistence of the 'missing elements' containing 43 and 61 protons (Suess and Jensen, 1951). The differences in pairing energies are caused by the deformation of nuclei in the high spin regions. Indications of the existence of a correlation of mass ranges of deformed nuclides with those of even-odd effects in N_R yields have been discussed.

5. Concluding Remarks

The study of empirical solar system abundances in the past has allowed major advances in our understanding of nuclear properties (the shell model), as well as in the mechanisms of element synthesis. It now appears that the abundance peak at $A \approx 163$ in the REE region may be related to properties of the nuclear energy surfaces in the deformed mass region. Therefore, the study of nuclide abundances could provide one more significant contribution to the study of nuclear properties. Furthermore, the existence of an even-odd-A difference in the N_R-abundances is interpreted as a relict in the R-process component, and most probably the result of neutron emission during beta-decay to the stability valley.

Acknowledgements

We thank D. Zeh for many stimulating discussions and H. Beer for sending us a preprint of the neutron cross-section compilation. We also thank M. Carter for her compilation and her assistance with the figures. Work was in part supported by NASA grant NAG 9–41 to one of us (K.M.).

The authors wish to dedicate this paper to Professor Hannes Alfvén on the occasion of his 80th birthday, with thanks for many critical and stimulating discussions.

References

Amiet, J. P. and Zeh, H. D.: 1968, *Z. Phys.* **217**, 485.
Anders, Edward and Ebihara, Mitsuru: 1982, *Geochim. Cosmochim. Acta* **46**, No. 11, 2363.
Bao, Z. Y. and Käppeler, F.: 1986, *Atomic Data and Nuclear Data Tables* **36**, 411.
Beer, H. and Macklin, R. L.: 1985, *Phys. Rev.* **C32**, 738.
Beer, H., Käppeler, F., Reffo, G., and Venturini, G.: 1983, *Astrophys. Space Sci.* **97**, 95.
Beer, H., Walter, G., Macklin, R. L., and Patchett, P. J.: 1984, *Phys. Rev.* **C30**, 464.
Bengtsson, R., Möller, P., Nix, J. R., and Zhang, Jing-ye: 1984, *Phys. Scripta* **29**, 402.
Browne, E. and Firestone, R. B.: 1986, in V. S. Shirley (ed.), *Table of Radioactive Isotopes*, Wiley, Interscience, New York.
Burbidge, E. M., Burbidge, G. R., Fowler, W. A., and Hoyle, F.: 1957, *Rev. Mod. Phys.* **29**, 547.
Cameron, A. G. W.: 1957, *Publ. Astron. Soc. Pacific* **69**, 201.
Cameron, A. G. W.: 1973, *Space Sci. Rev.* **15**, 121.

Cameron, A. G. W.: 1982, in C. A. Barnes, D. D. Clayton, and D. N. Schramm (eds.), *Essays in Nuclear Astrophysics*, Cambridge University, Cambridge.

Clayton, D. D. and Ward, R. A.: 1978, *Astrophys. J.* **224**, 1000.

Conrad, J.: 1976, 'Analyse des *s*-Prozesses bei der Kosmischen Synthese schwerer Elemente', Thesis, Heidelberg.

Crookes, W.: 1886, *Nature* **34**, 423.

Fowler, William A.: 1984, *Rev. Mod. Phys.* **56**, No. 2, Part I, 149.

Goldschmidt, V. M.: 1938, *Geochem. Verteilungsgesetze der Elemente IX*, Skrifter Norske Videnskaps–Akademien, Oslo, I. Mat. Natur. Kl., No. 4.

Hamilton, J. H.: 1985, *Prog. Particle Nucl. Phys.* **15**, 107.

Jensen, H. and Suess, H.: 1947, *Naturwissenschaften* **34**, 131.

Jochum, K. P., Seufert, H. M., Spettel, B., and Palme, H.: 1986, *Geochim. Cosmochim. Acta* **50**, 1173.

Käppeler, F., Beer, H., Wisshak, K., Clayton, D. D., Macklin, R. L., and Ward, Richard A.: 1982, *Astrophys. J.* **257**, 821.

Marti, K. and Zeh, D.: 1985, *Meteoritics* **20**, 311.

Palme, H., Suess, H. E., and Zeh, H. D.: 1981, *Landolt-Börnstein*, **VI**, 2a, Springer-Verlag, Berlin.

Ragnarsson, I. and Sheline, R. K.: 1984, *Phys. Scripta* **29**, 385.

Seeger, P. A., Fowler, W. A., and Clayton, D. D.: 1965, *Astrophys. J. Suppl.* **11**, 121.

Steinberg, E. P. and Wilkins, B. D.: 1978, *Astrophys. J.* **223**, 1000.

Suess, Hans E.: 1947, *Z. Naturforsch.* **2a**, 311.

Suess, H. E.: 1949, *Experientia* **5**, 266.

Suess, H. E.: 1987, *Chemistry of the Solar System*, John Wiley and Sons, Inc., New York.

Suess, H. and Jensen, H.: 1951, *Arkiv Fysik* **3**, 577.

Suess, H. E. and Urey, H. C.: 1956, *Rev. Mod. Phys.* **28**, 53.

von Weizsäcker, C. F.: 1937, *Z. Phys.* **38**, 176.

THE SOLUBILITY PROBLEM OF HEAVY ELEMENTS IN INTERNAL STELLAR PLASMA REVISITED*

IMRE RUFF

Laboratory of Theoretical Chemistry, Eötvös University, Budapest, Hungary

GEORGE MARX

Department of Atomic Physics, Eötvös University, Budapest, Hungary

and

DAVID S. DEARBORN

Lawrence Livermore Laboratory, University of California, Livermore, California, U.S.A.

(Received 23 July, 1987)

Abstract. Theoretical predictions of the phase separation of heavy elements in internal stellar plasmas, based on the classical Debye–Hückel and mean spherical approximation, have offered a solution to the 'solar neutrino puzzle'. Recent contributions, however, expressed some doubt about the correctness of these calculations. In the present paper, we challenge the latter conclusions by showing that the linearization of the Poisson–Boltzmann equation can be preformed to avoid the negativity of pair correlation functions, and that the importance of classical charge-charge effects is considerably greater than quantum effects in determining the internal energy of solar plasmas. Comparison between gravity and radiation pressure acting on phase-separated high-Z plasma 'droplets' supports the formation of a small 'iron core' in the center of the Sun confirming Rouse's suggestion that the frequencies of the non-radial g-modes and the five-min band of oscillations in the Sun can be explained only by the existence of such a core. The depletion of the Sun's interior of heavy elements results in a decreased opacity and, consequently a higher temperature which finally leads to a chlorine neutrino signal of about 2.5 solar neutrino unit in agreement with Davies's experimental result. Essentially the same high neutrino capture rate as given by the present standard Sun model is predicted for the future gallium experiment. This prediction is in contrast to the neutrino oscillation hypothesis in which a wide range of coupling parameters suppresses both chlorine and gallium signals.

1. Introduction

In the present paper we revisit the problem: Are the heavy elements (and especially iron, that of the highest cosmic abundance) soluble in plasmas of the stellar interior?

The importance of the question is due to the consequence which the eventual insolubility of iron would have on the solar model. If the interior of the Sun is depleted in iron, the opacity and, consequently, the temperature of the solar interior would be lower than assumed so far, as a result of which the fusion of boron nuclei via lithium (which is the main source of the neutrinos observed by Davies, 1964) would be reduced, and as a final consequence the old 'neutrino dilemma' (measured flux = 2.5 SNU, calculated flux $\cong$ 6 SNU) could be resolved.

This line of logic was first suggested by Alder and Pollock who modeled the solar

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

internal plasma by a two-component system of point charges and calculated the excess free enthalpy of mixing by two terms: i.e.,

$$\Delta G^{\text{ex}} = \Delta G^{\text{ex}}_{\text{id.g.}} + \Delta G^{\text{ex}}_{\text{ch.}} . \tag{1}$$

The ideal gas term contains only the entropy of mixing of the components bearing kinetic energy, i.e., protons, iron ions, and electrons (the formally two-component system has, in fact, three real components restricted by the electroneutrality theorem). The portion of excess free energy of mixing due to charge-charge interactions is taken into account by $\Delta G^{\text{ex}}_{\text{ch}}$. This was calculated by the DH theory treating H^+, Fe^{z+} ($z = 26$ or 24), and e^- as point charges (cf. Pollock and Alder, 1978).

Their conclusion was that, although the point in the phase diagram representing the internal solar conditions corresponding to the present standard model does not lie under the spinoidal curve and, thus, the internal solar plasma is not instable, it does lie under the coexistence curve and, thus, the plasma is not in equilibrium but oversaturated by iron in a concentration corresponding to its cosmic abundance.

Pitzer (1980) criticized the former model on the ground that the DH point charge model yields negative pair correlation functions for like-charged particles which is unphysical. Instead, he suggested to replace the DH like-charged pair correlation function with a zero-function in the short-range region where the DH $g(r)$'s would be negative. Indeed, this alteration in the model could lead to the conclusion that iron was soluble in hydrogen plasmas in the entire concentration range. However, by simply omitting some of the positive particles, the charge density is changed at the expense of the violation of the electroneutrality theorem. Thus Pitzer's result is not convincing.

The non-negativity of the pair correlation functions can be achieved in a self-consistent way for both the DH and the MSA theory, as we will show in more details in Section 2. The error committed by the latter theories in the internal energy is by all means less than necessary to reverse the conclusion on the solubility of iron in solar plasmas.

Simultaneously, Alder *et al.* (1980) revisited the problem by a more sophisticated method, the hypernetted-chain (HNC) approximation. They modeled the solar plasma as a mixture of two one-component plasmas (OCP), i.e., as a binary ionic mixture (BIM), while electrons were regarded to form a homogeneous neutralizing background. This is a usual assumption in plasma theory and is quite adequate if the information to be extracted are based essentially on the interactions between positive ions. It is, however, a deficient model if the thermodynamic properties of a real plasma are to be obtained, since such a model would be obviously erroneous in predicting the fraction of internal energy due to repulsion between electrons and to attraction between ions and electrons.

The model of Gombert and Deutsch (1984) is a successful improvement of the DH approximation. They used effective potentials in place of some Coulomb terms. The excess free enthalpy of mixing is a function similar to that obtained by the DH calculation, viz., it is *positive* in the relevant low concentration range of iron. If they did not overstep the region near $x_{\text{Fe}} = 0$ even with the first points closest to it, they would have found the small minimum in the excess free enthalpy of mixing and would have concluded that iron had to be insoluble. Unexplicably, they did the opposite in spite of the positivity of the free enthalpy of mixing in the broad middle region.

Ruff *et al.* (1985) attempted to improve the DH model by using the MSA approximation and by assuming small hard-core diameters for iron or hydrogen ions within which other particles are excluded. Electrons were handled as classical point charges. The MSA model proved to be quite insensitive to the choice of the hard-core diameters within 20 pm when, self-consistently, the ideal gas term was simultaneously corrected by the Percus–Yevick (PY) hard-core real gas model. Thus, the conclusion was that the DH model is a reasonable approximation within a non-quantum statistical framework.

Based on the former results, Ruff and Liszi (1985) calculated the phase diagram of the hydrogen-helium-iron ternary mixture. The free enthalpy of mixing surface (as a function of composition) is concave in the sense that there are no three points to which a definite triple-tangent plane can be fitted. If the hydrogen-helium mixture of the composition corresponding to solar conditions is handled to form a pseudo-two-component mixture with iron, the solubility of iron is lowered even further.

In a recent paper, Iyetomi and Ichimaru (1986) aimed at improving the Alder–Pollock–Hansen model by replacing the uniform neutralizing background by a polarizable one in the BIM. Electrons were treated in the random phase approximation (RPA). Although their calculations indicate a considerable decrease in the solubility of iron ions in comparison with the unpolarizable background, it is not sufficient to expect phase separation in the center of the Sun. In Section 2 an attempt is made to reproduce these results by a suitably parametrized MSA. The failure of the reproducibility of the internal energy obtained by the BIM in spite of a very good agreement between the corresponding pair correlation functions indicates some discrepancy in the energy calculation of the BIM.

Returning to the original problem of the 'neutrino dilemma', we should like to refer to preliminary results of Dearborn *et al.* (1987) obtained by re-running the Livermore program of the standard solar model assuming the reduction of the internal opacity as a consequence of the immiscibility of heavy elements with the hydrogen plasma. As discussed in Section 3, due to the change in opacity the chlorine capture rate may drop from about 6 to 2.5 solar neutrino units, thus the discrepancy between theory and experiment can be essentially eliminated.

The eventual phase separation of heavy elements raises intriguing questions concerning the fate of the plasma 'droplets' of heavy elements which were the primary results of phase separation. Did they sink due to their mass density being larger than that of the surrounding hydrogen plasma, and do stars have an 'iron core' as a consequence? Or were they lifted by the radiation pressure to the convective outer layer of the stars where the temperature and density is low enough to make them dissolve? Or do they float as a result of the two opposite forces somewhere in an inner layer which would require to readjust opacity calculations accordingly? In Section 4 we present calculations to answer these questions.

Although calculations on the diffusion of heavy elements *dissolved* in the internal solar plasma (cf. Aller and Chapman, 1960) indicate that their concentration increases gradually towards the center of the Sun reaching a total of about 12% for iron, a sufficiently reliable calculation on almost pure iron 'droplets' seems to be worth while considering its crucial consequences on star models. Namely, if they prove indeed to be sunk to the center, the prediction on phase separation would give support to the assumption of Rouse (1983) on the 'high-Z, iron-like' core of the Sun. As she has shown, this assumption is a 'necessary condition for a unique solution of the stellar structure equations' in agreement with the experimental neutrino flux. On the other hand, observations on the non-radial g-modes and the structure of the five-min band of the acoustic and gravitational waves of the Sun are in excellent agreement with Rouse's predictions based on the same assumption (cf. Rouse, 1985, 1986).

2. Non-Negativity of the Debye–Hückel Pair Correlation Functions

It is a widespread misbelief that the reason why the DH theory yields too low (large negative) values for the electrostatic part of the internal energy is due to the negativity (i.e., unphysical nature) of the pair correlation functions at small interionic separation distance. This was the essence of Pitzer's (1980) criticism of the paper of Pollock and Alders and it is used as an argument in several other papers as well (cf. Alder *et al.*, 1980; Gombert and Deutsch, 1984; Iyetomi and Ichimaru, 1986). This negativity of the $g(r)$ functions originates from the usual linearization of the Poisson–Boltzmann (PB) equation

$$\nabla^2 \psi_i(r) = -\frac{4\pi}{\varepsilon} \sum_j z_j e \exp(-z_j e \psi_i / k_B T) = -\frac{4\pi}{\varepsilon} \sum_j z_j e g_{ij}(r), \tag{2}$$

where

$$g_{ij}(r) = \exp(-z_j e \psi_i / k_B T); \tag{3}$$

in which $\psi_i(r)$ is the electric potential of mean force at a distance from an ith kind of ion; ε, the relative permittivity of the medium; e, the protonic charge; z, the ionic charge number; k_B, the Boltzmann constant; and T, the temperature. The well-known way to linearize this equation *via* the expansion of the exponential function yields $\kappa_D^2 \psi_i(r)$ for the right-hand side of Equation (2) (here κ_D is the inverse Debye length). Similarly, it seems to be self-consistent to write the $g(r)$ functions in the approximate form

$$g_{ij}(r) = 1 + z_j e \psi_i(r) \tag{4}$$

and, indeed, this function is negative when $\operatorname{sign}(z_i) = \operatorname{sign}(z_j)$ and $r \to 0$.

However, one can preserve the non-negativity of the g-functions *and* get a linearization of Equation (2), if the expansion into Taylor-series and the retaining of the leading term is carried out in an unconventional way as follows.

Let us take the definition

$$F_i(\psi_i) = \sum_j z_j e \exp(-z_j e \psi_i / k_B T), \tag{5}$$

so that we have trivially $\psi_i = \mathrm{inv}F_i[F_i(\psi_i)]$, and

$$g_{ij}(r) = \exp\{-z_j e\,\mathrm{inv}F_i[F_i(\psi_i)]/k_B T\}\,; \tag{6}$$

where invF_i is the inverse of function F_i in Equation (5). In the form of Equation (6) the pair correlation function is of course always positive. Substituting Equation (6) into Equation (2) we get

$$\nabla^2\{\mathrm{inv}F_i[F_i(\psi_i)]\} = -\frac{4\pi}{\varepsilon}F_i[\psi_i(r)]\,. \tag{7}$$

Expanding invF_i into a power series of F_i, and retaining only the leading term (with the inverse of the coefficient of the leading term in the expansion of F_i itself!), yields the linearized form in Equation (2) but now with F in place of ψ.

By this 'roundabout' way of linearizing the PB equation, the potential of mean force can be calculated in the simple DH approximation, this approximate value F_{ij} can be used to get ψ_i by the relationship $\psi_i = \mathrm{inv}F_i(F_i)$ which, in turn, is to be inserted into Equation (3) to get the non-negative (i.e., physically meaningful) pair correlation functions. In general, this procedure requires the solution of a nonlinear algebraic equation for ψ_i. For symmetrically charged binary $(z - z)$ plasmas (or electrolyte

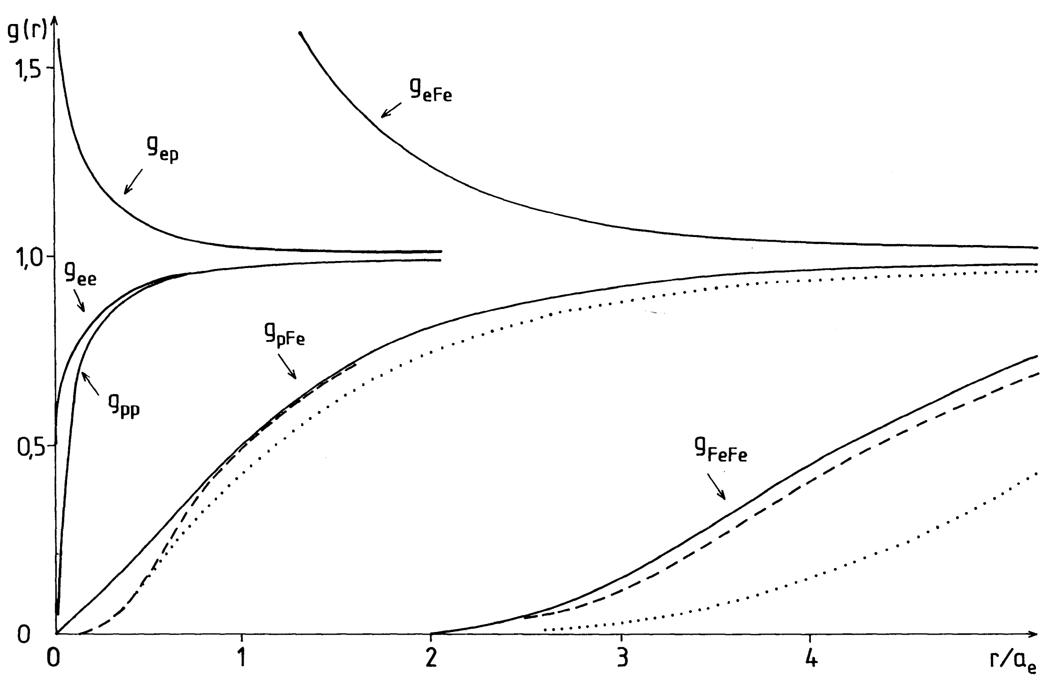


Fig. 1. Partial pair correlation functions of a mixture of hydrogen and iron plasma (molar fraction of iron = 2.25×10^{-5}, T = 14.66 MK, p = 1.042×10^5 Mbar). Full curves: modified Debye–Hückel theory; dotted curves: hypernetted chain theory of binary ionic mixture in an unpolarizable neutralizing background; broken curves: improved HCN–BIM theory with polarizable background.

solutions), however, the F_i is the sinh function that can be inverted in a simple way as we have shown earlier (cf. Ruff, 1982).

With non-negative DH g-functions at hand, we are able to compare them with those obtained by Iyemoto and Ichimura using a sohisticated HNC approximation including the bridge functions of the cluster expansion (Iyemoto and Ichimaru, 1986). In Figure 1 this comparison is made for a low-Γ plasma in which the linearized PB equation yields good values for the potential of mean force. It is seen that the deviation of the DH g-functions from the HNC ones is indeed rather small.

It is worth to note that the exact electroneutrality condition is retained in our approximation which shows in the fact that the like-like and like-unlike charged g-functions are *not* symmetrical with respect to the $g(r) = 1$ line as they were when the approximation had been made by Equation (4). Thus the charge density at a given r, which originates from the difference between these g-functions, as well as the electrostatic internal energy are a rather good approximation of the HNC results (for the case specified in the caption of Figure 1, $\Delta U^{ex}/k_B T$/particle = -0.0156 and -0.0164, respectively). Therefore, one may expect that even if the DH g-functions are less precise, their deviations from the accurate ones are such that the errors partly cancel in their *differences*.

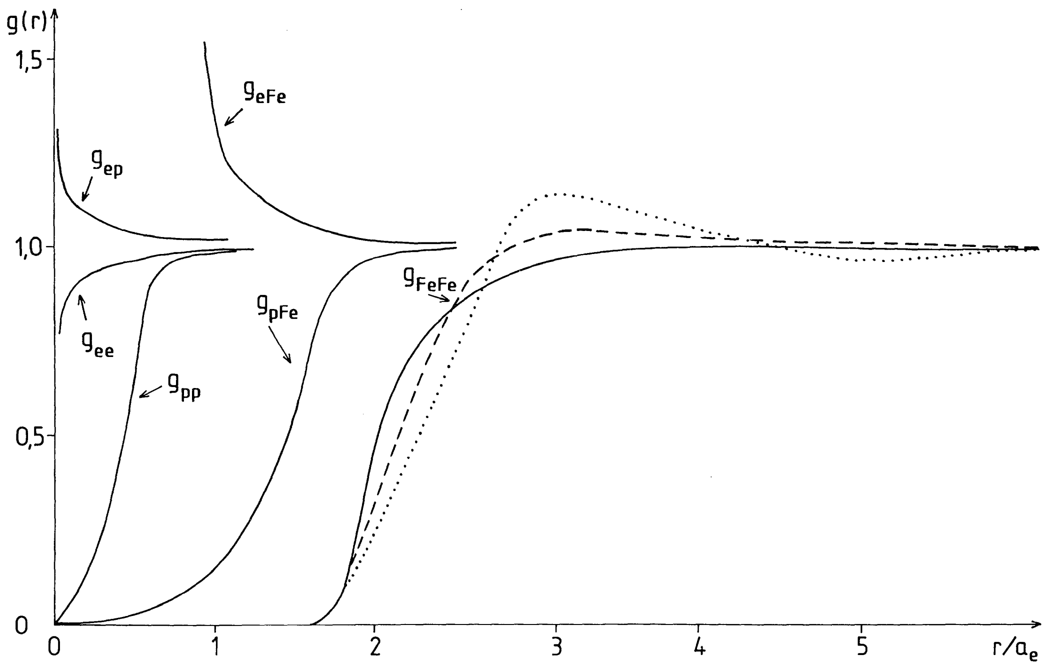


Fig. 2. Partial pair correlation functins of a mixture of hydrogen and iron plasma (molar fraction of iron = 0.999, T = 14.66 MK, p = 1.042 × 10^5 Mbar). Full curves: modified Debye–Hückel theory; dotted curve: hypernetted chain theory of a one-component iron plasma in an unpolarizable neutralizing background; broken curve: modified mean spherical approximation with part of the electrons bearing hard cores with a diameter of the thermal de Broglie radius.

It is more important, however, to make this comparison in the ion-rich phase, since the conditions in it cause the Γ-value of the subsystem of iron ions to become larger than unity which would increase the error of the DH theory.

The g-functions of the iron-rich phase are shown in Figure 2. Unfortunately, the HNC results for the polarizable BIM are not available for this composition, therefore, the comparison is made with the one-component plasma (OCP) model of a pure Fe^{24+} phase (cf. Springer *et al.*, 1973). In this case one has to be aware of the fact that the unpolarizable background of the OCP shields the repulsion between iron ions less efficiently than a polarizable one would do, whence the deviation of the DH curve from the OCP one is overestimated in comparison to a polarizable BIM. Although the g_{Fe} functions show larger deviations than before, the difference in internal energy between the DH and OCP models (Springer *et al.*, 1973) ($\Delta U^{\mathrm{ex}}/k_{\mathrm{B}}T$/particle = -18.2214 and -8.31, respectively) is much larger than what can be attributed to the difference in the potential energy due to an eventual increase in the repulsion between iron ions. The energy difference must rather be due to the non-uniform, polarized, distribution of the electrons in the DH model.

However, to reinforce the earlier statements on the insolubility of iron in internal solar plasmas, it would be necessary to answer some further questions. (1) Is the density of electrons in the close vicinity of iron ions so high that the spin interactions and the correlational motion of the electrons lessen their shielding ability around iron ions, thus leading to more OCP-like behaviour? If so, these effects can be modelled by MSA putting hard repulsive cores on the electrons. This, then, leads to question (2): Is the MSA internal energy as insensitive to hard core parameters on electrons as it proved to be to those on protons and iron ions (Ruff *et al.*, 1985)?

The DH model yields indeed electron densities high enough to expect their non-RPA behaviour but only within a distance (from the iron ion) of $\frac{1}{10}$ of the thermal de Broglie radius. The effect of this region on the internal energy, however, must be quite small due to the $4\pi r^2$ factor of the $g(r)$'s in it. Nevertheless, we have repeated our earlier calculations with two MSA models.

In one, we put hard cores with a diameter of the thermal de Broglie radius (194 pm for the center of the Sun) on all electrons. This is so large a value that, in the corresponding density, the pair correlation functions of iron ions and electrons show in-phase oscillations with increasing distance, yet $\Delta U^{\mathrm{ex}}/k_{\mathrm{B}}T$/particle dropped from -1.8214 only to -1.7879.

In another computation, we tried to model enhanced electron-electron repulsion by considering 'two kinds' of electrons in the MSA: some point-like and some bearing a hard core of thermal de Broglie radius in a proportion to yield the electron-electron pair correlation function shown in Figure 3. This is a step-wise approximation of the pair correlation function of the degenerate Fermi gas more appropriate in the dense region in the vicinity of iron ions than far from them.

Simultaneously we modified the MSA technique to avoid the negativity of the g-functions by imposing the structure functions obtained by the Fourier-transformation of the non-negative $g(r)$ functions of the DH approximation outlined above as a

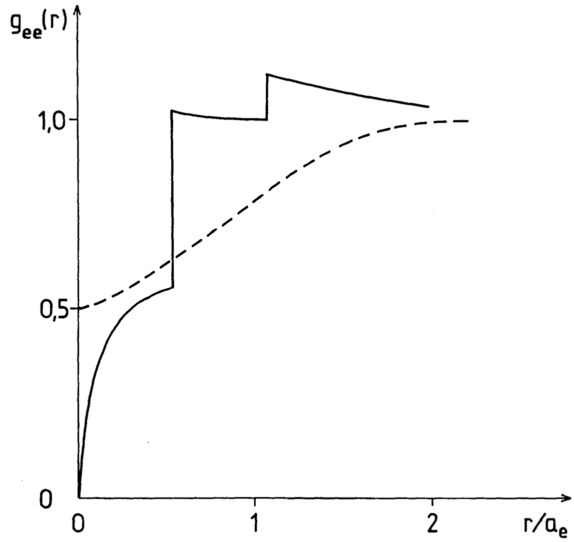


Fig. 3. Partial pair correlation functions of electrons in a pure iron plasma ($T = 14.66$ MK, $p = 1.042 \times 10^5$ Mbar). Full curve: modified MSA theory with part of the electrons bearing hard cores with a diameter of the thermal de Broglie radius; broken curve: electron correlation in a random phase approximation of a degenerate Fermi gas under the same conditions.

reference system in the calculation of the MSA pair correlation functions. This modification, of course, leaves the internal energy of the MSA unaltered.

The iron-iron pair correlation function obtained in this way is shown in Figure 2. It is seen that it runs in between the OCP and DH curves where one would expect to find the $g(r)$ of the polarizable BIM as well. Yet, the corresponding internal energy is hardly different from that of the DH model (-1.8111 instead of -1.8214).

Thus we can conclude, as the main conclusion of this section, that we could not find any substantial increase in the internal energy of the iron-rich phase that would result in the low (large negative) excess free enthalpy of mixing reported to have been obtained by the polarizable BIM model of Iyetomi and Ichimaru (1986). Instead, these calculation support our earlier assumption of the insolubility of heavy elements in the Sun.

3. The Solar Neutrino Signal

Nuclear fusion in stellar hydrogen runs mainly *via* the reaction chain

$$p^+ + p^+ \rightarrow D + e^+ + \nu, \tag{8}$$

$(p^+ + e^- + p^+ \rightarrow D + \nu$ makes less than $1\%)$

$$p^+ + D \rightarrow {}^3\mathrm{He} + \nu, \tag{9}$$

$${}^3\mathrm{He} + {}^3\mathrm{He} \rightarrow {}^4\mathrm{He} + p^+ + p^+ . \tag{10}$$

In case of considerable ^{4}He admixture, the last reaction is competing with the reaction:

$$^3\text{He} + {}^4\text{He} \rightarrow {}^7\text{Be} + \nu . \tag{11}$$

^{7}Be may capture e^- (weak process): i.e.,

$$e^- + {}^7\text{Be} \rightarrow {}^7\text{Li} + \nu , \tag{12}$$

$$p^+ + {}^7\text{Li} \rightarrow {}^8\text{Be} ; \tag{13}$$

or it may capture a p^+ overcoming a considerable Coulomb barrier: i.e.,

$$p^+ + {}^7\text{Be} \rightarrow {}^8\text{B} , \tag{14}$$

$$^8\text{B} \rightarrow {}^8\text{Be} + e^+ + \nu . \tag{15}$$

Both processes terminate in α decay

$$^8\text{Be} \rightarrow {}^4\text{He} + {}^4\text{He} . \tag{16}$$

At any rate, the global outcome is given by

$$4p^+ \rightarrow {}^4\text{He} + 2e^+ + 2\nu . \tag{17}$$

Dividing the observed solar luminosity $L_0 = 3.86 \times 10^{36}$ W by 14 MeV, one gets the neutrino current originating in the Sun. The current density at the Earth turns out to be $7 \times 10^{10}\ \nu\,\text{cm}^{-2}\,\text{s}^{-1}$.

Davies (1964) used chlorine target to catch the solar neutrinos: i.e.,

$$\nu + {}^{37}\text{Cl} \rightarrow {}^{37}\text{A}^* + e^- , \qquad {}^{37}\text{A}^* \rightarrow {}^{37}\text{Cl} + e^+ + \nu . \tag{18}$$

He observed a capture rate 2.1 ± 0.3 SNU (Davies, 1964) (1 solar neutrino unit = 10^{-36} capture atom^{-1} s^{-1}). This result is in disagreement with the standard solar model (cf. Kippenhahn and Thomas, 1965) (on the basis of an ideal gas approximation, the theoretical expectation was 6.5 ± 1.9 SNU (Bahcall, 1978)). In order to explain this discrepancy, special assumptions – like the instability of the neutrino – were suggested, but the observation of the neutrino flux from the Supernova 1987a disproved most of these exotic explanations. The classical idea of the decreased opacity of the solar plasma due to the depletion of heavy elements offers a rather attractive explanation of the solar neutrino puzzle (Ruff and Liszi, 1985; Dearborn *et al.*, 1987).

By making the modest assumption that iron is practically eliminated where $T \geqq 3 \times 10^6$ K (actually with a transition layer between 3×10^6 and 10^6 K), the Livermore code indicates a drop between 15 and 30% in the interior opacity (with a logarithmic decrease in the transition layer).

Comparison of the standard model with two models of reduced opacity is given in Table I and Figure 4. It is seen that a 30% decrease of internal opacity reduces the central temperature by 0.75×10^6 K (about 5%) which suppresses the chlorine neutrino signal from 5.9 SNU to 2.5 SNU in acceptable agreement with the experimental result of 2.1 ± 0.3 SNU.

In order to obtain the observed optical luminosity at a lower central temperature, a

TABLE I

Comparison of the standard solar model with reduced opacity models (Livermore model)

	SSM	ROM1	ROM2
Relative opacity	1	0.85	0.70
$X(R_\odot)$	0.731	0.760	0.791
$X(0)$	0.376	0.410	0.438
$T(0)$/MK	15.45	15.03	14.69
$\rho(0)/10^3$ kg m^{-3}	146.22	144.88	146.55
$p(0)/10^{16}$ N m^{-2}	2.36	2.32	2.39
Neutrino flux/SNU	5.852	3.72	2.538

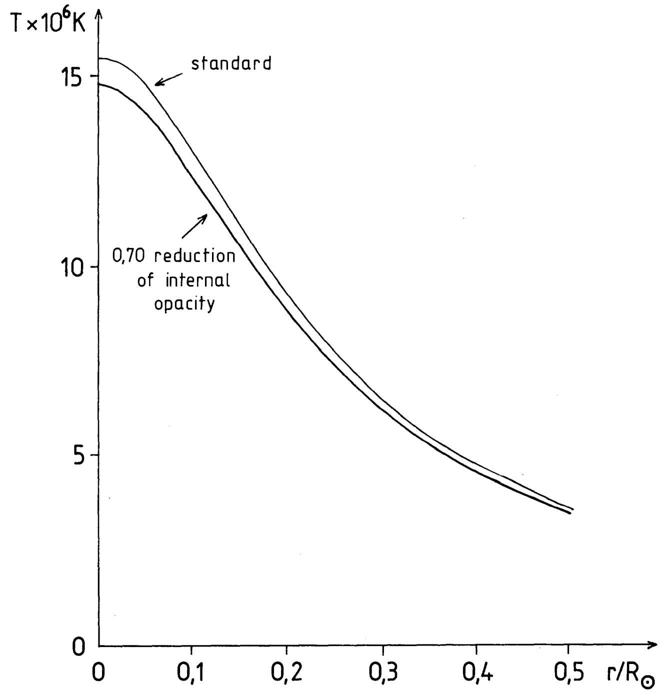


Fig. 4. Comparison of the temperature function of the standard solar model and the reduced opacity model.

higher initial hydrogen concentration must be assumed – 79% instead of 73% as input in the standard Sun model – but this can be tolerated within the accuracy of spectroscopical evidences.

Thus, taking into account also the uncertainties of solar evolution codes, the depletion of the solar interior from iron seems to explain the bulk of the chlorine-captures neutrino discrepancy.

In case of a gallium target, the capture reactions would be

$$\nu + {}^{71}\mathrm{Ga} \rightarrow {}^{71}\mathrm{Ge}^* + e^- , \qquad {}^{71}\mathrm{Ge}^* \rightarrow {}^{71}\mathrm{Ga} + e^+ + \nu , \tag{19}$$

for which the capture threshold is so low ($E_\nu > 0.236$ MeV) that – judging by the standard model – almost 70% of the capture rate is due to $p^+ + p^+$ and $p^+ + e^- + p^+$ neutrinos, an these reactions are slightly enhanced by the increased hydrogen concentration (Figures 5–7).

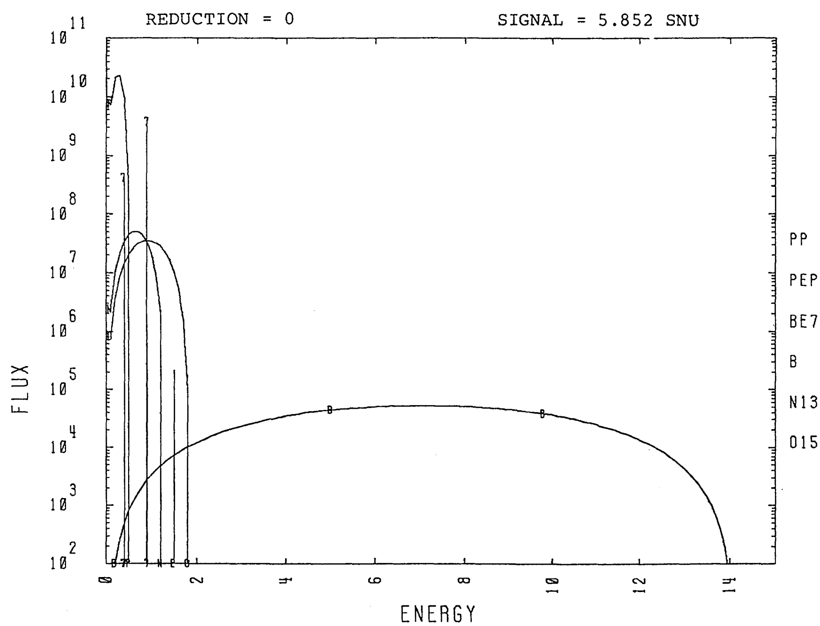


Fig. 5. The neutrino spectrum from the standard solar model.

As a consequence of iron depletion one would predict essentially the same high cupture rate as given by the standard Sun model in the future gallium experiment. This prediction is in contrast to the neutrino oscillation hypothesis in which a wide range of coupling parameters suppresses both chlorine and gallium signals.

4. Are There High-Z Cores in Stars?

Let us consider a two-component, phase-separated, stellar plasma in which the heavy atoms are modelled by iron, their most abundant representative. The hydrogen-rich plasma phase is modelled by pure hydrogen. The concentration of the iron-rich phase can be estimated to be as high as 0.99 expressed in the fraction of electrons compensating their charges (Ruff *et al.*, 1985). The latter is assumed to be 24 as given by the same equation. The mass density and the electron density of these two phases can be

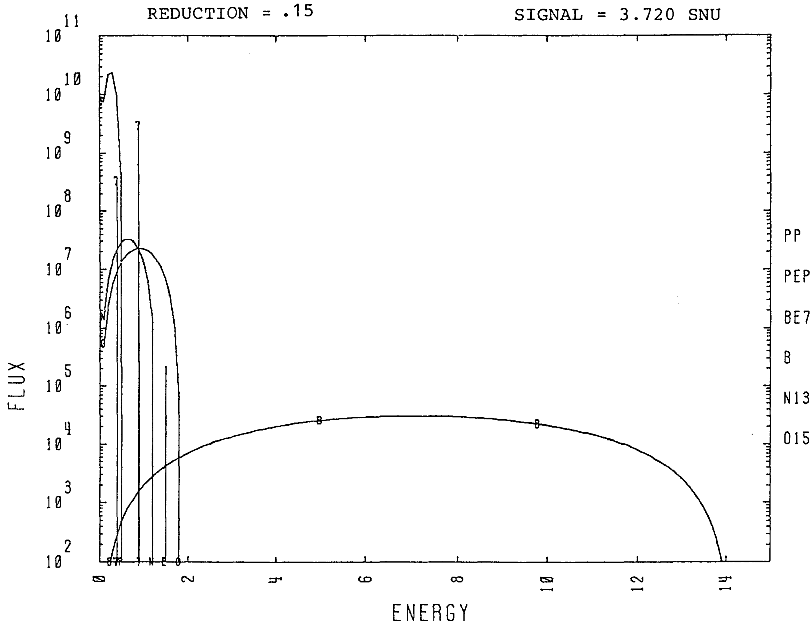


Fig. 6. The neutrino spectrum from the reduced opacity model (15% reduction).

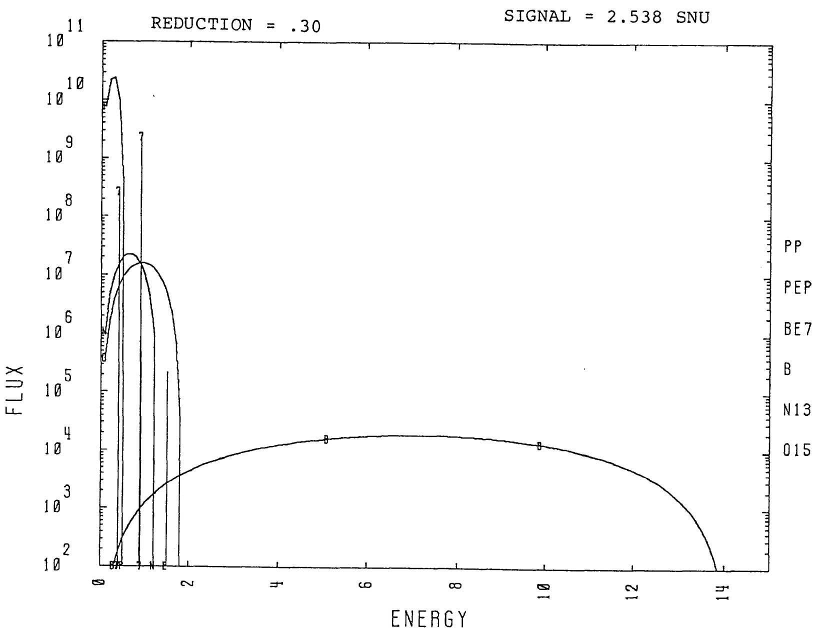


Fig. 7. The neutrino spectrum from the reduced opacity model (30% reduction).

calculated from isothermal-isobaric conditions. The difference in mass density between the iron 'droplets' and the hydrogen plasma surrounding them gives the gravitational force acting on a unit volume.

The radiation pressure exerted on the iron droplets, on the other hand, is approximated by the sum of momenta obtained by the particles in three kinds of interactions with photons: (i) bound-free transition (photoionization) of the K-shell electrons of iron, (ii) free-free transition (*bremsstrahlung*) of electrons in the field of iron ions, and (iii) Compton scattering (cf. Clayton, 1983). A hydrogen plasma contains as many electrons as protons. The iron plasma contains 24 times as many free electrons plus two times as many bonded ones as iron ions. Whereas the density of both plasmas is determined through the (nearly) ideal gas equation of state by the total number of particles irrespective of their charges. Thus the main difference between the two phases with respect to *bremsstrahlung* and Compton scattering is due to the difference in the density of electrons $\Delta\rho_e$. The volume force exerted by the radiation pressure is

$$F_{V,r} = (\Delta\rho_e)^{2/3} \int_0^\infty [\sigma_C(\nu) + \sigma_f(\nu) + \sigma_b(\nu)] \left(\frac{\partial E}{\partial \nu}\right)_T \left(\frac{\partial T}{\partial r}\right) d\nu , \tag{20}$$

where σ_C, σ_f, and σ_b are effective cross sections of the Compton scattering and the free-free and bound-free transitions, respectively. All cross sections depend on the frequency of the photons as given in the equations

$$\sigma_C = 2\pi r_0^2 \left\{\frac{1+\varepsilon}{\varepsilon^2}\left[\frac{2(1+\varepsilon)}{1+2\varepsilon} - \frac{1}{\varepsilon}\ln(1+2\varepsilon)\right] + \frac{1}{2\varepsilon}\ln(1+2\varepsilon) - \frac{1+3\varepsilon}{(1+2\varepsilon)^2}\right\}, \tag{21}$$

where $\varepsilon = h\nu/m_e c^2$ and $r_0 = e^2/4\pi\varepsilon_0 m_e c^2$ if m_e denotes the electron mass, c the speed of light, h the Planck constant, and ε_0 the vacuum permittivity.

$$\sigma_f = 3.69 \times 10^{-8}(Z^2 g_{Fe}\rho_{e,\,Fe} - g_H \rho_{e,\,H})/T^{12}\,\nu^3 , \tag{22}$$

$$\sigma_b = 2.82 \times 10^{29}\, Z^4 g(\rho, n, l, Z)/n^5\,\nu^3 , \tag{23}$$

where Z is the charge of iron ions, n the principle, and l the angular quantum number, while the g's are the corresponding Gaunt factors (Clayton, 1983).

Instead of taking into account the bound-bound transitions, we have simplified the problem by lowering the threshold of the bound-free transition to the energy difference between the $n = 2$ and $n = 1$ electrons in the 'helium-like' iron ion. By this the eventual forbiddenness of some transitions is disregarded which results in some overestimation in the opacity of the pure iron plasma. Another approximation contributing to this overestimation was the assumption that the 'droplets' are small enough not to decrease the intensity of the radiation while light passes through them. As the results below will show, they do not influence our final conclusions.

For the energy density distribution of photons $E(\nu)$ in Equation (20) we can use the Planck formula for black-body radiation which is considered to be a good approximation in the interior of stars: i.e.,

$$E(\nu) = (8\pi h\nu^3/c^3)/[\exp(h\nu/k_B T) - 1] . \tag{24}$$

Taking temperature, pressure, and density data from the standard model of the Sun (Kippenhahn and Thomas, 1965; Bahcall, 1978), the volume forces attracting the iron plasma phase towards the center $F_{V,g}$ and that pushing it towards the surface $F_{V,r}$ could be obtained as listed in Table II. It is seen that the force of gravity is always greater by orders of magnitude than the force of the radiation pressure even if the bound-bound transition is slightly overestimated in it. Thus it can be concluded that if phase separation occurs in the Sun, the heavy elements must accumulate in its central region.

TABLE II

Comparison of gravitation and radiation force exerted on iron droplets in the Sun

$R/R_\odot$	T/MK	p/dyn cm^{-2}	$\Delta\rho_m/g$ cm^{-3}	$(\Delta\rho_e)^{2/3}$ cm^{-2}	$F_{V,g}$/dyn cm^{-3}	$F_{V,r}$/dyn cm^{-3}
0.02	14.52	2.16×10^{17}	304.24	7.64×10^{16}	1.45×10^{7}	22.6
0.06	13.83	1.84×10^{17}	271.78	7.08×10^{16}	3.78×10^{7}	54.1
0.10	12.65	1.36×10^{17}	220.40	6.16×10^{16}	4.44×10^{7}	62.2
0.16	10.64	7.48×10^{16}	143.62	4.63×10^{16}	3.44×10^{7}	34.5
0.20	9.35	4.64×10^{16}	101.44	3.67×10^{16}	2.42×10^{7}	19.2
0.30	6.65	1.18×10^{16}	36.24	1.85×10^{16}	7.21×10^{6}	5.1
0.40	4.74	2.70×10^{15}	11.64	8.68×10^{15}	1.71×10^{5}	0.7
0.50	3.43	6.13×10^{14}	3.66	4.01×10^{15}	3.86×10^{5}	0.1

Nevertheless, there are still several problems to be solved. It has been demonstrated (Ruff and Liszi, 1985; Ruff *et al.*, 1985) that iron becomes soluble if its charge is decreased to about 16. This may be a result of the drop in temperature with increasing distance from the center of the star. In the solar model we have used in these calculations, this value is reached at $R/R_\odot \cong 0.9$, i.e., near the borderline of the convective region. In light of the foregoing discussion the question seems to be important whether the depth at which iron becomes soluble is larger than the depth of the convective layer or not. If it is, the convective layer is in contact with a region in which iron is present in its cosmic abundance and, thus, the two layers would be in equilibrium in this aspect. However, if it is not, the deeper, depleted layer would gradually extract iron (and, in general, the heavy elements) from the convective layer. This would make the heavy element content of the outer convective layer of stars and, consequently, the intensity of the corresponding spectral lines change with time. Since cosmic abundance tables are based mainly on spectrographic results obtained from the outer layer of the Sun, this possibility should be regarded as an additional caution against overestimating the precision of abundance values.

The phase separation, the formation of an iron core, the acoustic and gravitational vibrations, and the neutrino flux comprise a self-consistent system of proposed modifications to be carried out to resolve the present controversies between observations and the standard solar model.

References

Alder, B. J., Pollock, E. L., and Hansen, J. P.: 1980, *Proc. Nat. Acad. Sci. USA* **77**, 6272.

Aller, L. H. and Chapman, S.: 1960, *Astrohys. J.* **132**, 461.

Bahcall, J. N.: 1978, *Rev. Mod. Phys.* **50**, 881.

Clayton, D. D.: 1983, *Principles of Stellar Evolution and Nucleosynthesis*, University of Chicago Press, Chicago, pp. 185–232.

David, G., Marx, G., and Ruff, I.: 1984, *Proc. 11th Int. Conf. Neutrino Phys. Astrophys., Dortmund,* World Scientific Publ., Singapore, p. 254.

Davis, R., Jr.: 1964, *Phys. Rev. Letters* **12**, 303.

Dearborn, D. S., Marx, G., and Ruff, I.: 1987, *Prog. Theor. Phys.* **77**, 2.

Gombert, M. M. and Deutsch, C.: 1984, *Phys. Scripta* **T7**, 113.

Iyetomi, H. and Ichimaru, S.: 1986, *Phys. Rev.* **A34**, 3203.

Kippenhahn, R. and Thomas, H. C.: 1965, in H. H. Voigt (ed.), Landoldt-Börnstein, Group 6, Vol. 1, *Astronomy and Astrohysics*, Springer-Verlag, Berlin, p. 485.

Pitzer, K. S.: 1980, *Proc. US Nat. Acad. Sci.* **77**, 3103.

Pollock, E. L. and Alder, B. J.: 1978, *Nature* **275**, 41.

Rouse, C. A.: 1983, *Astron. Astrophys.* **126**, 102.

Rouse, C. A.: 1985, *Astron. Astrophys.* **149**, 65.

Rouse, C. A.: 1986, *Solar Phys.* **106**, 205, 217.

Ruff, I.: 1982, *Chem. Phys. Letters* **93**, 553.

Ruff, I. and Liszi, J.: 1985, *Chem. Phys. Letters* **116**, 335.

Ruff, I., Liszi, J., and Gömbös, K.: 1985, *Astrophys. J.* **289**, 409.

Springer, J. F., Pokrant, M. A., and Stevens, F. A., Jr.: 1973, *J. Chem. Phys.* **58**, 4863.

ELECTRON HEATING IN SUPERHIGH MACH NUMBER SHOCKS*

K. PAPADOPOULOS

Department of Physics and Astronomy, University of Maryland, College Park, MD, U.S.A.

(Received 15 July, 1987)

Abstract. Fluid and MHD models, as well as direct extrapolation of the Earth's bow shock measurements in the high Mach number (HMN) range ($3 \leq M_F \leq 12$) to the superhigh Mach number (SHMN) range ($M_F > 30$–40) predict that the downstream electron pressure p_{e2} is only a negligible fraction of the Rankine–Hugoniot downstream pressure p_2, i.e., $p_{e2}/p_2 \approx (M_F^2)^{-1}$. However, the interpretation of X-ray supernovae emissions, due to SHMN shock heating requires $p_{e2}/p_2 \approx 0(1)$. Following Alfvén we have used plasma physics experimental-theoretical data combined with magnetospheric observations to probe the physics of the SHMN shocks. It is shown below that inclusion of proper plasma physics considerations in the interaction of the reflected and transmitted ions and the electrons at the 'foot' of the shock leads to the surprising result that electron heating can dominate in the SHNM range. A stationary model of the shock structure is derived and shown to be the result of extrapolation of the high Mach number shock physics with incorporation of collective interactions at the foot.

1. Introduction

A major achievement of space plasma physics during the last 6 years has been the emergence of a clear understanding of supercritical quasi-perpendicular shocks. Supercritical shocks have a magnetosonic Mach number $M_F \gtrsim 2$–3, and result in the reflection upstream of the shock of a fraction of the incoming ions (Tidman and Krall, 1971; Kennel *et al.*, 1985; Papadopoulos, 1985; Sagdeev, 1966, 1979). Quasi-perpendicular refers to angles θ between the plasma flow velocity $\mathbf{U}_1$ and the ambient magnetic field $\mathbf{B}_1$, $\theta \equiv \cos^{-1}(\mathbf{U}_1 \cdot \mathbf{B}_1 / U_1 \mathbf{B}_1) > 40°$. The understanding was the result of utilizing the highly resolved measurements by the ISEE spacecrafts (Scudder *et al.*, 1986; Greenstadt *et al.*, 1980; Paschmann *et al.*, 1982) with new highly sophisticated hybrid codes (Leroy *et al.*, 1981; 1982; Goodrich, 1985; Forslund *et al.*, 1984; Quest, 1985). A ubiquitous feature of the supercritical shocks is the presence of a magnetic field overshoot. It was found that the size of the overshoot and the ion reflection process were self-regulated so as to produce downstream the conditions necessary for a pressure balanced stationary shock. An example of the simulation results is shown in Figure 1 which refers to the simulation of the November 12, 1977 bow shock crossing (Leroy *et al.*, 1981, 1982; Papadopoulos, 1985). Figure 1(a) shows the overall magnetic field profile from the simulations which was in excellent agreement with the observations. It shows a foot, a ramp, and an overshoot-undershoot transition before reaching downstream. Figures 1(b) and 1(c) show the $(v_x - x)$, $(v_y - x)$ phase space at the same time. It can be clearly seen that the incoming ion distribution is split into a transmitted and

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 535–547.

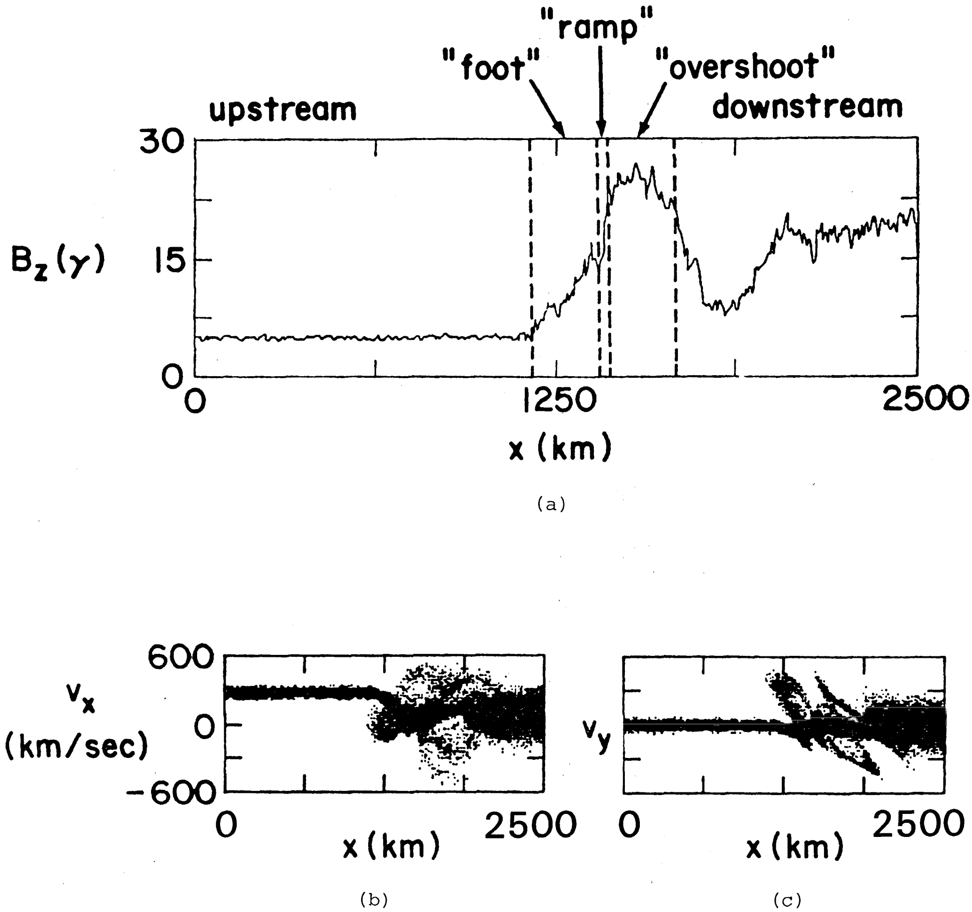


Fig. 1. HMN shock hybrid simulation results: (a) magnetic field profile at one instant of time; (b) and (c) $(v_x - x)$ and $(v_y - x)$ ion-phase space, respectively, at the same time.

a reflected population. The transmitted ions slow down and are adiabatically heated while traveling downstream (Figure 1(b)). The reflected ions perform a gyration upstream and are subsequently accelerated in the y-direction (surfatron acceleration) in front of the shock (Figure 1(c)). This process allows them to overcome the potential barrier and move downstream. Thus the overshoot acts like the barrier that separates the upstream and downstream states. The free energy created by the reflection process (and seen as a 'hole' in phase space downstream) is self-regulated by the size of the overshoot and the fraction of the reflected particles, so as to provide the ion pressure required by the Rankine–Hugoniot (R–H) conditions when thermalized downstream. Details of the above can be found in Leroy *et al.* (1982), Wu *et al.* (1984), and Papadopoulos (1985). An important consequence of the above picture is that for supercritical shocks most of the heating resides in the ions while very little electron heating is expected; this is analogous to viscous fluid shocks. It is in fact easy to show

(Papadopoulos, 1985; Leroy, 1983) that

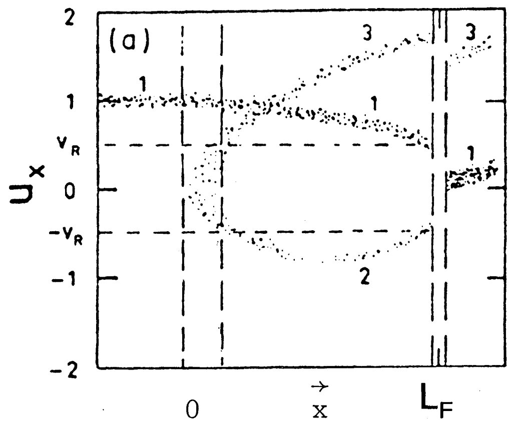
$$\frac{T_{e2}}{T_{e2} + T_{i2}} \sim \frac{1}{M_F^2}\,, \tag{1}$$

where T_{e2}, T_{i2} are the downstream electron and ion temperatures. This is consistent with the observations of the Earth's bow shock where most of the measurements indicate little more than adiabatic electron heating for the range of M_F occurring (i.e., $M_F \approx 2$–12).

This presents a major dilemma in attempting to extraplate our knowledge of high Mach number shocks (HMN) from the solar environment to astrophysics (McKee and Hollenback, 1980; Blandford and Eichler, 1987). Astrophysical shocks are expected to have much higher M_F. The best observational evidence of shocks with $M_F \gtrsim 10^2$–10^3 comes from studying the young supernovae remnants (SNRs) Cas A and Tycho. X-ray observations of Cas A with the Einstein X-ray Observatory show a shell of X-ray emission coincident with a plateau of radio emission and lying outside a complex bright X-region, which is interpreted as emission from a shocked interstellar medium with $T_e \approx 4$–5 keV. Observations with HEAO-1 indicate emission from Cas A with $T_e \simeq 10$ keV and an 8 keV component from Tycho. Even higher values of T_e might be required to account for the emission of the recently discovered supernova SN 1987A (H. Völk, private communication). Since the T_i to T_e collisional equipartition time is much longer than the SNR lifetime, one has to examine the possibility that superhigh Mach numbers (SHMNs), i.e., shocks with M_F much larger than the range of 10–20 encountered in the solar terrestrial environment, are controlled by different processes which can result in significant collisionless electron heating. It is the purpose of this contribution to indicate such a process and derive qualitatively the shock structure expected at SHMN shocks.

2. Problems in Extrapolating of the HMN Shock Physics to SHMN Shocks

A theoretical multifluid model of shocks presented recently by Leroy (1983) offers the best vehicle for extrapolating the shock physics to the SHMN range. The model is shown schematically in Figures 2(a) and 2(b). The ions are modeled as three waterbag fluids in the foot region, as two fluids in the ramp and downstream, while a single fluid model is adopted for the electrons. The model can account quantitatively for the self-regulated ion reflection and for many quantitative features of the simulations and observations, such as the values and scalings of the overshoot, the fraction of reflected ions vs M_F, etc. An important ingredient of the HMN shock physics is the overshoot and the associated structure of the electric potential; they control the fraction of reflected ions so that when they are transmitted downstream and thermalized they produce the necessary ion temperature to satisfy the R–H conditions. Relative to the downstream state of Figure 2 the free energy available for ion heating is the relative energy between beams 1 and 3. Scaling results from the three fluid models (Leroy, 1983) and the hybrid



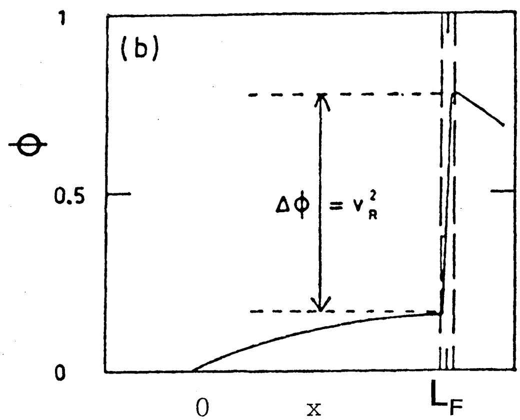


Fig. 2. Schematic representation of HMN shock simulation results according to Leroy's (1983) model. (a) $x - v_x$ ion-phase space, (b) normalized potential $2e\phi/\frac{1}{2}MU_1^2$ vs x. V_R is defined as $V_R^2 = \Delta\phi$. The point $x = 0$ represents the edge of the foot and L_F the shock ramp.

simulations (Leroy *et al.*, 1981, 1982; Goodrich, 1985; Papadopoulos, 1985; Quest, 1986) are shown in Figure 3 as a function of the Alfvénic Mach number M_A for an upstream value of the ion beta $\beta_{i1} \simeq 1$. The values of ϕ_{max} and $\Delta\phi$ which represents the ramp potential (Figure 3(a)) are normalized to $\phi \to 2e\phi/MU_1^2$. ϕ_{max} decreases from 0.8 at $M_A \approx 6$ to 0.5 at $M_A = 16$. Note that ϕ_{max} remains at all times below the value of $\phi \approx 0.9$ required to decelerate cold ions from U_1 to the downstream value of $U_2 \approx \frac{1}{4}U_1$ required by the R–H conditions. A larger fraction α of ions is reflected with increasing M_A (Figure 3(b)) despite the decrease in ϕ_{max}; the fraction of reflected ions $\alpha \approx 0.08$ at $M_A \approx$ 4–6 and increases to $\alpha = 0.21$ at $M_A \approx 16$. This is due to the fact that while ϕ_{max} decreases with M_A, the magnetic deflection in the v_y shown in Figure 1(c) increases by a factor of 3 from $M_A \approx 4$ to $M_A \approx 16$. The process is self-regulated so that the value of α and the value of ϕ_{max} combine to produce the pressure required by the R–H

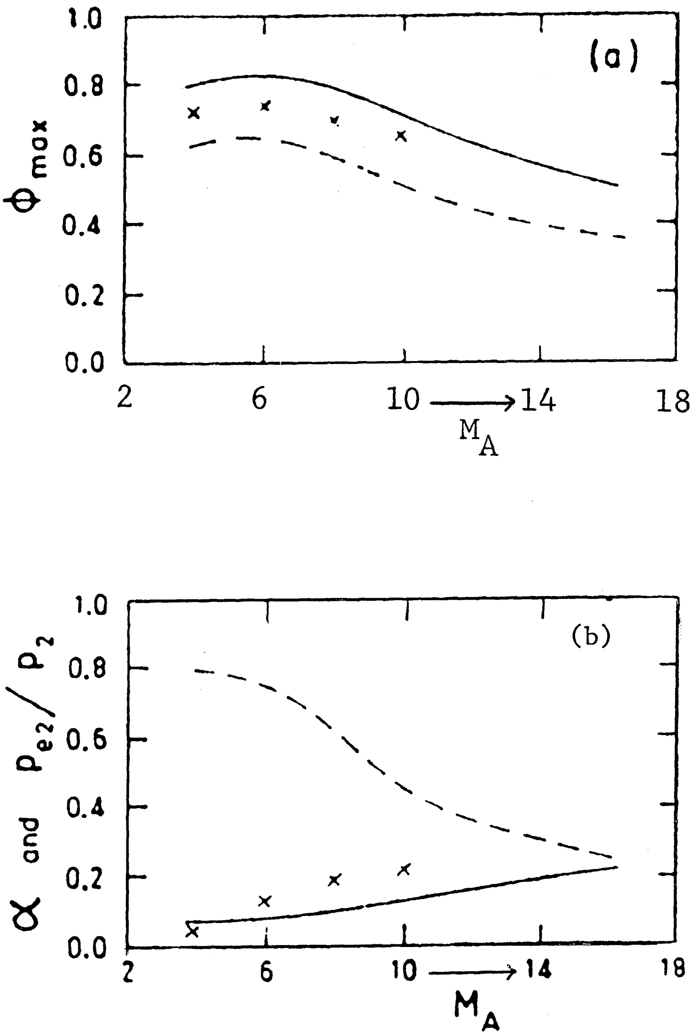


Fig. 3. Mach number scaling of the shock structure (Leroy, 1983). (a) ϕ_{max} (solid line) and $\Delta\phi$ (dotted line), (b) α (solid line), P_{e2}/P_2 (dotted line). Crosses represent simulation results for (a) ϕ_{max} and (b) α.

conditions when beams 1 and 3 thermalize downstream. Figure 3(b) also shows the ratio of the electron pressure p_{e2} downstream to the total pressure p_2 required by the R–H conditions. As noted in the previous section it decreases strongly with M_A. A final comment on the above model is that no HMN solutions were found by Leroy (1983) with $\alpha > 0.35$, while the simulations indicated an upper limit of $\alpha \approx 0.25$ which occurs in the range of $M_A \approx 12$–15.

The physical description given above is consistent with the hybrid simulations at the University of Maryland (Leroy *et al.*, 1981, 1982), the simulations at Los Alamos National Laboratory (LANL) (Forslund *et al.*, 1984; Quest, 1985) and the experimental

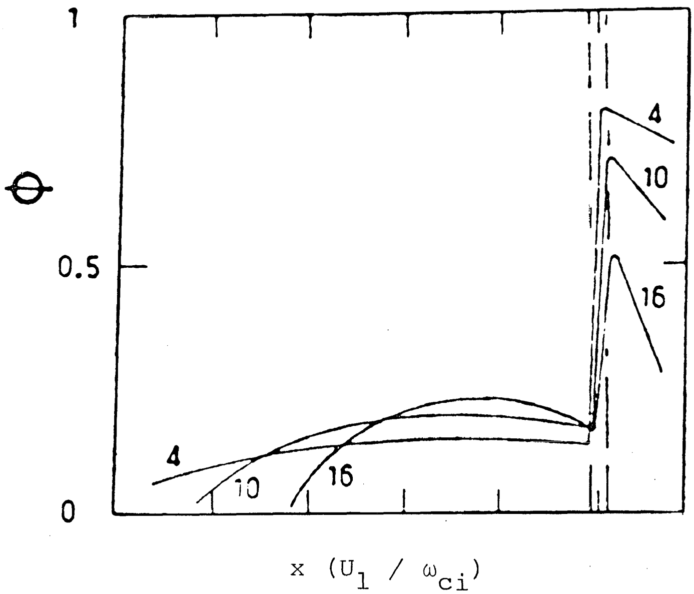


Fig. 4. Spatial profile of potential structure for various M_A numbers (Leroy, 1983).

measurements at the Earth's bow shock (Greenstadt *et al.*, 1980; Paschmann and Sckopke, 1983; Sckopke *et al.*, 1983) up to $M_F \approx 12$–13 or $M_A \approx 14$–16. In attempting to apply this model to higher Mach numbers a number of theoretical difficulties appear. Furthermore, the simulation results for high Mach numbers are sensitively dependent on the value of resistivity used and the resolution of the computational grid (Quest, 1985; Papadopoulos, 1985; Goodrich, 1985) and, therefore, physically unreliable. Finally the observational basis is rather poor (Moses *et al.*, 1984, 1985a, b; Russell *et al.*, 1982).

An early warning of the difficulties of aplying the above physical picture to $M_F \gtrsim 12$–13 (or $M_A \geq 14$–16) was given by Kennel *et al.* (1985). They, correctly, noted that since most of the downstream energy resides in the gyrating ions (i.e., Figure 2 beam 3), in order to satisfy the R–H pressure conditions downstream for $M_F \geq 12$–13 the value of α should be $\alpha \gtrsim 0.35$. However, the simulations (Leroy *et al.*, 1981, 1982), laboratory experiments (Chodura, 1975), and observations (Paschmann *et al.*, 1982) indicate saturation of the value of α at about $\alpha \approx 0.2$–0.25 (Figure 3(b)). Another difficulty noted by Papadopoulos (1985) and Goodrich (1985) is the fact that a finite ion temperature is required at the shock front, for a fraction of the incoming ions to be reflected. Otherwise all the incoming ions will either be reflected or transmitted. A measure of the finite ion temperature effect is the ratio (δ) of the ion thermal energy (nT_{i1}) to the potential $e\phi_{\max}$ which in normalized units can be written as

$$\delta \equiv \frac{(\beta_{i1})}{M_A^2} \frac{1}{\phi_{\max}} . \tag{2}$$

In the SHMN limit since $\phi_{\max}$ depends weakly on M_{A} (Figure 3(a)) and for constant β_{i1}, $\delta \rightarrow 0$. The reflection process cannot occur in a steady state fashion, since $\delta \rightarrow 0$ implies a cold incoming fluid and the potential can either reflect all of the ions or none. Such an unsteady behaviour was actually seen in the simulations (Leroy *et al.*, 1982) in the $M_{\mathrm{A}} \approx 4$–10 range for low β_{i1} values ($\beta_{i1} \approx 0.1$). The shock front had a limit cycle behaviour over an ion gyrotime (Ω_i^{-1}) with either complete transmission or reflection occurring in fractions of the cycle giving a time-averaged value of α consistent with Figure 3(b). The value of δ with $\beta_{i1} \approx 0.1$ and $M_{\mathrm{A}} \approx 10$ was of the order of 10^{-3}, which is comparable to the value of δ for a $\beta_{i1} = 1$, $M_{\mathrm{A}} \approx 30$ situation. This self-similar behaviour was observed in the simultions of shocks with M_{A} up to 22 by Quest (1985) and up to 25 at the University of Maryland (unpublished data). An average stationary state was established because the ratio of reflection to transmission time was regulated to produce the required balance. There are several reasons to question the reality of the limit cycle picture for SHMN shocks (Quest, 1985). First, contrary to the results for $M_{\mathrm{A}} \lesssim 14$–$16$, the results of the SHMN simulations were extremely sensitive to the precise value of the resistive length to the cell size ratio. Stationary states were established when the resistive length was comparable to the cell size. This implies that the presence of strong instabilities at the foot of the shock will invalidate the dynamic shock picture. Secondly, in the absence of electron dissipation and dispersion the shock-ramp thickness at the time of reflection is one cell; namely, it is controlled by numerics and not by physics. It should also be noted that the nonstationary picture of the shock is at odds with the, admittedly, limited data on $M_{\mathrm{A}} \gtrsim 20$ shocks (Russell *et al.*, 1982; Moses *et al.*, 1984, 1985) and will preserve the $T_e \sim (M_{\mathrm{A}}^2)^{-1}$ scaling. The above results, however, are very instructive guides in developing a picture of SHMN shocks which preserves the essential features of the stationary reflection of the three fluid model, and at the same time results in the substantial electron heating required for astrophysical SHMN shocks. This is presented in the next section.

3. Strong Electron Heating at the Foot of SHMN Shocks

In the HMN shock model discussed previously the ion heating is provided by thermalization of the free energy of the reflected ions. A closer examination of the computational as well as the theoretical results reveals another possible free energy source associated with the potential drop at the foot of the shock given by $\Delta\phi_F = \phi_{\max} - \Delta\phi$ (Figures 1(b) and 3(a)). The scaling of $\Delta\phi_F$ with M_{A} is shown in Figure 4 (Leroy, 1983). From Figure 4 it can be seen that both the ratio $\Delta\phi_F/\Delta\phi$ and the associated electric field at the foot increase with M_{A}. This potential induces a relative streaming velocity U_D between the electron fluid and the main ion beam (beam 1). Wu *et al.* (1983, 1984) made an in depth investigation of electron-ion instabilities in the shock foot region. It was found that in the HMN range the electron-ion instabilities were very weak, producing wave signatures and some electron acceleration but not significant heating. The ion acoustic (IA) instabiliity between the electrons and either the transmitted or the reflected ions was very weak due to the fact that $T_e/T_i \approx 1$ and the ratio of the reflected ion beam velocity

($\sim 2U_1$) to the electron thermal speed V_e was in the range $2U_1/V_e \ll 1$, where only very oblique ion-acoustic modes are weakly unstable. We demonstrate below that this is valid only in the HMN range and a direct extrapolation of the HMN scaling to SHMN indicates that a strong IA instability can be excited between the electrons and the transmitted ions which can produce large electron and moderate ion heating at the foot. It will be shown that as a consequence the basic physics of the stationary reflection dominated three fluid waterbag is preserved at the SHMN range with the addition of strong electron heating at the foot. As a result the downstream state can have $T_e/T_i \geq 1$.

Reviewing the IA instability (Ichimaru, 1973) we find that instability between the electrons and the transmitted ions requires a relative drift velocity U_D such that

$$\frac{U_D}{V_e} \geq \sqrt{1-\alpha}\,\sqrt{\frac{m}{M}} + (1-\alpha)^{3/2}\left(\frac{T_e}{T_i}\right)^{3/2} \exp\left[-(1-\alpha)\,\frac{T_e}{2T_i}\right]; \tag{3a}$$

or, taking $\alpha \approx 0.2$,

$$\frac{U_D}{V_e} \geq 0.9\,\sqrt{\frac{m}{M}} + 0.7\left(\frac{T_e}{T_i}\right)^{3/2} \exp\left[-\frac{0.4T_e}{T_i}\right], \tag{3b}$$

where m, (M) are the electron (ion) masses.

For values of $T_e/T_i > 10$, Equation (3b) becomes the often quoted 'instability condition' $U_D > c_s$ for strongly non-isothermal plasmas. For isothermal plasmas (i.e., $T_e/T_i \approx 1$) the condition for instability becomes $U_D > V_e$ which is similar to the Buneman instability (BI) condition (Papadopoulos, 1977). Therefore, triggering of the IA instability at the foot requires:

(i) A mechanism to heat electrons on a short-length scale at the edge of the foot to $T_e/T_i \gtrsim 10$.

(ii) That $U_D > c_s$ at the foot.

We focus next in the interaction between the reflected ion beam and the electrons. Following Davidson *et al.* (1970) we note that if

$$\frac{U_D}{V_e} \approx \frac{2U_1}{V_e} \geq 1\,, \tag{4}$$

a BI will be triggered and it will result in saturation by heating the electrons to

$$\Delta T_e \approx 2mU_1^2 \approx 4\left(\frac{m}{M}\right)\frac{1}{2}\,MU_1^2\,, \tag{5}$$

consistent with stability for $U_D \lesssim V_e$. The inequality (4) can be expressed as

$$M_F \geq \frac{1}{2}\,\frac{\beta_{e1}^{1/2}}{(1+\beta_1)^{1/2}}\left(\frac{M}{m}\right)^{1/2}. \tag{6}$$

Note that for $\beta_{e1} \ll 1$ the BI can be triggered even for values of M_F below 20. For values of β_1 and β_{e1} of order unity the triggering of the BI requires $M_F \geq 20$. From Equation (5) we see that the BI by itself transfers to electrons less than 1% of the shock energy $\frac{1}{2}MU_1^2$. It can, however, act as a catalyst to trigger an IA instability by increasing the T_e/T_i ratio. From Equation (5) dividing by T_i, assumed constant, we find that for an initial $T_e/T_i \approx 1$ the ratio of T_e/T_i at saturation will be given by

$$\frac{T_e}{T_i} = 1 + 4M_F^2\left(\frac{m}{M}\right)\frac{(1+\beta_1)}{\beta_{i1}} . \tag{7}$$

For $\beta_{e1} \approx \beta_{i1} \approx \frac{1}{2}$ and protons Equation (7) becomes

$$\frac{T_e}{T_i} \approx 1 + 6.5 \times 10^{-3} M_F^2 . \tag{8}$$

It is obvious from Equation (8) that strongly anisothermal plasmas can be produced for values of $M_F > 20$–25. For $M_F \approx 36$ the ratio T_e/T_i becomes large enough ($T_e/T_i > 10$) that the threshold condition is dominated by the first term of Equation (3b) – i.e.,

$$\frac{U_D}{V_e} > 0.9\left(\frac{m}{M}\right)^{1/2} ; \tag{9}$$

namely for $M_F \geq 36$ the IA instability will be excited if at the foot the relative drift U_D between the electrons and the transmitted ions is larger than the local sound speed. The values of U_D should be determined self-consistently on the basis of the potential drop at the foot. This will be pursued numerically in a more detailed paper. We simple mention here that condition (9) is essentially trivially satisfied for the SHMN case since a potential drop at the foot of even 1% of $\frac{1}{2}MU_1^2$ can produce it. This fact indicates that a marginal stability analysis (Lampe *et al.*, 1975), in which the potential adjusts itself to produce locally a value of $U_D/V_e \approx (m/M)^{1/2}$ as required by Equation (3b) for $T_e/T_i > 10$, might be a good avenue to explore semi-quantitatively the structure of a SHMN shock. We will do that in the next section as an illustration.

4. A Possible Physical Picture of SHMN Shocks – A Qualitative Example

We discuss below the physical picture of a SHMN shock that emerges from the previous discussion in the form of an example. The precise numbers and values of the various parameters should be taken only as indicative of the physical processes involved and their relative strength. For concreteness sake and without any loss of generality we consider the case of a SHMN shock with $\beta_{e1} \approx \beta_{i1} \approx \frac{1}{2}$, and value of $M_F \geq 36$ so that from Equation (8) the value of $T_e/T_i \gtrsim 10$ at the edge of the foot. For a strong shock the Hugoniot conditions for the downstream state give (Tidman and Krall, 1971)

$$\begin{aligned} &\frac{U_2}{U_1} \approx \frac{n_1}{n_2} \approx \frac{1}{4} , \\ &\frac{T_{e2} + T_{i2}}{\frac{1}{2}MU_1^2} \approx \frac{3}{16} . \end{aligned} \tag{10}$$

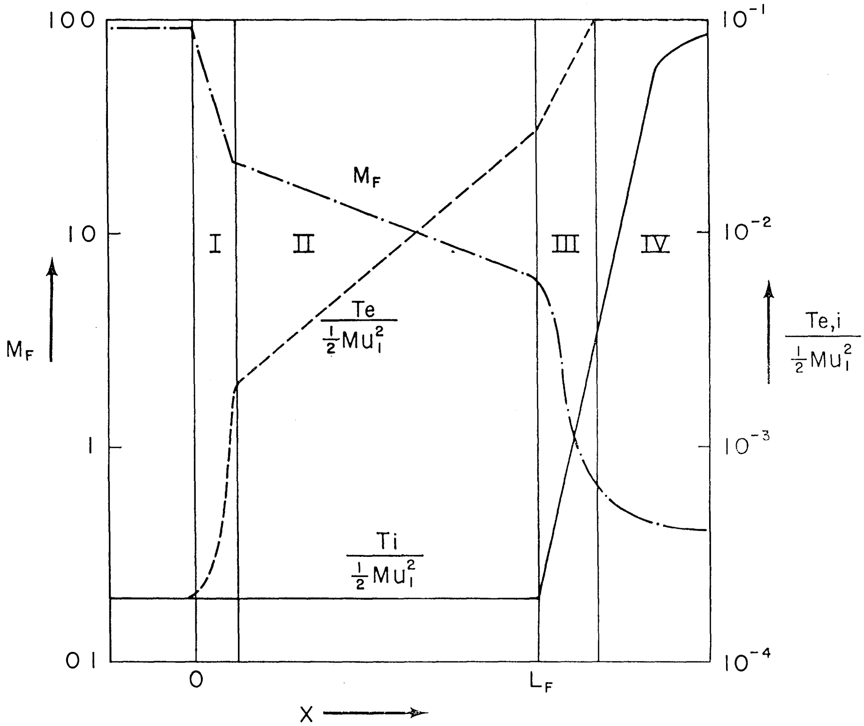


Fig. 5. Schematic of the expected spatial profile of a SHMN shock with strong electron heating. Plotted are M_F, $T_e/\frac{1}{2}MU_1^2$ and $T_i/\frac{1}{2}MU_1^2$ vs x. Region I represents the region of initial electron heating due to BI (edge of foot); region II has $T_e/T_i \gg 1$ and produces electron heating by the IA instability (main foot); region III represents the ramp with adiabatic electron and ion heating; region IV is the downstream state where the reflected ion energy becomes thermal.

This implies that electron heating will be dominant for downstream values of $T_{e2}/\frac{1}{2}MU_1^2 \approx \frac{1}{8}$, while the rest will be due to ion reflection with $\alpha \approx 0.2$. A schematic of the proposed shock structure is shown in Figure 5. The value of $x = 0$ is at the edge of the 'foot' of the shock, while the main transition ('ramp') is at $x = L_F$. Note that since L_F is controlled by the ion gyroradius at the foot $L_F \lesssim U_1/\Omega_i$. Figure 5 shows the expected profiles of $T_{e,i}/\frac{1}{2}MU_1^2$ and M_F vs x. For $x < 0$ M_F is constant, $T_e/T_i \approx 1$ and $T_e/\frac{1}{2}MU_1^2 \approx 2/M_F^2$. At the edge of the foot ($x \approx 0$) a strong BI sets in between the incoming electrons and the reflected beam. Following the BI theory (Davidson *et al.*, 1970; Lampe *et al.*, 1975; Papadopoulos, 1977) as given by Equations (4–8) the interaction results in a plasma with $T_e/\frac{1}{2}MU_1^2 \approx 2 \times 10^{-3}$, $M_F \approx 22$ and $T_e/T_i \gg 1$. This occurs on an extremely short scale-length of the order $(M/m)^{1/3}\,(U_1/\omega_e) \ll L_F$, where ω_e is the upstream plasma frequency. The proposed model relies in exciting the ion-acoustic instability by the relative streaming U_D between the transmitted ions and the electrons in region II. If we assume that the instability heats the electrons to a value T_e' in front of the ramp, i.e., $T_e' = T_e(L_F)$, the downstream electron temperature $T_{e2} \approx 4T_e'$ is due to adiabatic electron heating at the ramp. The factor 4 comes from assuming the electrons as two-dimensional due to their gyromotion (Papadopoulos, 1985). From the above we see that the required resitive heating at the foot should result in an electron temperature $T_e'/\frac{1}{2}MU_1^2 \approx \frac{1}{32}$, i.e., one-fourth of the $T_{e2} \approx \frac{1}{8}(\frac{1}{2}MU_1^2)$ required by the R–H conditions. Since at this point $\beta_{e1} \gg 1$, the $T_e'/\frac{1}{2}MU_1^2 \approx \frac{1}{32}$ value corresponds to $M_F' \approx 4$–6. These are illustrated in Figure 5. From the above we see that a vital issue of the model is the requirement that the L_F be long enough so that heating due to IA in region II produce the required values of T_e' and M_F'. We compute them by assuming that the potential drop at the foot adjusts itself to maintain marginal stability – i.e.,

$$\frac{U_D}{c_s} \approx \frac{U_D}{V_e}\left(\frac{M}{m}\right)^{1/2} \approx 1 . \tag{11}$$

The spatial heating rate is given by (Lampe *et al.*, 1975; Papadopoulos, 1977)

$$U_1 \frac{\mathrm{d}T_e}{\mathrm{d}x} = \nu^* U_D \,|\, U_D - V_e \,|\, m , \tag{12a}$$

$$\nu^* = \frac{\sqrt{\pi}}{32}\, \omega_e . \tag{12b}$$

Taking into account that following BI $\beta_{e1} \gg 1$, we can take

$$M_F \approx \frac{U_1}{c_s} . \tag{13}$$

From Equations (11)–(13) we find that

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{M_F^2} = \frac{1}{L_0 M_F^2} , \tag{14a}$$

$$L_0^{-1} \equiv \frac{2\sqrt{\pi}}{32}\frac{\omega_i}{U_1} \approx 10\, \frac{\omega_i}{U_1} ; \tag{14b}$$

where ω_i is the upstream ion-plasma frequency. Equation (14) gives

$$M_F(x) = M_F(0) \exp\left(- \frac{x}{2L_0}\right) . \tag{17}$$

In order to reduce M_F from $M_F(0) \approx 22$–25 to $M_F(L_F) \approx 4$–6 we should have

$$\frac{L_F}{L_0} \gtrsim 4 ; \tag{18}$$

or, equivalently, $\omega_i/\Omega_i > 40$. This condition is easily satisfied in most astrophysical plasmas.

5. Concluding Remarks

The paper, dedicated to Hannes Alfvén, is a token evidence that his preachings had and will always have a profound effect in the thinking and approach of subsequent generations of physicists and astrophysicists. He has for a long time advocated that progress can be achieved only by 'breaking down the scientific disciplinary barriers'. This, he argued, is particularly true for astrophysics. The so-called 'visual Universe', the Universe based on observations in the 'visual' octave of the electromagnetic spectrum supplemented by infrared and radio observations, is drastically different from the plasma universe (Alfvén, 1981, 1986a). Information about the plasma universe can be obtained by extraplation of laboratory experiments and in situ measurements of ionospheric and magnetospheric plasmas, but always with extreme care that we do not violate the fundamental principles of plasma physics. To quote him (Alfvén, 1986b),

"The lack of contact between Birkeland's and Langmuir's experimental-theoretical approach on the one hand, and the Chapman–Cowling mathematical-theoretical approach on the other, delayed progress in cosmic plasma physics by perhaps half a century. Many new concepts which came with the space age begin to be understood by magnetospheric physicists, but have not yet reached the textbooks in astrophysics – a delay of one or two decades, often more, as seen in the preceding section. Very few if any deny that (at least by volume) more than 99% of the Universe consists of plasma; but students in astrophysics are kept ignorant even of the existence of important plasma phenomena."

In this paper we attempted to follow his approach. Starting from the understanding of HMN supercritical shocks and current driven instabilities as emerged from laboratory experiments, in situ magnetospheric observations and computer simulations, we proceeded to explore the SHMN regime of astrophysical shocks such as occur in supernovae. The preliminary results of the paper demonstrated that while a 'blind' extrapolation of the HMN results would have predicted predominance of ion heating an extrapolation that 'respected' the fundamental plasma physics concepts gave the surprising result that electron heating predominance returns for sufficiently high Mach numbers. This is supported by the observational evidence. We feel that this paper is only the beginning of understanding the physics of SHMN shocks and we hope that it can set the stage for scientific debate half as controversial as the introduction by Alfvén of concepts such as the critical ionization velocity and double layers.

Acknowledgements

This paper was the result of a 'casual' remark by Dr H. Völk at the International Conference on 'Collisionless shocks', Balatonfüred, Hungary, June 1–5, 1986. I would like to express my thanks to the organizers Prof. K. Szegö and Academician R. Z. Sagdeev for their kind invitation. Discussions on the subject with Acad. R. Z. Sagdeev and a very critical reading of the manuscript by Dr P. Cargill are gratefully acknowledged. The referees' (Drs E. C. Whipple and S. Rogers) remarks have also contributed in improving the clarity of the paper. This work was supported by NASA NAGW-81 and NSF ATM 85-00794.

References

Alfvén, H.: 1981, *Cosmic Plasma*, D. Reidel Publ. Co., Dordrecht, Holland.

Alfvén, H.: 1986a, in H. Garret, J. Feynman, and S. Gabriel (eds.), *Space Technology Plasma Issues*, NASA-JPL Publ. 86–49, p. 373.

Alfvén, H.: 1986b, in H. Garret, J. Feynman, and S. Gabriel (eds.), *Space Technology Plasma Issues*, NASA-JPL Publ. 86–49, p. 373.

Blandford, R. and Eichler, D.: 1987, *Phys. Reports* (in press).

Chodura, R.: 1975, *Nucl. Fusion* **15**, 55.

Davidson, R. C., Kral, N. A., Papadopoulos, K., and Shanny, R.: 1970, *Phys. Rev. Letters* **24**, 579.

Forslund, D. W., Quest, K. B., Brackbill, J. U., and Lee, K.: 1984, *J. Geophys. Res.* **89**, 2142.

Goodrich, C. C.: 1985, 'Collisionless Shocks in the Heliosphere', in R. G. Stone and B. T. Tsurutani (eds.), *Geophys. Monograph* **34**, p. 59.

Greenstadt, E. W., Russell, C. T., Gosling, J. T., Bame, S. J., Paschmann, G., Parks, G. K. Anderson, K. A., Scarf, F. L., Anderson, R. R., Gurnett, D. A., Lin, R. P., Lin, C. S., and Reme, H.: 1980, *J. Geophys. Res.* **85**, 2124.

Ichimaru, S.: 1973, *Basic Principles of Plasma Physics*, W. A. Benjamin, Reading, Mass., pp. 139–142.

Kennel, C. F., Edminston, J. P., and Hada, T.: 1985, 'Collisionless Shocks in the Heliosphere', in R. G. Stone and B. T. Usurutani (eds.), *Geophys. Monograph* **34**, p. 59.

Lampe, M., Manheimer, W. M., and Papadopoulos, K., *et al.*: 1975, NRL Memo Rep. 3076, Naval Research Laboratory, Washington, D.C., pp. 1–82.

Leroy, M.: 1983, *Phys. Fluids* **26**, 2742.

Leroy, M. M., Goodrich, C. C., Winske, D., Wu, C. S., and Papadopoulos, K.: 1981, *Geophys. Res. Letters* **8**, 1269.

Leroy, M. M., Winske, D., Goodrich, C. C., Wu, C. S., and Papadopoulos, K.: 1982, *J. Geophys. Res.* **87**, 5081.

McKee, C. F. and Hollenbach, D. J.: 1980, *Ann. Rev. Astron. Astrophys.* **18**, 219.

Moses, S. K., Coroniti, F. V., Kennel, C. F., and Scarf, F. L.: 1984, *Geophys. Res. Letters* **11**, 183.

Moses, S. L., Coroniti, F. V., Kennel, C. F., and Scarf, F. L.: 1985a, *Geophys. Res. Letters* **12**, 183.

Moses, S. L., Coroniti, F. V., Kennel, C. F., and Scarf, F. L.: 1985b, *Geophys. Res. Letters* **12**, 609.

Papadopoulos, K.: 1977, *Rev. Geophys. Space Phys.* **15**, 113.

Papadopoulos, K.: 1985, 'Collisionless Shocks in the Heliosphere' in R. G. Stone and B. T. Tsurutani (eds.), *Geophys. Monograph* **34**, p. 59.

Paschmann, G. and Sckopke, N.: 1983, in G. Haerendel and B. Battrick (eds.), *Topics in Plasma-Astro and Space Physics*, Max-Planck Inst., München, p. 139.

Paschmann, G., Sckopke, N., Bame, S. J., and Gosling, J. T.: 1982, *Geophys. Res. Letters* **9**, 881.

Quest, K. B.: 1985, *Phys. Rev. Letters* **54**, 1872.

Quest, K. B.: 1986, *J. Geophys. Res.* **91**, 8805.

Russell, C. T., Hoppe, M. N., and Livsey, W. A.: 1982, *Nature* **296**, 45.

Sagdeev, R. Z.: 1966, in M. A. Leontovich (ed.), *Reviews of Plasma Physics*, Vol. 4, Consultants Bureau, New York, pp. 23–90.

Sagdeev, R.: 1979, *Rev. Mod. Phys.* **51**, 1.

Sckopke, N., Paschmann, G., Bame, G., Gosling, S. J., and Russel, C. T.: 1983, *J. Geophys. Res.* **88**, 6121.

Scudder, J. D., Mangeney, A., Gosling, J. T., Paschmann, G., Wu, C. S., Russell, C. T., Rodriguez, P., Anderson, R. R., Bame, S. J., Fitzenreiter, R. J., Harvey, C., Hoppe, M., Lacombe, C., Ogilvie, K. W., and Sckopke, N.: 1986, *J. Geophys. Res.* **91**, A10, 11, 019–11, 052.

Tidman, D. A. and Krall, N. A.: 1971, *Shock Waves in Collisionless Plasmas*, John Wiley, New York, 1–85.

Wu, C. S., Zhou, Y. M., Tsai, S. T., Guo, Winske, D., and Papadopoulos, K.: 1983, *Phys. Fluids* **26**, 1259.

Wu, C. S., Winske, D., Zhou, Y. M., Tsai, S. T., Rodriguez, P., Tanaka, M., Papadopoulos, K. Akimoto, K., Lin, C. S., Leroy, M. M., and Goodrich, C. C.: 1984, *Space Sci. Rev.* **37**, 63.

'DEAD' PULSARS: COSMIC-RAY SOURCES*

K. D. COLE

Department of Physics, La Trobe University, Bundoora, Victoria, Australia

(Received 15 June, 1987)

Abstract. 'Dead' pulsars outnumber 'live' pulsars by a factor of 10^4. It is estimated that there are 3×10^9 of them in our Galaxy. The exospheres of the atmosphere of 'dead' pulsars are characterised by cosmic-ray energies per particle, as the result of accretion of cold particles from interstellar space. Velocities of particles in the exosphere tend to be 'Maxwellianised' by collisions there. The temperature of the exosphere from which particles escape is of the order of 10^{12} K while the temperature of the photosphere closer to the surface of the pulsar is of the order of 10^7 K. Collisions in the exosphere result in Jeans's type escape of cosmic rays with GeV energies at infinity. Two braod ranges of conditions for the exospheres are considered (a) with no magnetic fields involved, and (b) with magnetic fields. Similar conclusions are reached regarding the escape of cosmic rays. Conditions are delineated such that the exospheres of 'dead' pulsars might be major sources of cosmic rays.

1. Dead Pulsars

The minimum population of observable pulsars in our Galaxy is estimated to be 3×10^5 while the maximum lifetime of most pulsars is 3×10^6 yr (Smith, 1977). It may be inferred from this that if pulsars are not annihilated at the end of their pulsing life there are at least 3×10^9 'dead' pulsars in our Galaxy (assumed here to be 3×10^{10} yr old). These object then far outnumber 'live' pulsars by a factor of 10^4. On the assumption that the number of pulsars in any galaxy is proportional to its mass, it follows that the number of 'dead' pulsars in the Universe is about 10^{21}. What are the atmospheres of such objects like?

2. Cosmic-Ray Atmospheres of 'Dead' Pulsars

Consider the atmosphere of a star in which the gravitational force on a particle dominates the electrostatic force engendered by the magnetic field and the star's rotation. Then with the usual symbols (in c.g.s.)

$$\frac{GMm}{r^2} \gg \frac{er\Omega B}{c} \tag{1}$$

or

$$\Omega B < \frac{cGMm}{er^3} .$$

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 549–556.

Such atmospheres complement those assumed for pulsars for which the inequality signs are reversed (e.g., Smith, 1977). These stars may be the product of the evolution of pulsars when they have finished their pulsing phase. For a typical neutron star with $M = M_\odot$ ($\approx 2 \times 10^{33}$ g), inequality (1) implies $\Omega B \ll 1.2 \times 10^4$ for protons and 6.7 for electrons.

We suppose in the first analysis that the kinetic energy density of plasma far exceeds the magnetic energy density so that no magnetic forces have to be taken into account. Such a star will accrete matter from nearby interstellar space and generate an atmosphere which has energy per particle in its 'exosphere' in the cosmic-ray range. This is so because the energy ε of a cold particle of rest mass m_p falling from infinity to radius r from the centre of the star is (McVittie, 1964)

$$\varepsilon = m_p c^2/(1 - 2GM/rc^2)^{1/2} . \tag{2}$$

For a 'dead' pulsar, here assumed to be of radius $a = 10$ km, ε would be 1.3 GeV near its surface. $2GM/c^2$ (the Schwarzschild radius) is ≈ 3 km. The equation represents the upper limit to the mean energy per particle at the base of the exosphere. This corresponds to temperatures of 8.6×10^{12} K for $r \approx 10$ km. Below the exosphere collisions will cause heat conduction that will ensure that there is a temperature gradient from the exosphere to the surface of the dead pulsar where the temperature may be less than 10^7 K (see Section 5).

Let us estimate the scale height of the density of the exosphere by two extreme assumptions. With the first assumption that radial flow of matter is continuous and total accretion of the matter occurs on the surface of the 'dead' pulsar, the flux F across any concentric sphere of radius r is given by

$$F = 4\pi r^2 (2GM/r)^{1/2} \, n = \text{constant} , \tag{3}$$

where n is the density, if few or negligibly few, collisions occur (i.e., down to the base of exosphere). The exosphere then has a density scale height of $2r/3$. This suggests that the scale height of the exosphere is approximately the radius a of the 'dead' pulsar. Later in Section 5 we show that the exosphere is within a distance a of the surface.

In reality the capture would not be precisely by radial flux because the motion of the 'dead' pulsar through the interstellar particles confers on each of them an angular momentum with respect to the pulsar. Therefore, particles would perform curved orbits about the pulsar suffering collisions when entering regions in the exosphere of the pulsar where the particle density is sufficiently great.

At the other extreme we suppose a cosmic-ray exosphere exists around the 'dead' pulsar in hydrostatic equilibrium and is described by the Oppenheimer–Volkoff (OV) equation (Misner *et al.*, 1973), thus

$$\frac{\mathrm{d}p}{\mathrm{d}r} = -\frac{G(\rho c^2 + p)\,(M/c^2 + 4\pi r^3 p/c^4)}{t(r - 2GM/c^2)} , \tag{4}$$

where p = pressure and ρ = density. With the assumption of a tenuous and relativistic exposphere we put $4\pi r^3 p/c^2 \ll M$ and $p \approx \rho c^2/3$.

Then

$$\frac{c^2\,\mathrm{d}\rho}{3\,\mathrm{d}r} = \frac{\mathrm{d}p}{\mathrm{d}r} = -\,\frac{4\rho GM/3}{r^2(1 - 2GM/rc^2)}\,. \tag{5}$$

It follows that, in this exosphere,

$$\frac{\rho(r)}{\rho(r_0)} = \left[\left(\frac{r - 2GM/c^2}{r_0 - 2GM/c^2}\right)\left(\frac{r_0}{r}\right)\right]^2. \tag{5a}$$

Also the exosphere has a characteristic height of

$$L = \frac{c^2 r^2(1 - 2GM/rc^2)}{4GM}\,. \tag{6}$$

Again if r is of the order of a, so is L.

The exosphere of 'dead' pulsars clearly constitute reservoirs of cosmic rays throughout the Universe.

It will be assumed from here on that the density scale height of the pulsar's accreted exosphere is of the order of a.

3. Collisions in the Pulsar's Exosphere

Collisions in the pulsar's exosphere will tend to randomise velocities of the particle causing particles from infinity to change their magnitude and direction and tend to produce a relativistic Maxwellian velocity distribution, by accelerating some and decelerating others. Also a negative temperature gradient in the atmosphere will be produced from the exosphere to the pulsar surface (see Section 5).

'Thermalisation' of the velocity of every incoming particle would be ensured on the first encounter with the atmosphere only if the density n there were equal to or in excess of n_{crit} given by

$$n_{\mathrm{crit}} H\sigma = 1\,, \tag{7}$$

where H is the scale height of the atmosphere and $\sigma(\sim 3 \times 10^{-26}\ \mathrm{cm}^2)$ is the collision cross section for protons in the 1–50 GeV energy range (Perkins, 1972). If collective plasma phenomena effectively increase σ, n_{crit} would correspondingly be reduced.

For exospheric densities in which $n < n_{\mathrm{crit}}$ the probability of 'thermalisation' of a particle's velocity being effected in its first orbit about the 'dead' pulsar is reduced. However, the probability of particles acquiring greater than escape velocities by collisions in such an exosphere is still finite.

The escape flux may be estimated by

$$F_{\mathrm{esc}} = \phi n^2\, 2\pi a^3 \sigma f\,, \tag{8}$$

where ϕ ($\approx cn$) is the flux of cosmic rays at the lowest altitude of escape, $4\pi a^3$ an estimate of the volume in which takes place the 'last' collision before escape and f the fraction of the surface of the star from over which escape is allowed by magnetic fields. A factor of 2 reduction is applied to allow only for up-going particles after collision. Then

$$\begin{aligned} F_{\text{esc}} &= 2\pi a^3 c\sigma n^2 f \quad \text{if} \quad n \le n_{\text{crit}}, \\ &= 5 \times 10^3 n^2 f. \end{aligned} \tag{9}$$

4. Cosmic-Ray Flux Needed from 'Dead' Pulsars

The mass density of cosmic rays in the Universe is approximately $10^{-33}\,\text{g cm}^{-3}$ (Misner *et al.*, 1973) or 3.55×10^{52} g total. In a Universe of characteristic size 2×10^{28} cm and for cosmic rays of characteristic energy GeV, this would imply a rate of production over the life of the Universe ($\sim 3 \times 10^{10}$ yr) of about 6.6×10^{57} particles s^{-1}. It follows there would need to be approximately $1.3 \times 10^{54}\, n^{-2} f^{-1}$ cosmic-ray 'dead' pulsar to account for them. Equating this number to 10^{21}, the number of 'dead' pulsars in the Universe makes

$$n^2 f = 1 \times 10^{33}. \tag{10}$$

If the 'dead' pulsar had no magnetic field then $f = 1$ and $n \approx 3 \times 10^{16}\,\text{cm}^{-3}$ in the exosphere. In this case $F_{\text{esc}} = 5 \times 10^{36}\,\text{s}^{-1}$.

To provide the energy for this escape flux would require at least a similar capture rate of particles from interstellar space. If the atmosphere were approximately static a density of the order of $10^{16}\,\text{cm}^{-3}$ should have been built up within a radius above the surface of the star. In this case the accretion rate to the star would have to be of the same order at least as the escape flux but obviously in excess of it. On the other hand, at the other extreme of possibility for formation of the atmosphere by fall of particles from interstellar space to the star, the necessary influx to build up a density of $3 \times 10^{16}\,\text{cm}^{-3}$ above the surface requires, according to Equation (3), a total accretion rate of $10^{40}\,\text{s}^{-1}$. This is 3 orders of magnitude greater than the escape flux and would imply that in the lifetime of the Universe the mass of the neutron star would approximately double. In this case the necessary flux could be supplied by an interstellar density of 1 proton cm^{-3} at a distance of 5×10^{16} cm or 17 light days from the dead pulsar. Moreover, it would be possible during this process that the mass of the dead pulsar would increase to such an extent ($M > 1.7\, M_\odot$) that they would collapse into black holes. The very highest of energies of cosmic rays are expected to be emitted from the atmosphere of such objects when their radii approach or become less than the Schwarzschild values.

It is interesting to note that in this case the number of black holes in our Galaxy would be expected to be of the order of 10^9, an inference drawn independently from other considerations (Misner *et al.*, 1973).

5. Exosphere and 'Photosphere' of the 'Dead' Pulsar

In addition to the exosphere built up by accretion of matter from interstellar space a photosphere would be generated close to the surface of the pulsar, i.e., well below the exosphere. The temperature and mean energy of particles in the exosphere would be of the order of 10^{12} K and 1 GeV, respectively. However, in the 'photosphere' the temperature would have an upper limit set by radiation balancing incoming accreted energy in the capture flux. Assuming $f = 1$, a capture rate of $10^{40}\ \mathrm{s}^{-1}$ means $8 \times 10^{26}\ \mathrm{cm}^{-2}\ \mathrm{s}^{-1}$. Then, assuming Planck's law, $\sigma T^4_{\mathrm{phot.}} = 8 \times 10^{26}\ \mathrm{GeV\ cm}^{-2}\ \mathrm{s}^{-1}$, produces a photospheric temperature upper limit ($T_{\mathrm{phot.}}$) of 10^7 K. There is a steep temperature gradient between the photosphere and exosphere.

Let us examine this gradient. For the atmosphere below the exosphere we are dealing with temperatures in the range $10^7\ \mathrm{K} \lesssim T \lesssim 10^{11}$ K. To a good approximation then Equation (4) yields assuming $p \ll \rho c^2$ and $4\pi r^2/c^2 \ll M$,

$$\frac{\mathrm{d}p}{\mathrm{d}r} = -\frac{GM\rho}{r(r - 2GM/c^2)} \,. \tag{11}$$

Putting $p = nkT$,

$$\frac{1}{nkT}\frac{\mathrm{d}}{\mathrm{d}r}(nkT) = -\frac{GMm}{kTr(r - 2GM/c^2)} \,; \tag{12}$$

where m is the mass of a proton. For $M = M_\odot$ this equation yields the following atmospheric scale heights H of pressure.

TABLE I

T (K)	10^7	10^8	10^9	10^{10}	10^{11}	10^{12}	10^{13}
H (cm)	5	50	500	5×10^3	5×10^4	5×10^5	10^6

The entries for $T = 10^{12}$ and 10^{13} K in this table are derived from Equation (6).

This table suggests that the 'photosphere' of the 'dead' pulsar is within 10 cm of the surface while the 'exosphere' from which cosmic rays of GeV energies escape is about 10 km thick above the surface.

As a first crude approximation to the altitude profile of density in the 'dead' pulsar's atmosphere we assume $\mathrm{d}T/\mathrm{d}z$ is constant with altitude z in the part below the exosphere. On this basis the following model of the atmosphere of a 'dead' pulsar is arrived at.

The exosphere is not cooled by radiation to temperatures characteristic of the photosphere because the optical depth in the exosphere is much greater than the scale height there. A somewhat similar temperature relationship with a different cause, obviously exists in the exospheres of the Earth and the Sun (2×10^3 K and 2×10^6 K, respectively) in comparison to their photospheres (3×10^2 K and 6×10^3 K, respectively).

Model of atmosphere of 'dead' pulsar

	Photosphere	Transition region					Exosphere
Z (cm)	5	50	500	5×10^3	5×10^4	5×10^5	10^6
T (K)	10^7	10^8	10^9	10^{10}	10^{11}	10^{12}	10^{13}
				$(p = nkT)$			$(p = \rho c^2/3)$
p (dynes cm^{-2})	6×10^{15}	2×10^{14}	8×10^{14}	3×10^{14}	1×10^{14}	4×10^{13}	1.5×10^{13}
n (cm^{-3})	4×10^{24}	1×10^{23}	2×10^{20}	7×10^{18}	3×10^{17}	3×10^{16}	

6. Magnetic Field

If accretion of protons from interstellar space accumulated interstellar magnetic field in the exosphere of the 'dead' pulsar, the magnetic field in this exosphere would be open in the direction of the interstellar field.

In this case a steady state could exist in which there is radial capture of material and magnetic flux from interstellar space, and an equal or smaller escape loss from the 'polar' regions. f would still be approximately unity.

It may be suggested that in the process of capture of magnetic flux by the star a relationship such as 'n/B^α is constant' may apply. For, in the case of movement across a unidirectional magnetic field $\alpha = 1$ and for movement orthogonal to lines of a dipole magnetic field $\alpha = \frac{4}{3}$. Magnetic fields become inhibiting on the escape process if in the exosphere of the 'dead' pulsar

$$B^2/8\pi \gtrsim n\varepsilon . \tag{13}$$

If ε is 1 GeV and $n \sim 3 \times 10^{16}$ cm^{-3}, $B = 3 \times 10^7$ G. If starting with interstellar densities of 1 cm^{-3} and magnetic fields of the order of 10^{-6} G, a density of 3×10^{16} cm^{-3} and magnetic field 3×10^7 G at the 'dead' pulsar would imply $\alpha = 1.15$ which seems not unreasonable. In this case a mechanism would be needed for destroying the magnetic field near the surface of the star. This would need to occur at a rate of about $B^2a^2c/2$ or 1.3×10^{37} ergs s^{-1} in the upper limit, worst, case of $B = 3 \times 10^7$ G. This is three orders of magnitude greater than the energy of the escaping cosmic-ray flux. The energy released by destruction of magnetic flux could be readily radiated from photosphere of temperature 10^7 K (as in Section 5). For values of $B < 3 \times 10^7$ G in the exosphere of the star the value of f would approach 1.

The destruction of incoming magnetic flux from interstellar space would be assisted by the process of magnetic reconnection if the dead pulsar had an ordered magnetic field like a dipole and of strength of the order of 10^7 G in its photosphere. Alternatively, if the dead pulsar has no such magnetic field, collective processes in the exosphere may lead to disorder of the magnetic field and destruction of it at X-type neutral points. It is inferred from this discussion that with or without the involvement of magnetic field the escape flux of cosmic rays will be of similar order, i.e., $f \approx 1$. The bonus provided by magnetic fields is to assist the energisation process of cosmic rays over and above that provided by proton-proton collisions in the exosphere.

7. The Electric Field

It is of interest to estimate the electric field E in the atmosphere of a 'dead' pulsar. Including it in the separate OV equations for ions (protons) and electrons in hydrostatic equilibrium we have

$$n_i eE - \nabla p_i = \frac{4n_i m_i GM/3}{r^2(1 - 2GM/rc^2)}, \tag{14}$$

$$-n_e eE - \nabla p_e = \frac{4n_e m_e GM/3}{r^2(1 - 2GM/rc^2)}. \tag{15}$$

Addition of (14) and (15) gives, assuming electrical neutrality,

$$\nabla p = -\frac{4\rho GM/3}{r^2(1 - 2GM/rc^2)}. \tag{16}$$

Assuming the term on the right-hand side of Equation (15) is small compared to other terms on account of the relative smallness of the electron mass, and that $p_e \approx p_i$, we find that

$$eE = \frac{2m_i GM/3}{r^2(1 - 2GM/rc^2)}. \tag{17}$$

It follows from this that the total potential drop across the 'exosphere' of our model 'dead' pulsar would be approximately 1 GeV.

8. Discussion and Conclusions

The atmospheres of 'dead' pulsars considered here are vastly different from those of 'live' pulsars in which the magnetic field is of the order of 10^{12} G and the temperatures of what little atmosphere there is is of the order of 10^6 K (e.g., Smith, 1977). Such a magnetic field would exclude the capture of the atmosphere such as is suggested here for 'dead' pulsars.

It is concluded that 'dead' pulsars in which the magnetic field has decreased to values of the order of 10^7 G or below support cosmic-ray exospheres energised by gravitational capture of interstellar gas. Collisions in the exosphere cause velocity spread of captured particles which can escape from the exosphere with cosmic-ray energies at infinity. The photosphere of the dead pulsars close to the surface has a temperature of only 10^7 K. Of the particles captured from interstellar space approximately 1 to 0.1% are returned as cosmic rays through Jeans-type escape from the exosphere. The bulk of the energy captured is radiated from the photosphere, while the residual particles add to the mass of the 'dead' pulsar perhaps transforming it to candidacy for a black hole in the lifetime of the Universe.

There are 10^4 more 'dead' pulsars than 'live' ones (i.e., ones radiating like conventionally known pulsars). For this reason alone they should be objects of considerable astrophysical interest. Conditions have been delineated for their exospheres to be major sources of cosmic rays.

References

McVittie, G. C.: 1964, *The International Astrophysics Series*, Chapman and Hall Ltd, London.
Misner, C. W., Thorne, K. S., and Wheeler, J. A.: 1973, *Gravitation*, W. H. Freeman and Co., San Francisco.
Perkins, D. H.: 1972, *Introduction to High Energy Physics*, Addison-Wesley, London.
Smith, F. G.: 1977, *Pulsars,* Cambridge Univ. Press, Cambridge.

THE EFFECTS OF MUTUAL IRRADIATION ON THE OBSERVED RADIAL VELOCITY OF THE COMPONENTS OF CLOSE BINARY SYSTEMS*

ZDENĚK KOPAL

Department of Astronomy, University of Manchester, England

(Received 31 July, 1987)

Abstract. The aim of the present paper will be to investigate the effects, on the observed radial velocities of the components of close binary systems, of atmospheric motions caused by mutual irradiation of the two stars. Such motions can (and, in general, will) produce systematic differences between the observed radial velocity and that of the centre of mass of the respective star – differences varying with with the phase and thus giving rise to spurious deformations of the star's radial-velocity curves due to orbital motion. A failure to separate the two could (and, in general, will) vitiate the physical elements deduced from these curves – such as the masses or absolute dimensions of the components and of the shape of their orbit; but in order to do so, an investigation of atmospheric motions invoked by irradiation becomes a necessary prerequisite.

In the Introduction following this abstract, the problem at issue will be described in general terms, and phenomena outlined which should arise in this connection (together with the observations indicating their presence). In Section 2, general expressions for the radial velocity at any point of stellar surface arising from atmospheric motions will be formulated; while Section 3 will isolate such velocities for components of close binary systems as are produced by mutual irradiation of their mates, in terms of hydrodynamical equations of radiative transfer describing the problem. In Sections 4 and 5, the effects of non-rotational motions on the observed radial velocities will be specified, and hydrodynamical equations formulated which specify atmospheric convection caused by irradiation of each component of a close binary by its mate. Linearized versions of such equations will be constructed in Section 6; while Section 7 contains an evaluation of the effects which such gas streams exert on the observed radial velocity of the stars.

In the concluding Section 8 applications to practical cases are carried out. It will be shown that no reliable spectroscopic elements of close binary systems (including the masses and absolute dimensions of their components) can be obtained until the effects of atmospheric convection caused by mutual irradiation have been accounted for to permit us to convert the observed radial velocities (influenced as they are by the motion of as in which they originate) to those of the centre of mass of the respective stars.

1. Introduction

The observations of radial-velocity changes (as disclosed by Doppler shifts of different lines in the spectra) have come to complement the results of photometric observers of close binary systems ever since 1889, when H. C. Vogel discovered Algol to be also a 'spectroscopic binary', whose conjunctions coincided with the minima of light (cf. Vogel, 1890).

The basic methods for an interpretation of the observed radial-velocity changes in terms of the elements of the Keplerian motion were largely established between 1890–1910 by a considerable number of investigators, among whom Wilsing (1893) or Lehmann-Filhés (1894), followed by Schwarzschild (1900), Nijland (1903), Zurhellen

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

(1907, 1908), Curtis (1908), King (1908), or Plummer (1908) deserve especial mention; for a succint brief summary of their work, see Section 3 of Kopal (1980a). However, in all work referred to above, the 'founding fathers' of our subject saw no reason to treat, not only wide, but also close binaries as systems consisting of light points, and tacitly identified the radial velocities deduced from Doppler shifts with radial components of the respective Keplerian motions. For wide binaries, whose components can be regarded as spherical and exhibiting distribution of brightness on their apparent discs which remain symmetrical with respect to their projected centres of mass, a straightforward application of the theory of Keplerian motion becomes indeed expressible in closed form.

However, since the 1930's it has gradually become obvious that the axial rotation and mutual tidal action is bound to cause the components of close binaries to deviate from spherical form; and to render the distribution of brightness over their apparent discs not only non-uniform, but variable in the course of each orbital cycle.

In more specific terms, if V stands for the magnitude of the radial component of the velocity-vector at any surface pont of a distorted star, the effect δV of V on the observed radial velocity of the star should be represented by the ratio

$$\delta V \equiv \frac{\int V \, \mathrm{d}l}{\int \mathrm{d}l} \tag{1.1}$$

of the local value of V averaged over the star's apparent disc (or, for eclipsing stars, a visible fraction thereof), with respect to the light element

$$\mathrm{d}l = J \cos \gamma \, \mathrm{d}\sigma \tag{1.2}$$

at that pont, of brightness J and surface element $\mathrm{d}\sigma$ inclined at an angle γ to the line-of-sight.

The denominator

$$\int \mathrm{d}l \equiv \mathfrak{L} \tag{1.3}$$

on the right-hand side of Equation (1.1) is obviously identical with the total luminosity of the respective component – constant if the star is spherical, and varying with phase during eclipses (if any) or as a result of its distortion.

The explicit form of $\mathfrak{L}$ has already been investigated under very general conditions (cf. Kopal, 1942; or 1979, Chapter II) and need not to be reproduced in this place; while in order to evaluate the numerator on the right-hand side of (1.1), we should remember that the surface element $\mathrm{d}\sigma$ is given by

$$\cos \beta \, \mathrm{d}\sigma = r^2 \sin \theta \, \mathrm{d}\theta \, \mathrm{d}\phi \tag{1.4}$$

in spherical polar coordinates θ and ϕ, where r stands for the radius-vector connecting the centre of mass of the respective star with an arbitrary point of its (distorted) surface, and β is the angle between that radius-vector and surface normal to the element $d\sigma$.

A study of the effects, on observed radial velocities, invoked by axial rotation of distorted stars commended with a note entitled, 'A Source of Spurious Eccentricity in Spectroscopic Binaries', in which Sterne (1941) ponted out that such effects produced by second-harmonic tidal distortion can simulate those of orbital eccentricity if the apsidal line of the respective orbit were parallel with the line-of-sight. Sterne's work was subsequently generalized by the present writer (Kopal, 1945) by a more general method to include the effects of the third and fourth tidal harmonic distortion, and extended by Kitamura and Kopal (1965) to all higher harmonics (odd as well as even) up to the eighth – applicable to close binary systems in full light as well as within eclipses.

In doing so we found that (outside eclipses) the effects of odd-harmonic tides will influence the observed amplitudes of radial-velocity changes (and, through them, the absolute masses and dimensions of the respective system); while even-harmonic tides simulate spuriously eccentric orbits characterized by a longitude of periastron of $90°$ or $270°$.

As long as the respective partial tides are of the equilibrium type, their effects on the observed radial velocities remain, however, quantitatively small. Even in very close systems they can make the amplitudes of radial velocity changes of their components differ from those corresponding to the orbital motions of their centres of mass by a few – at most, several – percent; and simulate a spurious eccentricity of their orbits attaining (say) 0.1, but scarcely more. Their magnitudes are, therefore, insufficient to account – e.g., for a spurious eccentricity e as large as 0.47 (and longitude of periastron of $\omega = 25°$) exhibited by U Cephei (cf. Carpenter, 1930) – a result subsequently confirmed by Struve (1944). Moreover, Neubauer and Struve (1945) detected another eclipsing system – RZ Scuti – in which the same phenomenon is even more conspicuous. In such cases, additional phenomena – supplementary to equilibrium tides – must be identified to account for the observed facts; and, fortunately, they are not hard to find: namely the effects caused by mutual *irradiation* of the components in close binary systems (commonly referred to as the 'reflection effect').

First, it should be stressed that such an effect is bound to make the observed radial velocity of each component differ from that caused by orbital motion merely because of the fact that the varying phase of illumination will make the 'centre of light' of the apparent irradiated disc oscillate about the projected mass-centre even if the stars were spherical, and its projected disc a circle (cf. Kopal, 1943; or Chapter V of Kopal, 1959). Moreover, while a high rate of axial rotation of stars unaffected by tides will merely widen the lines in their spectra without deforming their symmetry, phase-effects due to irradiation will produce asymmetric lines even if tidal distortion is absent.

Furthermore – as is well known – the mean radial velocities of the components at any phase can be deduced from Doppler shifts of the cores of spectral lines (in absorption or emission) formed in the semi-transparent outer fringe ('exosphere') of the star in question. However, to identify these with the Doppler shifts due to orbital motion entails a tacit assumption that *the gaseous layers in which the measured spectral lines originate* *are at rest with respect to the centre of mass of the respective star.* Quite apart from the situations that may obtain for single stars, however, can this be the case with binary systems in which their components irradiate each other from close proximity?

This question is one of great importance; for systems are indeed known – such as the already-mentioned U Cep (cf. Carpenter, 1930; Struve, 1944) or RZ Sct (Neubauer and Struve, 1945; Struve, 1949) – whose real orbits (as evidenced by photometric data on these eclipsing variables) are essentially *circular*, but whose radial-velocity curves are *asymmetric* to such an extent as to simulate an orbital eccentricity e in excess of 0.4. Equal spacing of the primary and secondary minima of light of these eclipsing systems leaves no room for doubt that the observed asymmetry of the radial-velocity curves cannot be caused by orbital eccentricity, but to some other reason producing systematic differences between the radial velocity of the centre of mass of the respective star; and the most obvious is the motion of gas relative to it in which the observed spectral lines originate.

Struve (and others) sought the origin of such anomalous Doppler shifts in the absorption of gas streams between components of the respective systems. Such an explanation – while not out of the question – encounters serious physical difficulties (for their description, cf., e.g., pp. 422–424 of Kopal, 1978; or pp. 535–558 of Kopal, 1981) encouraging further inquiry into alternative possibilities to account for the observed facts. Physical problems arising in connection with gas-streams detached from the stars would indeed be simplified if such streams were located in the *atmospheres* themselves of the respective stars, and kept in motion by absorption of heat incident on their surface by their mates.

As is well known, the first theoretical formulation of such a problem was given by Kirbiyik and Smith (1976). However, because of physical complexities underlying such a problem these investigators were unable to arrive at any definite conclusions regarding the magnitude of the Doppler shifts which would affect the observed radial velocities as a result of 'reflection-produced' exospheric motions. Their discussion did not, however, rule out a possibility that such effects could become quite large – much larger than those produced by surface distortion.

The second investigator of this problem was the present writer (cf. Kopal, 1980a, c), who did so on the assumption that convective motions caused by irradiation possess spheroidal symmetry (cf. Section 2 of Kopal, 1980b) with respect to the line joining the centres of the two stars; but the radial parts of their amplitudes were still regarded as arbitrary. This could again be true in case of free spheroidal oscillations of the components, but would have nothing to do with the effects of irradiation. That atmospheric effects caused by such an irradiation should be symmetric with respect to the orbital radius-vector is only to be expected, because this vector constitutes also the axis of symmetry of the illuminating light beam. Thermal effects of incident radiation are, however, bound to give rise to a global system of gas currents which are not symmetrical with the line-of-sight, and not limited solely to the 'daylight' hemisphere when the illuminating source is above the horizon.

For should axial rotation of the irradiated star be synchronized with the orbit, its entire surface would be divided into two hemispheres of permanent 'day' or 'night'; with warm gas rising above the daytime hemisphere and sinking over those parts of the surface where the illuminating source of light is below the horizon. Should, however, asynchronism exist between rotation and revolution, each point on the surface of the illuminated star would experience an alternation of days and nights – giving rise to atmospheric streaming which is bound to persist even at places where the intake of heat by the companion star is cut-off by onset of night. A 'dynamical meteorology' in the atmospheres of the components of close binary systems – as distinct from those of single stars whose external energy input is negligible – represents still a very largely virgin field, which we shall consider in Section 3 of the present paper.

2. Coordinate Systems of the Problem

However, before we do so, we must establish a proper analytical basis for such an investigation by introducing four distinct systems of rectangular coordinates, with origin at the centre of mass of our binary systems, specified by the following definitions:

(1) The (unprimed) axes XYZ will represent a system of *inertial* coordinates ('space axes') of directions fixed in space in such a way that the XY-plane coincides with the *invariable plane of the system*; while the Z-axis is perpendicular to it.

(2) The singly-primed axes $X'Y'Z'$ will stand for a system of rectangular coordinates *rotating* with the body ('body-axes'), defined so that $X'Y'$-plane represents the (instantaneous) *equator* of the rotating star, inclined by an angle θ to the inertial XY-plane and intersecting it at the angle ϕ (see Figure 1).

(3) The doubly-primed axes $X''Y''Z''$ will hereafter represent a system of *revolving* rectangular coordinates, in which the X''-axis is constantly coincident with the radius-

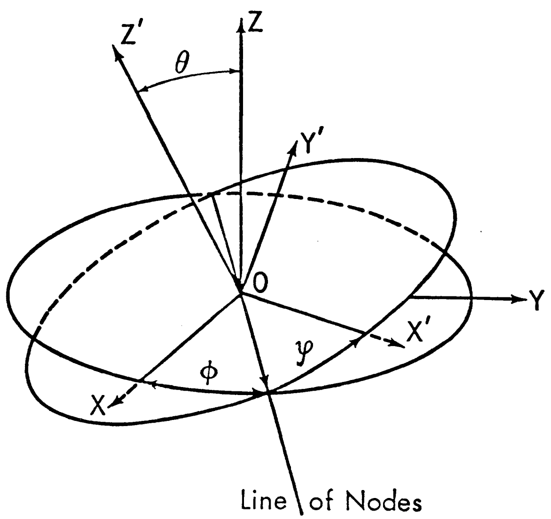


Fig. 1. Definition of Eulerian angles.

vector between the origin and the centre of mass of the revolving star, and $Z'' = 0$ represents the (instantaneous) position of the orbital plane.

(4) The triply-primed axes $X''' Y''' Z'''$ will, lastly, represent another system of moving coordinates, in which the Z'''-axis coincides with the *line-of-sight*, and the X'''-axis is made to coincide with the projection of the X''-axis on the plane tangent to the celestial sphere passing through the origin of coordinates (i.e., $Z''' = 0$).

As is well known, a transformation of coordinates from the inertial (space) to the rotating (body) axes is governed by the matrix equation

$$\begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \begin{Bmatrix} a'_{11} & a'_{12} & a'_{13} \\ a'_{21} & a'_{22} & a'_{23} \\ a'_{31} & a'_{32} & a'_{33} \end{Bmatrix} \begin{Bmatrix} x' \\ y' \\ z' \end{Bmatrix} \tag{2.1}$$

(and its inverse follows from a transposition of the columns and rows of the square matrix on the right-hand side), where the direction cosines

$$\left.\begin{aligned} a'_{11} &= \cos\varphi\cos\phi - \sin\varphi\sin\phi\cos\theta, \\ a'_{12} &= -\sin\varphi\cos\phi - \cos\varphi\sin\phi\cos\theta, \\ a'_{13} &= \sin\phi\sin\theta; \end{aligned}\right\} \tag{2.2}$$

$$\left.\begin{aligned} a'_{21} &= \cos\varphi\sin\phi + \sin\varphi\cos\phi\cos\theta, \\ a'_{22} &= -\sin\varphi\sin\phi + \cos\varphi\cos\phi\cos\theta, \\ a'_{23} &= -\cos\phi\sin\theta; \end{aligned}\right\} \tag{2.3}$$

$$\left.\begin{aligned} a'_{31} &= \sin\varphi\sin\theta, \\ a'_{32} &= \cos\varphi\sin\theta, \\ a'_{33} &= \cos\theta; \end{aligned}\right\} \tag{2.4}$$

where the Eulerian angles ϕ, θ, φ are defined by a scheme illustrated on the accompanying Figure 1.

A transformation of the inertial to revolving coordinates is similarly governed by the matrix equation

$$\begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \begin{Bmatrix} a''_{11} & a''_{12} & a''_{13} \\ a''_{21} & a''_{22} & a''_{23} \\ a''_{31} & a''_{32} & a''_{33} \end{Bmatrix} \begin{Bmatrix} x'' \\ y'' \\ z'' \end{Bmatrix} \tag{2.5}$$

where the direction cosines

$$\left.\begin{aligned} a''_{11} &= \cos u\cos\Omega - \sin u\sin\Omega\cos i, \\ a''_{12} &= -\sin u\cos\Omega - \cos u\sin\Omega\cos i, \\ a''_{13} &= +\sin\Omega\sin i; \end{aligned}\right\} \tag{2.6}$$

$$\left.\begin{aligned} a''_{21} &= \cos u \sin\Omega + \sin u \cos\Omega \cos i\,, \\ a''_{22} &= -\sin u \sin\Omega + \cos u \cos\Omega \cos i\,, \\ a''_{23} &= \qquad\qquad - \qquad\quad \cos\Omega \sin i\,; \end{aligned}\right\} \tag{2.7}$$

$$\left.\begin{aligned} a''_{31} &= \sin u \sin i\,, \\ a''_{32} &= \cos u \sin i\,, \\ a''_{33} &= \cos i\,; \end{aligned}\right\} \tag{2.8}$$

where Ω denotes the angle of the nodes (i.e., of intersection of the $Z = 0$ and $Z'' = 0$ planes measured from the X-axis); i, the inclination of the orbital ($Z'' = 0$) to the invariable ($Z = 0$) plane of the system; and u, the angle between the line of the nodes and the instantaneous position of the radius vector*.

Accordingly, a transformation from the rotating to the revolving axes obeys the matrix equation

$$\begin{Bmatrix} x' \\ y' \\ z' \end{Bmatrix} = \begin{Bmatrix} b''_{11} & b''_{12} & b''_{13} \\ b''_{21} & b''_{22} & b''_{23} \\ b''_{31} & b''_{32} & b''_{33} \end{Bmatrix} \begin{Bmatrix} x'' \\ y'' \\ z'' \end{Bmatrix}, \tag{2.9}$$

where the direction cosines b''_{ij} are given by

$$\begin{Bmatrix} b''_{1j} \\ b''_{3j} \\ b''_{3j} \end{Bmatrix} = \begin{Bmatrix} a'_{11} & a'_{21} & a'_{31} \\ a'_{12} & a'_{22} & a'_{32} \\ a'_{13} & a'_{23} & a'_{33} \end{Bmatrix} \begin{Bmatrix} a''_{1j} \\ a''_{2j} \\ a''_{3j} \end{Bmatrix} \tag{2.10}$$

for $j = 1, 2, 3$.

Lastly, let the transformation between the doubly- and triply-primed systems of rectangular coordinates be given by the matrix equation

$$\begin{Bmatrix} x'' \\ y'' \\ z'' \end{Bmatrix} = \begin{Bmatrix} l_2 & l_1 & l_0 \\ m_2 & m_1 & m_0 \\ n_2 & n_1 & n_0 \end{Bmatrix} \begin{Bmatrix} x''' \\ y''' \\ z''' \end{Bmatrix}, \tag{2.11}$$

where l_0, m_0, n_0 are the direction cosines of the line-of-sight (i.e., the Z'''-axis) in the doubly-primed system of coordinates. Moreover, as the X'''-axis has been defined as a projection of the X''-axis on the $Z''' = 0$ plane, it follows that the direction cosines of

* The reader may note that, if, in Equations (2.2)–(2.4) defining the singly-primed direction cosines a'_{ij} we set $\phi = \Omega$, $\rho = u$, and $\theta = i$, they become identical with the doubly-primed direction cosines a''_{ij}.

the X'''-axis in the doubly-primed system of coordinates on the right-hand side of Equation (2.11) are

$$l_0, \qquad l_1 = 0, \qquad l_2 = \sqrt{1 - l_0^2}; \tag{2.12}$$

while the remaining direction cosines m_j and n_j ($j = 1, 2$) follow in terms of those for $j = 0$ from the orthogonality conditions

$$\left.\begin{aligned} l_0 l_1 + m_0 m_1 + n_0 n_1 &= 0, \\ l_0 l_2 + m_0 m_2 + n_0 n_2 &= 0, \\ l_1 l_2 + m_1 m_2 + n_1 n_2 &= 0; \end{aligned}\right\} \tag{2.13}$$

which combined with (2.12) yield

$$m_1 = -n_0/l_2, \qquad n_1 = m_0/l_2 \tag{2.14}$$

and

$$m_2 = -l_0 m_0/l_2, \qquad n_2 = -l_0 n_0/l_2. \tag{2.15}$$

A transformation of the inertial into triply-primed coordinates can be performed with the aid of the matrix equation

$$\begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \begin{Bmatrix} a'''_{11} & a'''_{12} & a'''_{13} \\ a'''_{21} & a'''_{22} & a'''_{23} \\ a'''_{31} & a'''_{32} & a'''_{33} \end{Bmatrix} \begin{Bmatrix} x''' \\ y''' \\ z''' \end{Bmatrix}, \tag{2.16}$$

where

$$\begin{Bmatrix} a'''_{1j} \\ a'''_{2j} \\ a'''_{3j} \end{Bmatrix} = \begin{Bmatrix} a''_{11} & a''_{12} & a''_{13} \\ a''_{21} & a''_{22} & a''_{23} \\ a''_{31} & a''_{32} & a''_{33} \end{Bmatrix} \begin{Bmatrix} l_{3-j} \\ m_{3-j} \\ n_{3-j} \end{Bmatrix}, \tag{2.17}$$

$j = 1, 2, 3$.

Which ones of the 12 axes specifying our four rectangular systems of coordinates are inertial (i.e., possess directions which are invariant in time)? Only four: namely, X, Y, Z, and Z'''. The singly- and doubly-primed axes rotate and revolve in space – and so do the X'''- and Y'''-axes of the triply-primed system – because the Eulerian angles ϕ, θ, and ϕ contained in the direction cosines a'_{ij} as well as the Keplerian elements Ω, i, and u contained in the a''_{ij}'s are, in general, functions of the time. An exception is the angle of inclination of the invariable ($Z = 0$) plane of the binary system to the plane $Z''' = 0$ tangent to the celestial sphere, which will hereafter be denoted by I. Since both Z-and Z'''-axes are inertial, the angle I should be independent of the time (or its value may change but slowly as our terrestrial observing station and the respective binary may change their relative positions due to their peculiar motions in the Galaxy).

As was already stated before, the inertial XYZ system of coordinates has been fixed so that the XY-plane coincides with the invariable plane of the system; but the directions of the X- and Y-axes in this plane have not been specified so far. In order to remove this arbitrariness, *let us constrain the X-axis to lie in the ZZ'''-plane*. If so, however, the direction cosines of the Z'''-axis in the inertial frame of reference will be given by

$$a'''_{13} = \sin I, \qquad a'''_{23} = 0, \qquad a'''_{33} = \cos I; \tag{2.18}$$

and the cosines l_0, m_0, n_0 of angles between the revolving X''-, Y''-, and Z''-axes and the line-of-sight Z''' can be expressed as

$$\left.\begin{aligned} l_0 &= a''_{11}\sin I + a''_{31}\cos I, \\ -m_0 &= a''_{12}\sin I + a''_{32}\cos I, \\ n_0 &= a''_{13}\sin I + a''_{33}\cos I, \end{aligned}\right\} \tag{2.19}$$

which on insertion for a''_{1j} and a''_{3j} from Equations (2.6) and (2.8) yield

$$l_0 = (\cos u \cos\Omega - \sin u \sin\Omega \cos i)\sin I + \sin u \sin i \cos I, \tag{2.20}$$

$$m_0 = (\sin u \cos\Omega + \cos u \sin\Omega \cos i)\sin I - \cos u \sin i \cos I, \tag{2.21}$$

$$n_0 = \sin\Omega \sin i \sin I + \cos i \cos I. \tag{2.22}$$

If, moreover, we define new angles ψ and j by setting

$$\left.\begin{aligned} l_0 &= \cos\psi \sin j, \\ m_0 &= \sin\psi \sin j, \\ n_0 &= \cos j, \end{aligned}\right\} \tag{2.23}$$

the angle j stands evidently for *instantaneous* inclination of the orbital plane to the celestial sphere (obtainable for eclipsing systems from a solution of their light curves for geometrical elements, by methods expounded in Kopal's 1979 monograph); and ψ, for the 'phase angle' in elliptical orbit. As is well known, in close binary systems the angles Ω and i in Equations (2.20)–(2.22) will, in general, be functions of the time – Ω secularly regressing and i oscillating on account of nutation (cf. Kopal, 1978; Chapter V) – so that, in accordance with Equation (2.22), the actual inclination j of the orbital plane to the celestial sphere should oscillate between $I \pm i$ as Ω runs from 0 to 2π; and will remain secularly constant if $i = 0$ (which can, in turn, be true only if the equators of both components of a close binary system are coplanar with the orbit).

3. Radial Motions on Rotating Stellar Discs

With geometric preliminaries necessary for the treatment of our problem now completed, let us turn our attention first to the motion of the *centre of mass* of one (say, the primary) component about the centre of mass of the system. Consistent with our preceding definitions, the doubly-primed coordinates x'', y'', z'' of the mass centre of the revolving star will be given by

$$x'' = r, \qquad y'' = 0, \qquad z'' = \theta; \tag{3.1}$$

where r denotes the radius-vector of the absolute orbit of that star; and, moreover, its distance z'' from the plane normal to the celestial sphere which passes through the mass-centre of the system follows from Equations (2.11) and (2.19) is given by

$$z'' \equiv l_0 x'' = r(a''_{11} \sin I + a''_{31} \cos I), \tag{3.2}$$

where the angle I between the (inertial) planes $Z = 0$ and $Z'' = 0$ can be treated as a constant. Accordingly, the velocity V_* of the star in its orbit along the line-of-sight – i.e., the radial velocity of its centre of mass – should then follow as the time-derivative of Equation (3.2) in the form

$$\begin{aligned} V_* &\equiv \gamma + \dot{z}'' = \\ &= \gamma + \dot{r}(a''_{11} \sin I + a''_{31} \cos I) + r(\dot{a}''_{11} \sin I + \dot{a}''_{31} \cos I), \end{aligned} \tag{3.3}$$

where γ stands for the radial velocity of the centre of mass of the system in space (constant in the absence of any other body), and the dots above r and a_{ij} denote ordinary time-derivatives of the respective quantities.

An evaluation of these derivatives is classical (cf., e.g., pp. 330–334 of Kopal, 1980a); and its results furnish the basis for all well-known methods of the computation of orbital elements of 'spectroscopic' binary systems whose components can be regarded as light points. Should, however, the rotating stars exhibit discs of finite angular diameter, the distance of any point on the star's surface from a plane tangent to the line-of-sight and passing through the star's centre of mass will be given by

$$z'' = a''_{13}x + a''_{23}y + a''_{33}z, \tag{3.4}$$

where the a''_{13}'s stand for the direction cosines of the line-of-sight. Moreover, since the orientation of the latter is invariant in space, a time-differentiation of Equation (3.4) yields for the radial velocity V of any arbitrary point on the star's surface the expression

$$V \equiv \dot{z}'' = a''_{13}\dot{x} + a''_{23}\dot{y} + a''_{33}\dot{z}; \tag{3.5}$$

where by a differentiation of the relations (2.1) and of the direction cosines (2.2)–(2.4) we find (cf. Chapter IV of Kopal, 1978) that

$$\dot{x} \equiv u = z\omega_y - y\omega_z + u'_0, \tag{3.6}$$

$$\dot{y} \equiv v = x\omega_z - z\omega_x + v'_0, \tag{3.7}$$

$$\dot{z} \equiv w = y\omega_x - x\omega_y + w'_0, \tag{3.8}$$

where the angular velocities $\omega_{x,y,z}$ of axial rotation are related with the time-derivatives of the Eulerian angles θ, ϕ, and φ by

$$\omega_x = \dot{\theta}\cos\phi + \dot{\varphi}\sin\theta\sin\phi\,, \tag{3.9}$$

$$\omega_y = \dot{\theta}\sin\phi - \dot{\varphi}\sin\theta\cos\phi\,, \tag{3.10}$$

$$\omega_z = \qquad + \dot{\varphi}\cos\theta \qquad + \dot{\phi}\,; \tag{3.11}$$

and

$$\begin{Bmatrix} u'_0 \\ v'_0 \\ w'_0 \end{Bmatrix} = \begin{Bmatrix} a'_{11} & a'_{12} & a'_{13} \\ a'_{21} & a'_{22} & a'_{23} \\ a'_{31} & a'_{32} & a'_{33} \end{Bmatrix} \begin{Bmatrix} u' \\ v' \\ w' \end{Bmatrix}, \tag{3.12}$$

where

$$u' \equiv \dot{x}'\,, \qquad v' \equiv \dot{y}'\,, \qquad w' \equiv \dot{z}'\,, \tag{3.13}$$

are the *body* velocity components in the *rotating* (single-primed) coordinates. Accordingly, the quantities u'_0, v'_0, and w'_0 on the right-hand sides of Equations (3.6)–(3.8) as defined by Equation (3.12) represent the *body* velocity components in the direction of the *fixed* space axes X, Y, and Z. Should our configuration rotate as a *rigid* body, the u', v', w' – and, therefore, u'_0, v'_0, w'_0 on the right-hand sides of Equations (3.6)–(3.8) – would be identically zero; and can become finite only if the body in question is *deformable* (or, for spectroscopic binaries, if the medium in which spectral lines are formed is in motion relative to the centre of mass of the respective configuration).

The coordinates x, y, z factoring the angular velocities ω_x, ω_y, ω_z, on the right-hand sides of Equations (3.6)–(3.8) refer to the inertial space axes. In order to facilitate the tasks awaiting us in the subsequent two sections of this paper we find it, however, of advantage to rewrite Equations (3.5)–(3.8) in terms of the doubly-primed (revolving) and triply-primed coordinates and velocity components. If we ignore – for the time being – the velocity components u'_0, v'_0, w'_0 of non-rotational origin and convert x, y, z in x'', y'', z'' with the aid of the transformation equations (2.5), the radial velocity component $V_{\rm rot}$ arising from rigid-body rotation can be expressed as

$$V_{\rm rot} = \omega_{x''}(n_0 y'' - m_0 z'') + \omega_{y''}(l_0 z'' - n_0 x'') + \omega_{z''}(m_0 x'' - l_0 y'')\,, \tag{3.14}$$

where

$$\begin{Bmatrix} \omega_{x''} \\ \omega_{y''} \\ \omega_{z''} \end{Bmatrix} = \begin{Bmatrix} a''_{11} & a''_{21} & a''_{31} \\ a''_{12} & a''_{22} & a''_{32} \\ a''_{13} & a''_{23} & a''_{33} \end{Bmatrix} \begin{Bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{Bmatrix} \tag{3.15}$$

are the angular velocity components about the doubly-primed axes, and the direction cosines l_0, m_0, n_0 of the line-of-sight in the same coordinates continue to be given by Equations (2.20)–(2.22).

If, lastly, we wish to rewrite the expression (3.14) for V_{rot} in terms of the triply-primed coordinate system x''', y''', z''', a resort to the transformation equation (2.11) discloses that, in rectangular coordinates of the latter type,

$$n_0 y'' - m_0 z'' = l_1 x''' - l_2 y''' , \tag{3.16}$$

$$l_0 z'' - n_0 x'' = m_1 x''' - m_2 y''' , \tag{3.17}$$

$$m_0 x'' - l_0 y'' = n_1 x''' - n_2 y''' , \tag{3.18}$$

where the direction cosines $l_{1,2}$, $m_{1,2}$, and $n_{1,2}$ continue to be given by Equations (2.12) and (2.14)–(2.15); so that

$$V_{\text{rot}} = \omega_{x''}(-l_2 y''') + \omega_{y''}(m_1 x''' - m_2 y''') + \omega_{z''}(n_2 x''' - n_2 y''') , \tag{3.19}$$

since $l_1 = 0$.

The foregoing expressions (3.14) or (3.19) for V_{rot} are exact for any (not necessarily constant) values of the angular velocities $\omega_{x''}$, $\omega_{y''}$, $\omega_{z''}$ of our body about the revolving axes. Should, however, the equators of the respective components coincide with the plane of their orbit, then the Eulerian angle θ in Equations (3.9)–(3.11) vanishes – a fact which renders $\omega_x = \omega_y = 0$. Since, moreover, then also $i = 0$, it follows that $\omega_{x''} = \omega_{y''} = 0$ as well and $\omega_{z''} = \omega_z$ – as a result of which Equations (3.14) and (3.19) reduce to

$$V_{\text{rot}} = \omega_z(m_0 x'' - l_0 y'') = \omega_z(n_1 x''' - n_2 y''') , \tag{3.20}$$

of which extensive use will be made in Sections 5 and 6.

By this the tasks set forth in the present section are, however, not yet complete; for in order to construct a complete expression for the radial velocity V at any time and any surface point of the components of close binary systems we must adjoin to V_{rot} the contribution $V^{(+)}$ arising from possible deformability or atmospheric motions of the respective star, given by

$$V^{(+)} = a'''_{13} u'_0 + a'''_{23} v'_0 + a'''_{33} w'_0 , \tag{3.21}$$

which in view of Equations (2.18) and (3.12) can be rewritten as

$$\begin{aligned} V^{(+)} &= u'_0 \sin I + w'_0 \cos I = \\ &= (a'_{11} u' + a'_{12} v' + a'_{13} w') \sin I + (a'_{31} u' + a'_{32} v' + a'_{33} w') \cos I , \end{aligned} \tag{3.22}$$

where the singly-primed direction cosines a'_{ij} continue to be given by Equations (2.2)–(2.4).

For $\theta = 0$, these direction cosines reduce to

$$a'_{11} = \cos(\varphi + \phi) , \qquad a'_{12} = -\sin(\varphi + \phi) , \qquad a'_{13} = 0 ; \tag{3.23}$$

while

$$a'_{31} = 0\,, \qquad a'_{32} = 0\,, \qquad a'_{33} = 1\,. \tag{3.24}$$

As a result, Equation (3.22) simplifies into

$$V^{(+)} = \{u' \cos(\varphi + \phi) - v' \sin(\varphi + \phi)\} \sin I + w' \cos I\,, \tag{3.25}$$

where u', v', w' are body velocity-components of non-rotational origin, arising from possible motions in the medium in which spectral lines used to measure the Doppler shifts originate.

4. Effects of Non-Rotational Motions

In order to investigate such motions, let us transform first the 'body' velocities

$$\begin{Bmatrix} u' \\ v' \\ w' \end{Bmatrix} = \begin{Bmatrix} b''_{11} & b''_{12} & b''_{13} \\ b''_{21} & b''_{22} & b''_{23} \\ b''_{31} & b''_{32} & b''_{33} \end{Bmatrix} \begin{Bmatrix} u'' \\ v'' \\ w'' \end{Bmatrix}, \tag{4.1}$$

where the direction cosines b''_{ij} are given by Equations (2.10) of Section 2; in doing so we find that

$$a'_{11}u' + a'_{12}v' + a'_{13}w' = a''_{11}u'' + a''_{12}v'' + a''_{13}w'' \tag{4.2}$$

and

$$a'_{31}u' + a'_{32}v' + a'_{33}w' = a''_{31}u'' + a''_{32}v'' + a''_{33}w''\,, \tag{4.3}$$

where the directions cosines a''_{ij} continue to be given by Equations (2.6)–(2.8) of Section 2.

Next, let us change over from the rectangular velocity components u'', v'', w'' to the polar velocity components U, V, W by means of the matrix equation

$$\begin{Bmatrix} u'' \\ v'' \\ w'' \end{Bmatrix} = \begin{Bmatrix} \cos\phi \sin\theta & \cos\phi\cos\theta & -\sin\phi \\ \sin\phi\sin\theta & \sin\phi\cos\theta & \cos\phi \\ \cos\theta & -\sin\theta & 0 \end{Bmatrix} \begin{Bmatrix} U \\ W \\ V \end{Bmatrix}, \tag{4.4}$$

where θ and ϕ are spherical polar coordinates in the revolving system. If so, it follows that

$$\begin{aligned} a''_{11}u'' &+ a''_{12}v'' + a''_{13}w'' = \\ &= U\{\cos\Omega\cos(\phi + u) - \sin\Omega\sin(\phi + u)\cos i]\sin\theta + \sin\Omega\sin i\cos\theta\} + \\ &+ V\{[\cos\Omega\cos(\phi + u) - \sin\Omega\sin(\phi + u)\cos i]\cos\theta - \sin\Omega\sin i\sin\theta\} - \\ &- W\{\cos\Omega\sin(\phi + u) + \sin\Omega\cos(\phi + u)\cos i\} \end{aligned} \tag{4.5}$$

and

$$a''_{31}u'' + a''_{32}v'' + a''_{33}w'' =$$
$$= U\{\sin(\phi + u)\sin i \sin\theta + \cos i \cos\theta\} +$$
$$+ V\{\sin(\phi + u)\sin i \cos\theta - \cos i \sin\theta\} +$$
$$+ W\{\cos(\phi + u)\sin i\}, \tag{4.6}$$

where the angles Ω and i specify the position of the orbital plane in space (cf. Section 2) and u, the true anomaly of the revolving body measured from the nodal passage.

If, moreover, the angle i between the (instantaneous) orbital and invariable plane of the system is small enough to be ignored, the foregoing Equations (2.6) and (2.7) reduce to

$$a''_{11}u'' + a''_{12}v'' + a''_{13}w'' = U\cos(\Omega + u + \phi)\sin\theta +$$
$$+ V\cos(\Omega + u + \phi)\cos\theta - W\sin(\Omega + u + \phi) \tag{4.7}$$

and

$$a''_{31}u'' + a''_{32}v'' + a''_{33}w'' = U\cos\theta - V\sin\theta. \tag{4.8}$$

Within this scheme of our approximation, the non-rotational contribution to non-orbital radial velocity $V^{(+)}$ arising from bodily motion will, in accordance with Equation (3.21), be of the form

$$V^{(+)} = U\{\cos(\Omega + u + \phi)\sin I \sin\theta + \cos I \cos\theta\} +$$
$$+ V\{\cos(\Omega + u + \phi)\sin I \cos\theta - \cos I \sin\theta\} -$$
$$- W\{\sin(\Omega + u + \phi)\sin I\}. \tag{4.9}$$

and its effect δV on the observed radial velocity V should, in accordance with Equation (1.1), be given by

$$\delta V = \frac{\int V^{(+)}\,\mathrm{d}l}{\int \mathrm{d}l}. \tag{4.10}$$

5. Equations of the Problem

In order to establish the explicit form of the non-orbital part $V^{(+)}$ of the radial velocity representing the effects of irradiation of one star by its mate, we must first specify the body velocity-components U, V, and W arising from this cause. The problem at issue has been considered already by the present writer (cf. Kopal, 1980b) in Clairaut coordinates (in which the radial coordinate r is replaced by the equipotential a of the level surfaces of the respective configuration) rotating with an angular velocity ω.

If so, the *linearized* version of the respective equations of motion is represented by Equations (3.19)–(3.21) in the paper referred to above (Kopal, 1980b) implies a tacit assumption that the velocity components U, V, W caused by the heating effects of incident light are small enough for their squares and cross-products to be negligible. What is the magnitude that can be assigned to them to render such a linearization legitimate? Consistent with the applications considered in Section 8, if our unit of length is adopted as 10^{12}–10^{13} cm, that of mass 10^{34} g, and of the time such that the gravitation constant $G = 1$, the unit of velocity becomes 260–300 km s^{-1} – so that if the velocity components U, V, W are less than this limit, their squares and cross-products (as well as the changes P', ρ', ψ' in pressure, density, and gravitational potential caused by such motions) can be ignored – the more legitimately so, the stronger such inequalities become. Conversely, at the top of the atmosphere which peters out in empty space through an 'exosphere', the velocity of escape from the exosphere becomes 370–410 km s^{-1} (which will impose a limit on the tabulation presented in Section 8).

Let us assume, however, in what follows that the actual velocities of atmospheric gas streams are such as to render linearization of the Eulerian equations of hydrodynamics legitimate in the form given by Equations (3.19)–(3.21) of Kopal (1980b) to which the reader is referred for fuller details.

In the present case these equations admit, however, of further simplifications. Thus if the components of our system are separated far enough to be regarded as spheres, it is legitimate to insert in Equations (3.19)–(3.21) of Kopal (1980b) $a \equiv r_1$ so that

$$r_a = 1 \quad \text{and} \quad r_\theta = r_\phi = 0\,. \tag{5.1}$$

Furthermore, if the motions in an atmosphere irradiated from above are limited to a thin layer absorbing all incident light, such an atmosphere can be treated as plane-parallel one; and all terms in the equations of our problem divided by r neglected. Lastly, let the internal structure of the irradiated star be such that its gravity is controlled by the deep interior, and the gravitational perturbations ψ' caused by atmospheric gas stremas can likewise be ignored. If so, however, Equations (3.19)–(3.21) in Kopal (1980b) reduce to

$$\frac{\partial U}{\partial t} - 2\omega W \sin\theta = -\frac{g_0}{\rho_0}\,\rho' - \frac{1}{\rho_0}\frac{\partial P'}{\partial r}\,, \tag{5.2}$$

$$\frac{\partial V}{\partial t} - 2\omega W \cos\theta = 0\,, \tag{5.3}$$

$$\frac{\partial W}{\partial t} + 2\omega(U\sin\theta + V\cos\theta) = 0\,; \tag{5.4}$$

where ρ_0 signifies the unperturbed distribution of density; g_0 that of gravitational acceleration; and the terms factored by 2ω on the left-hand sides represent the effects of the Coriolis force; while t denotes the time.

The foregoing Equations (5.2)–(5.4) contain five dependent variables (i.e., $U, V, W; \rho', P'$); and, therefore, two additional equations are required to render their solution determinate. If the Eulerian equations (5.2)–(5.4) of motion represent a conservation of the momentum, the conservation of mass requires (subject to the above-mentioned approximations) that

$$\frac{\partial \rho'}{\partial t} = - \frac{\partial}{\partial r} (\rho_0 U) ; \tag{5.5}$$

while the perturbations P' in pressure caused by external irradiation can be related with the other four variables by an appeal to the conservation of energy as follows.

Let I denote the intensity of radiation at any surface point of the illuminated star; and ζ, γ, the spherical polar coordinates in the triply-primed system $X'' Y'' Z''$ as defined in Section 2. If so, the time-dependent equation of radiative transfer in a spherically-symmetric medium can be expressed as

$$\frac{1}{c} \frac{\partial I}{\partial t} + \cos\gamma \frac{\partial I}{\partial r} + \frac{\sin^2\gamma}{r} \frac{\partial I}{\partial \cos\gamma} = k\rho_0 (B - I) , \tag{5.6}$$

where $k\rho_0$ denotes the absorption coefficient per unit mass in the respective medium; t, the time; and c is the velocity of light.

Moreover, the condition of radiative equilibrium – requiring that all incident light be re-radiated – discloses that the emissivity (source function)

$$4\pi B = \oint I \sin\gamma \, \mathrm{d}\gamma \, \mathrm{d}\zeta + \pi S \, e^{-\tau \sec\alpha} , \tag{5.7}$$

where the integral on the right-hand side should be extended over the entire sphere, and

$$\pi S \equiv \frac{L_2}{R^2} , \tag{5.8}$$

where L_2 stands for the luminosity of the illuminating source; and R, its distance; moreover, α dneotes the angle of incidence of the illuminating flux S and, as such, is defined by the equation

$$r \cos\alpha \equiv x'' = a''_{11} x + a''_{21} y + a''_{31} z = l_2 x'' + l_0 z'' , \tag{5.9}$$

i.e.,

$$\cos\alpha = \cos\gamma \cos\varepsilon + \sin\gamma \sin\varepsilon \cos\zeta , \tag{5.10}$$

where γ continues to denote the angle of reflection; and ε, the angle between the radius-vector of the binary orbit and the line-of-sight is defined by

$$\cos\varepsilon \equiv l_0 , \qquad \sin\varepsilon \equiv l_2 . \tag{5.11}$$

Lastly, τ stands for the optical depth of the respective atmospheric layer, defined by

$$d\tau = -k\rho_0\, dr\,, \qquad \tau = \int_r^{r_*} k\rho_0\, dr\,; \tag{5.12}$$

where r_* corresponds to the 'top' of the atmosphere, at which $\tau(r_*) = 0$.

Equation (5.6) holds good exactly for the time-dependent and spherically-symmetric case. The first term on its left-hand side should, however, be generally negligible for the binary problem we have in mind, and the third likewise becomes ignorable to a plane-parallel approximation. If, moreover, we change over from r to τ as our independent variable by the first part of (5.12) and abbreviate $\cos\gamma \equiv \mu$, Equation (5.6) will reduce to an ordinary differential equation of the form

$$\mu\frac{dI}{d\tau} = I - B\,, \tag{5.13}$$

where the emissivity B is related with the intensity I by the condition (5.7) of radiative equilibrium, reducing now to

$$B = \int_{-1}^{1} I(\tau, \mu)\, d\mu + \tfrac{1}{4}S\, e^{-\tau\sec\alpha}\,; \tag{5.14}$$

subject to the boundary conditions requiring that, at the top of the atmosphere (i.e., for $\tau = 0$),

$$I(0, \mu) = 0 \tag{5.15}$$

for $-1 \leq \mu \leq 0$ (i.e., over the hemisphere receiving no radiation from outside); while, at the bottom of it,

$$\lim_{\tau\to\infty} e^{-\tau} I(\tau, \mu) = 0\,. \tag{5.16}$$

Next, let us integrate the time-independent equation of radiative transfer with respect to $d\mu$ over the entire sphere: we obtain

$$\frac{1}{r^2}\frac{\partial}{\partial r}(r^2 F) = k\rho_0 S\, e^{-\tau\sec\alpha}\,, \tag{5.17}$$

where

$$\mathrm{F} \equiv \int_{-1}^{+1} \mathrm{I}\mu\, d\mu\,, \tag{5.18}$$

plane-parallel case ($r \to \infty$) reduces to

$$\frac{\partial F}{\partial r} = k\rho_0 S\, e^{-\tau\sec\alpha}\,, \tag{5.19}$$

where the incident flux S continues to be given by Equation (5.8), its angle of incidence α by (5.10), and the optical depth τ by (5.12).

Now on the assumption of local thermodynamic equilibrium, the flux

$$F = \frac{ac}{4\pi} T^4, \tag{5.20}$$

where T defines the local temperature, and $a = 7.56 \times 10^{-15}$ erg cm^{-3} deg^{-4} the Stefan constant. On inserting (5.20) in (5.19), we find that the temperature gradient in an irradiated atmosphere can be expressed as

$$\frac{\partial T}{\partial r} = \frac{k\rho_0(\pi S)}{ac} e^{-\tau \sec \alpha}; \tag{5.21}$$

and if we abbreviate the effective luminosity L_1 of the illuminated star by

$$L_1 = \pi a c r^2 T^4, \tag{5.22}$$

Equation (5.19) combined with (5.8) can be rewritten as

$$\frac{r}{T} \frac{\partial T}{\partial r} = \pi(k\rho_0 r) \left(\frac{r}{R}\right)^2 \frac{L_2}{L_1} e^{-\tau \sec \alpha}, \tag{5.23}$$

where the product $k\rho_0 r$ stands for the radius of the illuminated star expressed in terms of the mean-free-path of its atmospheric gas.

As is well known, the variables of state P, ρ, T in perfect gas are connected by the Boyle's law

$$P = \frac{\Re}{\mu} \rho T, \tag{5.24}$$

where (for neutral hydrogen) $\Re/\mu = 8.314 \times 10^7$ erg deg^{-1}; and of them only the temperature T depends *explicitly* on the time t (through the angle α of incidence for changes caused by irradiation). Moreover, since – by partial differentiation of Equation (5.24) with respect to r – then

$$\frac{\partial P'}{\partial r} = \frac{P_0}{T_0} \frac{\partial T'}{\partial r} = \frac{\pi(k\rho r)P_0}{r} \left(\frac{r}{R}\right)^2 \frac{L_2}{L_1} e^{-\tau \sec \alpha}; \tag{5.25}$$

which by partial differentiation with respect to t discloses that

$$\frac{\partial^2 P'}{\partial r\, \partial t} = \left\{\frac{\pi P_0}{r} (k\rho r) \left(\frac{r}{R}\right)^2 \frac{L_2}{L_1}\right\} \frac{\partial}{\partial t} (e^{-\tau \sec \alpha}). \tag{5.26}$$

Let us now return to the equations of motion (5.2)–(5.4) at the beginning of this section, and avail ourselves for their solution to the Boussinesq approximation (cf. Chandrasekhar, 1961), in accordance with which all three velocity components U, V, W

can be regarded as functions of t only. Moreover, within the scheme of this approximation (and by an appeal to Equation (5.5) of continuity),

$$-\frac{\partial}{\partial t}\left\{g_0\rho' + \frac{\partial P'}{\partial r}\right\} = g_0 \frac{\partial \rho_0}{\partial r} U - \\ - \left\{\pi P_0(k\rho_0)\left(\frac{r}{R}\right)^2 \frac{L_2}{L_1}\right\}\frac{\partial}{\partial t} e^{-\tau \sec\alpha}; \tag{5.27}$$

so that if we differentiate (5.2) partially with respect to the time, the first one of the system of Equations (5.2)–(5.4) becomes non-homogeneous in the velocity components and of the form

$$\frac{\partial^2 U}{\partial t^2} - 2\omega \frac{\partial W}{\partial t} \sin\theta - \frac{g_0}{\rho_0}\frac{\partial \rho_0}{\partial r} U = -G \frac{\partial}{\partial t}\{e^{-\tau\sec\alpha}\}, \tag{5.28}$$

in which we have abbreviated

$$G \equiv \pi k P_0 \left(\frac{r}{R}\right)^2 \frac{L_2}{L_1} = a_*^2 \left\{\frac{\pi}{\gamma}(k\rho_0)\left(\frac{r}{R}\right)^2 \frac{L_2}{L_1}\right\} \text{cm s}^{-2}, \tag{5.29}$$

where γ stands for the ratio of specific heats in the atmosphere, and $\gamma P_0/\rho_0 \equiv a_*^2$ denotes the square of the local velocity of sound.

The foregoing equation (5.28) together with Equations (5.3) and (5.4) represent together a system of fourth order in the velocity components U, V, and W, the coefficients of which can (within the scheme of our approximation) be treated as independent of the time.

6. Solution of the Equations

The first step in the solution of our equations will be to separate the dependent variables in (5.3), (5.4), and (5.28) in the alternative form

$$\left\{\frac{\partial^4}{\partial t^4} + A\frac{\partial^2}{\partial t^2} + B\right\}U = G\left\{\frac{\partial^3}{\partial t^3} + 4\omega^2\cos^2\theta\frac{\partial}{\partial t}\right\}e^{-\tau\sec\alpha}, \tag{6.1}$$

$$\left\{\frac{\partial^4}{\partial t^4} + A\frac{\partial^2}{\partial t^2} + B\right\}V = -G\left\{4\omega^2\sin\theta\cos\theta\frac{\partial}{\partial t}\right\}e^{-\tau\sec\alpha}, \tag{6.2}$$

$$\left\{\frac{\partial^4}{\partial t^4} + A\frac{\partial^2}{\partial t^2} + B\right\}W = -G\left\{2\omega\sin\theta\frac{\partial^2}{\partial t^2}\right\}e^{-\tau\sec\alpha}; \tag{6.3}$$

where we have abbreviated

$$A = (2\omega)^2 + \left(-\frac{g}{\rho}\frac{\partial \rho}{\partial r}\right) \tag{6.4}$$

and

$$B = (2\omega \cos \theta)^2 \left(-\frac{g}{\rho} \frac{\partial \rho}{\partial r} \right), \tag{6.5}$$

in which zero subscripts of g and ρ wil hereafter be dropped (for ρ diminishing outwards, both A and B are positive quantities).

Particular integrals of Equations (6.1)–(6.3) due to the non-homogeneous term on the right-hand side of Equation (5.28) can be expressed in a closed form as

$$\begin{aligned} U(\tau, t) = G\Big\{ & N_1 e^{-\sqrt{\lambda_1} t} \int e^{\sqrt{\lambda_1} t} \left[\frac{\partial^3}{\partial t^3} + 4\omega^2 \cos^2 \theta \frac{\partial}{\partial t} \right] e^{-\tau \sec \alpha} \mathrm{d}t + \\ & + N_2 e^{\sqrt{\lambda_1} t} \int e^{-\sqrt{\lambda_1} t} \left[\frac{\partial^3}{\partial t^3} + 4\omega^2 \cos^2 \theta \frac{\partial}{\partial t} \right] e^{-\tau \sec \alpha} \mathrm{d}t + \\ & + N_3 e^{-\sqrt{\lambda_2} t} \int e^{\sqrt{\lambda_2} t} \left[\frac{\partial^3}{\partial t^3} + 4\omega^2 \cos^2 \theta \frac{\partial}{\partial t} \right] e^{-\tau \sec \alpha} \mathrm{d}t + \\ & + N_4 e^{\sqrt{\lambda_2} t} \int e^{-\sqrt{\lambda_2} t} \left[\frac{\partial^3}{\partial t^3} + 4\omega^2 \cos^2 \theta \frac{\partial}{\partial t} \right] e^{-\tau \sec \alpha} \mathrm{d}t \Big\}, \end{aligned} \tag{6.6}$$

$$\begin{aligned} V(\tau, t) = -2\omega^2 G \sin 2\theta \Big\{ & N_1 e^{-\sqrt{\lambda_1} t} \int e^{\sqrt{\lambda_1} t} \frac{\partial}{\partial t} (e^{-\tau \sec \alpha}) \mathrm{d}t + \\ & + N_2 e^{\sqrt{\lambda_1} t} \int e^{-\sqrt{\lambda_1} t} \frac{\partial}{\partial t} (e^{-\tau \sec \alpha}) \mathrm{d}t + \\ & + N_3 e^{-\sqrt{\lambda_2} t} \int e^{\sqrt{\lambda_2} t} \frac{\partial}{\partial t} (e^{-\tau \sec \alpha}) \mathrm{d}t + \\ & + N_4 e^{\sqrt{\lambda_2} t} \int e^{-\sqrt{\lambda_2} t} \frac{\partial}{\partial t} (e^{-\tau \sec \alpha}) \mathrm{d}t \Big\} \end{aligned} \tag{6.7}$$

and

$$\begin{aligned} W(\tau, t) = -2\omega G \sin \theta \Big\{ & N_1 e^{-\sqrt{\lambda_1} t} \int e^{\sqrt{\lambda_1} t} \frac{\partial}{\partial t} (e^{-\tau \sec \alpha}) \mathrm{d}t + \\ & + N_2 e^{\sqrt{\lambda_1} t} \int e^{-\sqrt{\lambda_1} t} \frac{\partial}{\partial t} (e^{-\tau \sec \alpha}) \mathrm{d}t + \\ & + N_3 e^{-\sqrt{\lambda_2} t} \int e^{\sqrt{\lambda_2} t} \frac{\partial}{\partial t} (e^{-\tau \sec \alpha}) \mathrm{d}t + \\ & + N_4 e^{\sqrt{\lambda_2} t} \int e^{-\sqrt{\lambda_2} t} \frac{\partial}{\partial t} (e^{-\tau \sec \alpha}) \mathrm{d}t \Big\}; \end{aligned} \tag{6.8}$$

where

$$\lambda_{1,2} = \tfrac{1}{2}\{-A \pm \sqrt{A^2 - 4B}\} \tag{6.9}$$

and

$$N_{1,2} = \frac{\mp 1}{2\sqrt{\lambda_1}(\lambda_1 - \lambda_2)}, \tag{6.10}$$

$$N_{3,4} = \frac{\pm 1}{2\sqrt{\lambda_2}(\lambda_1 - \lambda_2)}. \tag{6.11}$$

Both λ_1 and λ_2 are, in general, negative quantities (such that $|\lambda_2| \geq |\lambda_1|$). The reader may also note that their difference

$$\lambda_1 - \lambda_2 = (A^2 - 4B)^{1/2} = [(4\omega^2)^2 - 2(4\omega^2)x\cos 2\theta + x^2]^{1/2}, \tag{6.12}$$

where we have abbreviated

$$x \equiv -\frac{g}{\rho}\frac{\partial\rho}{\partial r} \quad \text{(positive)}. \tag{6.13}$$

Moreover, the reciprocal of (6.12) assumes the form

$$\frac{1}{\lambda_1 - \lambda_2} = \frac{1}{4\omega^2}\sum_{j=0}^{\infty}\left(\frac{x}{4\omega^2}\right)^j P_j(\cos 2\theta), \tag{6.14}$$

in which the series on the right-hand side of the preceding equation converges with sufficient rapidity if $x/4\omega^2$ is a small quantity.

Let us return now to the foregoing Equations (6.6)–(6.8) for U, V, W; and integrate them by parts: we find that

$$\begin{aligned} U(\tau, t) = \frac{G}{2(\lambda_1 - \lambda_2)}\Big\{&(\lambda_1 + 4\omega^2\cos^2\theta)\, e^{-\sqrt{\lambda_1}\,t}\int e^{-\tau\sec\alpha + \sqrt{\lambda_1}\,t}\,\mathrm{d}t + \\ &+ (\lambda_1 + 4\omega^2\cos^2\theta)\, e^{\sqrt{\lambda_1}\,t}\int e^{-\tau\sec\alpha - \sqrt{\lambda_1}\,r}\,\mathrm{d}t - \\ &- (\lambda_2 + 4\omega^2\cos^2\theta)\, e^{-\sqrt{\lambda_2}\,t}\int e^{-\tau\sec\alpha + \sqrt{\lambda_2}\,r}\,\mathrm{d}t - \\ &- (\lambda_2 + 4\omega^2\cos^2\theta)\, e^{\sqrt{\lambda_2}\,t}\int e^{-\tau\sec\alpha - \sqrt{\lambda_2}\,r}\,\mathrm{d}t\Big\}, \end{aligned} \tag{6.15}$$

$$V(\tau, t) = -\frac{\omega^2 G \sin 2\theta}{\lambda_1 - \lambda_2}\left\{e^{-\sqrt{\lambda_1}\,t}\int e^{-\tau\sec\alpha + \sqrt{\lambda_1}\,t}\,\mathrm{d}t + \right.$$

$$+ e^{\sqrt{\lambda_1}\,t}\int e^{-\tau\sec\alpha - \sqrt{\lambda_1}\,t}\,\mathrm{d}t - e^{-\sqrt{\lambda_2}\,t}\int e^{-\tau\sec\alpha + \sqrt{\lambda_2}\,t}\,\mathrm{d}t -$$

$$\left. - e^{\sqrt{\lambda_2}\,t}\int e^{-\tau\sec\alpha - \sqrt{\lambda_2}\,t}\,\mathrm{d}t\right\} \tag{6.16}$$

and

$$W(\tau, t) = -\frac{\omega G \sin\theta}{\lambda_1 - \lambda_2}\left\{-\sqrt{\lambda_1}\,e^{-\sqrt{\lambda_1}\,t}\int e^{-\tau\sec\alpha + \sqrt{\lambda_1}\,t}\,\mathrm{d}t + \right.$$

$$+ \sqrt{\lambda_1}\,e^{\sqrt{\lambda_1}\,t}\int e^{-\tau\sec\alpha - \sqrt{\lambda_1}\,r}\,\mathrm{d}t +$$

$$+ \sqrt{\lambda_2}\,e^{-\sqrt{\lambda_2}\,t}\int e^{-\tau\sec\alpha + \sqrt{\lambda_2}\,t}\,\mathrm{d}t -$$

$$\left. - \sqrt{\lambda_2}\,e^{\sqrt{\lambda_2}\,t}\int e^{-\tau\sec\alpha - \sqrt{\lambda_2}\,t}\,\mathrm{d}t\right\}. \tag{6.17}$$

Moreover, by abbreviating

$$e^{\pm\sqrt{\lambda_j}\,t}\int e^{-\tau\sec\alpha \mp \sqrt{\lambda_j}\,t}\,\mathrm{d}t \equiv \mathfrak{P}_{\pm}^{(i)}, \quad j = 1, 2, \tag{6.18}$$

Equations (6.15)–(6.17) can be rewritten more compactly as

$$U(\tau, t) = \frac{G}{2(\lambda_1 - \lambda_2)}\,\{(\lambda_1 + 4\omega^2\cos^2\theta)(\mathfrak{P}_{\times}^{(1)} + \mathfrak{P}_{-}^{(1)}) -$$

$$- (\lambda_2 + 4\omega^2\cos^2\theta)(\mathfrak{P}_{+}^{(2)} + \mathfrak{P}_{-}^{(2)})\}, \tag{6.19}$$

$$V(\tau, t) = -\frac{\omega^2 G \sin 2\theta}{\lambda_1 - \lambda_2}\,\{(\mathfrak{P}_{+}^{(1)} + \mathfrak{P}_{-}^{(1)}) - (\mathfrak{P}_{+}^{(2)} + \mathfrak{P}_{-}^{(2)})\}, \tag{6.20}$$

$$W(\tau, t) = -\frac{\omega G \sin\theta}{\lambda_1 - \lambda_2}\,\{\sqrt{\lambda_1}\,(\mathfrak{P}_{+}^{(1)} - \mathfrak{P}_{-}^{(1)}) - \sqrt{\lambda_2}\,(\mathfrak{P}_{+}^{(2)} - \mathfrak{P}_{-}^{(2)})\}. \tag{6.21}$$

Should we neglect $\tau\sec\alpha$ in comparison with $\sqrt{\lambda_j}\,t$, it follows from (6.18) that

$$\mathfrak{P}_{\pm}^{(j)} = \frac{\mp 1}{\sqrt{\lambda_j}}, \tag{6.22}$$

so that

$$\mathfrak{P}_+^{(j)} + \mathfrak{P}_-^{(j)} = 0\,, \tag{6.23}$$

rendering by (7.19) and (7.20)

$$U(0, t) = V(0, t) = 0 \tag{6.24}$$

at all times. Moreover,

$$\mathfrak{P}_+^{(j)} - \mathfrak{P}_-^{(j)} = -\frac{2}{\sqrt{\lambda_j}}\,; \tag{6.25}$$

and since both λ_1 and λ_2 are negative, their square-roots $\sqrt{\lambda_j}$'s are imaginary. However,

$$\sqrt{\lambda_1}\,(\mathfrak{P}_+^{(1)} - \mathfrak{P}_-^{(1)}) - \sqrt{\lambda_2}\,(\mathfrak{P}_+^{(2)} - \mathfrak{P}_-^{(2)}) = 0\,, \tag{6.26}$$

rendering [by Equation (7.21)]

$$W(0, t) = 0\,. \tag{6.27}$$

Having expressed the velocity components U, V, W by Equations (6.19)–(6.21) in terms of known functions of the time, let us return to Equation (2.10) for non-orbital radial velocity $V^{(+)}$ arising from the motion of any surface point relative to the star's centre of mass, which on insertion for the respective velocity components as given by Equations (6.19)–(6.22) above assumes the form

$$\begin{aligned}
V^{(+)} = {} & \frac{G}{2(\lambda_1 - \lambda_2)}\,(\mathfrak{P}_+^{(1)} + \mathfrak{P}_-^{(1)})\,\{(\lambda_1 + 4\omega^2)\,\times \\
& \times \cos I \cos\theta'' + \lambda_1 \sin I \sin\theta'' \cos(\Omega + u + \phi)\} - \\
& - \frac{G}{2(\lambda_1 - \lambda_2)}\,(\mathfrak{P}_+^{(2)} + \mathfrak{P}_-^{(2)})\,\{(\lambda_2 + 4\omega^2)\cos I \cos\theta'' + \\
& + \lambda_2 \sin I \sin\theta'' \cos(\Omega + u + \phi)\} + \\
& + \frac{\omega G}{\lambda_1 - \lambda_2}\,\{\sqrt{\lambda_1}\,(\mathfrak{P}_+^{(1)} - \mathfrak{P}_-^{(1)}) - \sqrt{\lambda_2}\,(\mathfrak{P}_+^{(2)} - \mathfrak{P}_-^{(2)})\}\,\times \\
& \times \sin I \sin\theta'' \sin(\Omega + u + \phi)\,,
\end{aligned} \tag{6.28}$$

where (cf. Section 2) the angle I denotes the inclination of the invariable plane of our binary system to the celestial sphere, while Ω and i specify the position of the orbital plane in space, and u is the true anomaly of the revolving body measured from the nodal passage.

Our next task should be to rewrite the periodic functions on the right-hand side of Equation (6.28) in terms of the direction cosines of the triply-primed system of revolving coordinates (introduced likewise in Section 2), the z'''-axis coincides with the line-of-sight. In order to do so, let us assume (to simplify matters) that the orbital plane coincides with the invariable plane of the system – an assumption which renders the angle i of Section 2 to vanish.

If so, and if, moreover, we appeal to Equation (2.11) expressing the relations between the direction cosines λ'', μ'', ν'' and L, M, N of an arbitrary radius-vector in the doubly-primed and triply-primed system of coordinates in the form

$$\left.\begin{aligned} \lambda'' &= L_0 L + l_2 N, \\ \mu'' &= m_0 L + m_1 M + m_2 N, \\ \nu'' &= n_0 L + n_1 M + n_2 N, \end{aligned}\right\} \tag{6.29}$$

in which l_0, m_0, n_0 stands for the direction cosines of the line-of-sight as given by Equations (2.23) it follows that (for $i = 0$), $\Omega + u = \psi$ and, accordingly,

$$\cos(\Omega + u + \phi)\sin I \sin\theta'' = (\lambda'' \cos 2\psi - \mu'' \sin 2\psi)\sin I, \tag{6.30}$$

$$\sin(\Omega + u + \phi)\sin I \sin\theta'' = (\lambda'' \sin 2\psi + \mu'' \cos 2\psi)\sin I, \tag{6.31}$$

while $\cos I \cos\theta'' = \nu'' \cos I$.

By taking advantage of the foregoing relations we find that Equation (6.28) for $V^{(+)}$ can be rewritten as

$$\begin{aligned} V^{(+)} = {} & \frac{G\mathfrak{P}_+^{(1)}}{2(\lambda_1 - \lambda_2)} \{(\lambda_1 \cos 2\psi + 2\sqrt{\lambda_1}\,\omega \sin 2\psi)\,\lambda'' \sin I - \\ & - (\lambda_1 \sin 2\psi - 2\sqrt{\lambda_1}\,\omega \cos 2\psi)\mu'' \sin I + (\lambda_1 + 4\omega^2)\,\nu'' \cos I\} + \\ & + \frac{G\mathfrak{P}_-^{(1)}}{2(\lambda_1 - \lambda_\nu)} \{(\lambda_1 \cos 2\psi - 2\sqrt{\lambda_1}\,\omega \sin 2\psi)\,\lambda'' \sin I - \\ & - (\lambda_1 \sin 2\psi + 2\sqrt{\lambda_1}\,\omega \cos 2\psi)\mu'' \sin I + (\lambda_1 + 4\omega^2)\,\nu'' \cos I\} - \\ & - \frac{G\mathfrak{P}_+^{(2)}}{2(\lambda_1 - \lambda_2)} \{(\lambda_2 \cos 2\psi + 2\sqrt{\lambda_2}\,\omega \sin 2\psi)\,\lambda'' \sin I - \\ & - (\lambda_2 \sin 2\psi - 2\sqrt{\lambda_2}\,\omega \cos 2\psi)\mu'' \sin I + (\lambda_2 + 4\omega^2)\,\nu'' \cos I\} - \\ & - \frac{G\mathfrak{P}_-^{(2)}}{2(\lambda_1 - \lambda_2)} \{(\lambda_2 \cos 2\psi - 2\sqrt{\lambda_2}\,\omega \sin 2\psi)\,\lambda'' \sin I - \\ & - (\lambda_2 \sin 2\psi + 2\sqrt{\lambda_2}\,\omega \cos 2\psi)\mu'' \sin I + (\lambda_2 + 4\omega^2)\,\nu'' \cos I\}, \end{aligned} \tag{6.32}$$

where ψ signifies the phase angle of the revolving star in its orbit; and I, the inclination of the latter to the celestial sphere; while the parameters $\lambda_{1,2}$ depend – via Equations (6.4)–(6.5) and (6.9) – on the physical structure and rotation of the irradiated atmosphere; and G (see Equation (5.29)) on the intensity of its illumination.

Moreover, the expressions (6.18) for $\mathfrak{P}^{(j)}$ in (6.32) can be further simplified if we set

$$\lambda_j = -s_j^2 \quad \text{so that} \quad \sqrt{\lambda_j} = is_j\,, \tag{6.33}$$

where i stands for the imaginary unit. If, accordingly, we set

$$e^{\pm\sqrt{\lambda_j}\,t} = \cos s_1 t \pm i \sin s_i t\,, \tag{6.34}$$

the expressions $\mathfrak{P}_+^{(j)} \pm \mathfrak{P}_-^{(j)}$ in Equations (6.19)–(6.21) can be rewritten in terms of s_j as

$$\mathfrak{P}_+^{(j)} + \mathfrak{P}_-^{(j)} = 2\left\{\cos s_j t \int e^{-\tau\sec\alpha} \cos s_i t\, \mathrm{d}t + \right.$$
$$\left. + \sin s_j t \int e^{-\tau\sec\alpha} \sin s_j t\, \mathrm{d}t\right\} \tag{6.35}$$

and

$$\sqrt{\lambda_j}\,(\mathfrak{P}_+^{(j)} - \mathfrak{P}_-^{(j)}) = -2s_j\left\{\sin s_j t \int e^{-\tau\sec\alpha} \cos s_j t\, \mathrm{d}t - \right.$$
$$\left. - \cos s_j t \int e^{-\tau\sec\alpha} \sin s_j t\, \mathrm{d}t\right\}. \tag{6.36}$$

At the top of the exosphere (optical depth $\tau = 0$) the foregoing expressions verify Equations (6.24) and (6.27) at all times; and if $\tau = \infty$ (a level no longer reached by the radiation of the illuminating star),

$$U(\infty, t) = V(\infty, t) = W(\infty, t) = 0 \tag{6.37}$$

as well.

7. Radial Velocity

In order to establish the contributions δV to the radial velocity arising from atmospheric convection between $0 < \tau < \infty$, we have to evaluate the ratios on the right-hand side of Equation (4.11); both integrals in the numerator and denominator to be taken between the limits

$$\int \equiv \int_{-1}^{1} \int_{0}^{\sqrt{1-M^2}} - \int_{-1}^{1} \int_{0}^{-l_0\sqrt{1-M^2}}, \tag{7.1}$$

defining the geometry of the irradiated crescent of the illuminated spherical star, where the element of integration

$$\mathrm{d}l \equiv J \cos\gamma\, \mathrm{d}\sigma = r^2 S \cos\alpha \cos\gamma \sin\theta''\, \mathrm{d}\theta''\, \mathrm{d}\phi''\,, \tag{7.2}$$

in accordance with Lambert's law, where the incident flux S continues to be given by Equation (5.8).

In order to perform the actual integrations of $V^{(+)}\,\mathrm{d}l$ and $\mathrm{d}l$ between the limits on the right-hand side of (7.1) let us convert the respective integrands from doubly-primed to triply-primed coordinates, in which

$$\left.\begin{aligned} &\cos\alpha \equiv l_0 Z + l_2 N\,, \\ &\cos\gamma \equiv L\,, \\ &\mathrm{d}\sigma = \frac{\mathrm{d}M\,\mathrm{d}N}{L}\ ; \end{aligned}\right\} \tag{7.3}$$

and insert in the expression (6.32) for $V^{(+)}$ from (6.29) in terms of the triply-primed direction cosines L, M, N. Moreover, let us assume, in what follows, that the optical depth τ at which the spectral lines used to establish the observed Doppler shifts is a small quantity (of the order of 0.01) – not an unreasonable assumption – and on the strength of this assumption approximate

$$e^{-\tau\sec\alpha}\cos\alpha = \cos\alpha - \tau + \cdots\,. \tag{7.4}$$

If so, the integration of $V^{(+)}\,\mathrm{d}l$ between the limits laid down by (7.1) becomes straightforward; and the outcome becomes

$$\begin{aligned} \delta V = &\ \frac{2}{3}\,\tau\,\frac{G\cos^2 I}{\lambda_1 - \lambda_2} \sum_{j=1}^{2} (-1)^j (\lambda_j + 4\omega^2)\,\times \\ &\times \left\{\cos s_j t \int (\pi - \varepsilon)\cos s_j t\,\mathrm{d}t + \sin s_j t \int (\pi - \varepsilon)\sin s_j t\,\mathrm{d}t\right\} - \\ &- \frac{2}{3}\,\tau\,\frac{G\sin^2 I}{\lambda_1 - \lambda_2} \sum_{j=1}^{2} (-1)^j s_j^2\,\times \\ &\times \left\{\cos s_j t \int (\pi - \varepsilon)\cos 3\psi\cos s_j t\,\mathrm{d}t\,+\right. \\ &\left. +\,\sin s_j t \int (\pi - \varepsilon)\cos 3\psi\sin s_j t\,\mathrm{d}t\right\} - \\ &- \frac{2}{3}\,\tau\,\frac{G\sin I}{\lambda_1 - \lambda_2} \sum_{j=1}^{2} (-1)^j s_j^2\,\times \\ &\times \left\{\cos s_j t \int l_2\cos 2\psi\cos s_j t\,\mathrm{d}t + \sin s_j t \int l_2\cos 2\psi\sin s_j t\,\mathrm{d}t\right\} - \\ &- \frac{4}{3}\,\tau\,\frac{\omega G\sin^2 I}{\lambda_1 - \lambda_2} \sum_{j=1}^{2} (-1)^j s_j\,\times \\ &\times \left\{\sin s_j t \int (\pi - \varepsilon)\sin 3\psi\cos s_j t\,\mathrm{d}t\,-\right. \end{aligned}$$

$$- \cos s_j t \int (\pi - \varepsilon) \sin 3\psi \sin s_j t \, dt \Big\} -$$

$$- \frac{4}{3} \frac{\omega \tau G \sin I}{\lambda_1 - \lambda_2} \sum_{j=1}^{2} (-1)^j s_j \times$$

$$\times \left\{ \sin s_j t \int l_2 \sin 2\psi \cos s_j t \, dt - \cos s_j t \int l_2 \sin 2\psi \sin s_j t \, dt \right\}, \quad (7.5)$$

where $l_2 \equiv \sin \varepsilon$.

For constant ρ_0 (when, by Equations (6.4) and (6.5), $A = 4\omega^2$ and $B = 0$)

$$\begin{aligned} &\lambda_1 = 0, \qquad \lambda_2 = -4\omega^2; \\ &\lambda_1 + 4\omega^2 = 4\omega^2, \qquad \lambda_2 + 4\omega^2 = 0; \end{aligned} \quad (7.6)$$

and

$$s_1 = 0, \qquad s_2 = 2\omega. \quad (7.7)$$

If so, Equation (7.5) reduces to

$$\begin{aligned} \delta V = & -\frac{2}{3} \tau G \cos^2 I \int (\pi - \varepsilon) \, dt - \\ & -\frac{4}{3} \tau G \sin^2 I \left\{ \cos 2\omega t \int (\pi - \varepsilon) \cos(3\psi + 2\omega t) \, dt + \right. \\ & \left. + \sin 2\omega t \int (\pi - \varepsilon) \sin(3\psi + 2\omega t) \, dt \right\} - \\ & -\frac{4}{3} \tau G \sin I \left\{ \sin 2\omega t \int l_2 \sin(2\psi + 2\omega t) \, dt + \right. \\ & \left. + \cos 2\omega t \int l_2 \cos(2\psi + 2\omega t) \, dt \right\}. \end{aligned} \quad (7.8)$$

If, lastly, our binary system happens to be an eclipsing variable (so that it is legitimate to set $\sin^2 I \approx 1$) whose components describe circular orbits (such that the phase angle $\psi = nt$ and $\varepsilon = \psi$, $l_2 = \sin \psi$), Equation (7.8) reduces further to

$$\begin{aligned} \delta V = & -\frac{2}{3} \tau \frac{G}{n} \left\{ \cos 2\kappa\psi \int (\pi - \psi) \cos(3 + 2\kappa)\psi \, d\psi + \right. \\ & + \sin 2\kappa\psi \int (\pi - \psi) \sin(3 + 2\kappa)\psi \, d\psi + \\ & + \sin 2\kappa\psi \int \sin(2 - 2\kappa)\psi \sin \psi \, d\psi + \\ & \left. + \cos 2\kappa\psi \int \cos(2 + 2\kappa)\psi \sin \psi \, d\psi \right\}, \end{aligned} \quad (7.9)$$

where we have abbreviated

$$\pm\kappa \equiv \frac{\omega}{n} ; \tag{7.10}$$

the algebraic sign of the left-hand side depends on whether the axial rotation is direct or retrograde.

The remaining integrals on the right-hand side of Equation (7.9) can be easily evaluated; and the outcome discloses that

$$\delta V = -\frac{\tau G}{3n}\left\{\frac{\cos\psi}{1+2\kappa} + \frac{2(\pi-\psi)}{3+2\kappa}\sin 3\psi - \frac{5+2\kappa}{(3+2\kappa)^2}\cos 3\psi\right\}, \tag{7.11}$$

representing the extent to which the radial-velocity curves of close binaries describing circular orbits should deviate from a simple sinusoid as a result of mutual irradiation of their components.

8. Discussion of the Results

In the preceding sections of this paper we have investigated the theoretical radial-velocity changes δV of the components of close binary systems, of non-orbital origin, which arise from gas motions (convection) in their atmospheres due to the irradiation of each star by its mate; and whose Doppler shifts superpose upon those caused by orbital motion. The magnitude of these effects will depend on the amplitudes of such non-orbital changes; and the aim of this concluding section of this paper will be to estimate this magnitude.

In the simplest form represented by Equation (7.11), its amplitude will be of the order of

$$\frac{\tau G}{3n} \simeq \frac{\pi}{3}\frac{\tau a_*^2}{r\gamma n}(k\rho_0 r)\left(\frac{r}{R}\right)^2\frac{L_2}{L_1} \tag{8.1}$$

byn (5.29), where the optical depth τ at which the spectral lines of greatest interest originate will be adopted as 0.01. Moreover, for a typical close binary the average orbital period is of the order of a few days; hence, its mean daily motion n can be estimated to $10^{-4}\ \mathrm{s}^{-1}$. The average radius of the semi-transparent fringe of the illuminated star may be set to be equal to 10^{11} cm; and as the size of the relative orbit R of its mate is likely to be between 10^{11}–10^{13} cm, we shall not go too far wrong by estimating

$$\frac{r}{R} \simeq \frac{1}{3}, \tag{8.2}$$

and the ratio of the luminosities of the two stars, to

$$\frac{L_2}{L_1} \simeq 1 . \tag{8.3}$$

Lastly, the sound-speed a_* in the semi-transparent fringe of the illuminated star can be approximated by 10 km s^{-1} for a ratio γ of specific heats of $\simeq 1.5$.

If so, it follows from Equation (8.1) that, for the above-adopted values of different parameters,

$$\frac{\tau G}{3n} \simeq 10^{-3} (k\rho_0 r) \text{ km s}^{-1} . \tag{8.4}$$

As was already mentioned before, the nondimensional product $k\rho_0 r$ is identical with the radius of the semi-transparent outer fringe of the respective star expressed in terms of the mean-free-path of the constituent gas. Accordingly, we must expect that

$$k\rho_0 r \gg 1 \tag{8.5}$$

if hydrodynamical treatment is to be applicable; for if the converse were true, we should find ourselves in the domain of particle mechanics. However, for

$k\rho_0 r$	$\tau G/3n$
10^3	1 km s^{-1}
10^4	10 km s^{-1}
10^5	100 km s^{-1}

etc.

For $k\rho_0 r \gg 10^5$ our tabulation can no longer be extended very far, as the radial velocity $\tau G/3n \simeq 300 \text{ km s}^{-1}$ would then exceed the velocity of escape from the gravitational field of the star, and gas moving with such speed would no longer dynamically belong to the respective system. On the other hand, the Doppler shifts corresponding to the velocity appropriate for $k\rho_0 r = 10^3$ would again be scarcely measurable with the means presently at our disposal. However, for $k\rho_0 r > 10^4$ *the respective velocity would become comparable with that of orbital motion of a mass-point*; and should we insist on interpreting the observations simply on this latter basis, the results so obtained would be devoid of much sense. This, in particular, is very probably the reason why the systems like U Cephei or RZ Scuti exhibit radial-velocity curves so conspicuously at variance with those that would be produced by the Keplerian motion; and the same may be true (albeit to a lesser extent) of many other systems as well.

Accordingly, the main message of the present investigation is to point out that conditions may easily exist in the atmospheres of the components of many close binaries in which a reduction of their observed radial velocities to the mass centres of their components becomes a matter of some complexity and the two cannot be automatically identified. The velocities furnished directly by the observations can be related with the orbital motion of the components only through the 'filter' of the outer semi-transparent fringe in which the spectral lines whose Doppler shifts we can measure originate. It is only if gas in this fringe is at rest with respect to the centre of mass of the star as a whole that the observed radial velocity can be identified without any ado with those arising from the Keplerian orbital motion. In the present paper we have, however, shown that the 'reflection effect' in close binary systems – making their components irradiate each other from close proximity – makes this unlikely, if not impossible, to be the case. If we do not heed this message, its neglect may cause the results obtained by short-cutting the process of velocity transfer between observations and the centre of mass to be systematically in error – not only as far as the shape of the orbit is concerned, but also its size – and, above all, the absolute masses of the constituent stars.

References

Carpenter, F. M.: 1930, *Astrophys. J.* **72**, 105.
Chandrasekhar, S.: 1961, *Hydrodynamic and Hydromagnetic Stability*, Oxford Univ. Press, Chapter II, Section 8.
Curtis, H. D.: 1908, *Publ. Astron. Soc. Pacific* **20**, 133.
King, W. F.: 1908, *Astrophys. J.* **27**, 125.
Kirbiyik, H. and Smith, R. C.: 1976, *Monthly Notices Roy. Astron. Soc.* **176**, 103.
Kitamura, M. and Kopal, Z.: 1968, in Z. Kopal (ed.), *Adv. in Astron. Astrophys.*, Academic Press, New York, Vol. 6, pp. 125–172.
Kopal, Z.: 1942, *Proc. Am. Phil. Soc.* **85**, 299.
Kopal, Z.: 1943, *Proc. Am. Phil. Soc.* **86**, 351.
Kopal, Z.: 1945, *Proc. Am. Phil. Soc.* **89**, 517.
Kopal, Z.: 1959, *Close Binary Systems*, Chapman-Hall and John Wiley, London and New York, Chapter V.
Kopal, Z.: 1978, *Dynamics of Close Binary Systems*, D. Reidel Publ. Co., Dordrecht, Holland.
Kopal, Z.: 1979, *Language of the Stars*, D. Reidel Publ. Co., Dordrecht, Holland, Chapter II.
Kopal, Z.: 1980a, *Astrophys. Space Sci.* **70**, 329.
Kopal, Z.: 1980b, *Astrophys. Space Sci.* **70**, 407.
Kopal, Z.: 1980c, *Astrophys. Space Sci.* **71**, 65.
Kopal, Z.: 1981, in E. B. Carling and Z. Kopal (eds.), *Photometric and Spectroscopic Binary Systems*, NATO ASI, Vol. 69, D. Reidel Publ. Co., Dordrecht, Holland, pp. 535–558.
Lehmann-Filhés, R.: 1894, *Astron. Nachr.* **136**, 17.
Neubauer, F. J. and Struve, O.: 1945, *Astrophys. J.* **101**, 240.
Nijland, A. N.: 1903, *Astron. Nachr.* **161**, 103.
Plummer, H. C.: 1908, *Astrophys. J.* **28**, 212.
Schwarzschild, K.: 1900, *Astron. Nachr.* **152**, 65.
Sterne, T. E.: 1941, *Proc. U.S. Nat. Acad. Sci.* **27**, 168.
Struve, O.: 1944, *Astrophys. J.* **99**, 222.
Struve, O.: 1949, *Monthly Notices Roy. Astron. Soc.* **109**, 487.
Vogel, H. C.: 1890, *Astron. Nachr.* **123**, 287.
Wilsing, J.: 1893, *Astron. Nachr.* **134**, 89.
Zurhellen, W.: 1907, *Astron. Nachr.* **175**, 245.
Zurhellen, W.: 1908, *Astron. Nachr.* **177**, 321.

MAGNETIC CONFINEMENT OF COSMIC CLOUDS*

MICHEL AZAR

Department of Electrical Engineering and Computer Sciences, University of California, San Diego, La Jolla, U.S.A.

and

W. B. THOMPSON

Department of Physics, University of California, San Diego, La Jolla, U.S.A.

(Received 27 July, 1987)

Abstract. The role of the magnetic field in the confinement or compression of interstellar gas clouds is reconsidered. The virial theorem for an isolated magnetized cloud in the presence of distant magnetic sources is reformulated in terms of moments of the internal and external currents, and an equilibrium condition is derived. This condition is applied to the interaction between isolated clouds for the simple- and artificial-case in which the field of each cloud is a dipole. With the simplest of statistical assumptions, the probability of any given cloud being compressed is calculated as $\sim 10\%$, the magnetic field acting as a medium which transmits the kinetic pressure between clouds. Even when compression occurs the magnetic pressure $\frac{1}{2}B^2$ may decrease on leaving the cloud surface.

1. Introduction

Do magnetic fields inhibit or enhance the contraction of a cosmic cloud? It has generally been agreed that gravitational collapse of such a cloud is initiated by an increase in the intercloud pressure. Such an increase can be induced by various triggering mechanisms (cf. Herbst and Assousa, 1978; Larson, 1978) (supernova explosions, density waves, cloud collisions, etc.). The role of the magnetic field in this process is to eventually halt the contraction and stabilize the cloud against gravitational collapse (cf. Muschovias, 1978).

The general belief is, therefore, that magnetic fields always obstruct the contraction process of cosmic clouds. This conclusion, based on the virial theorem (see Chandrasekhar and Fermi, 1953; Parker, 1958; Mestel, 1972) is valid if it is assumed that no external currents are present. On the other hand, magnetized clouds should interact through their magnetic fields, and on collision, the momentum transfer transmitted through the fields should compress the clouds. This may arise if the magnetic pressure $\frac{1}{2}B^2$ is greater outside than inside the clouds; but this, as we shall show, is not an essential condition.

To study this process we first reformulate the virial theorem for an isolated cloud interacting with others, but we assume that currents are confined entirely to the separate clouds, the field between them being an harmonic vacuum field. This assumption permits an equilibrium condition to be written in terms of moments of the internal and

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 587–614.

external current distributions alone. Indeed, the external moments can be written in terms of the Taylor expansion coefficients of the external field about the center of the system.

This equilibrium condition is then applied to an assembly of clouds each carrying a simple dipole field, and using the simplest statistical assumption, a random distribution of clouds, the probability that a given cloud will be compressed is shown to be that fraction of the volume occupied by the clouds, and can easily be of order 10%. Even in this case, however, the magnetic pressure $\frac{1}{2}B^2$ decreases on leaving the cloud surface.

The model considered does not pretend to be realistic, and indeed, although the virial condition gives a necessary condition for equilibrium, it is very far from sufficient and there is no claim that the conclusions are other than suggestive.

For clarity, the paper begins with an outline of the arguments, and a summary of the results, followed by the details of the calculation.

2. Summary

2.1. The virial theorem

We begin with a reformulation of the virial theorem for an isolated magnetized cloud in the presence of a distant current.

This is usually written

$$\int_V \left(3P + \frac{B^2}{8\pi}\right) \mathrm{d}^3 r = \int_S \mathbf{r} \left[\left(P + \frac{B^2}{8\pi}\right)\mathsf{I} - \frac{\mathbf{BB}}{4\pi}\right] \mathrm{d}\mathbf{S}\,, \tag{1}$$

where the volume integral extends over a volume including the entire cloud, and the surface integral is over the boundary of that volume. This result is obtained by forming the moment of $\mathbf{r}$, the vector from an arbitrary origin with the equation for state equilibrium

$$\int_V \mathbf{r} \cdot \left[\nabla\left(P + \frac{B^2}{8\pi}\right) - \frac{1}{4\pi}\,(\mathbf{B}\cdot\nabla)\mathbf{B}\right] \mathrm{d}^3 r = 0\,. \tag{2}$$

If the current distribution is localized, i.e., if all currents lie inside some small volume V_1, then the volume integral can be extended to some much larger volume V_2 and over the surface of V_2, the magnetic field becomes at least as small as that of a dipole and $B^2 \sim 1/r^6$, so that the surface integral in (1) vanishes, hence so must the positive definite volume integral; which implies that an isolated cloud cannot be confined by its own magnetic field.

If there are external currents, however, the surface integral depends on the distribution and value of these, and fields can transmit external pressures which confine a plasma. In general, the nature of the confinement depends on the details of the current and not much can be said. However, if the confined volume is in a vacuum so that no currents close the boundary surface, and if the volume is simply connected so that the surface is spherical some progress can be made.

In this case, the magnetic fields on the surface S can be derived from a harmonic potential, and that potential can be represented by a series of spherical harmonics; then the potential due to internal sources in S is

$$\phi_i = \sum_{l,m} \frac{q_{lm}}{r^{l+1}} \, y_{lm}(\theta, \varphi)\,; \tag{3}$$

while that due to external sources is

$$\phi_e = \sum_{l,m} \alpha_{lm} r^l y_{lm}(\theta, \varphi) \tag{4}$$

and, not surprisingly, the surface integral can be written in terms of the coefficients α_{lm} and q_{lm}. As we will show, the virial theorem may be written

$$\int\limits_{\text{all space}} \left(3P + \frac{B_i^2}{8\pi}\right) \mathrm{d}^3 r = \sum_{l=1}^{\infty} \sum_{m=-l}^{+l} (2l+1)(l+1)\,\mathrm{Re}(\alpha_{lm} q_{lm}^*)\,, \tag{5}$$

where $\mathbf{B}_i$ represents the field due to internal sources alone.

As is shown in Appendix A, the coefficients q_{lm} are related to the multipole moments of the current density $\mathbf{J}_i$ within the cloud, while the α_{lm} are the coefficients of the Taylor series of the external field expanded about the center of the cloud.

This scheme has been used to determine equilibrium conditions for plasmas confined by image fields in simply connected ideally conducting rigid vessels, which might be considered as a rough model of an isolated interstellar cloud surrounded by an ionized medium which is held in place by its own field, and ultimately by gravitation.

However, there may be more interest in a simple model which demonstrates how isolated clouds may interact statistically.

2.2. Magnetic interaction in an assembly of clouds

As an indication of how magnetic interaction may compress clouds we consider an extremely simple system, an assembly of magnetized clouds moving through a vacuum but interacting through their magnetic fields. This is, surely, rather artificial since an assembly of such clouds would certainly be moving through a diffuse background plasma, and currents might flow both within and between the clouds; however, the internal currents may be much larger than the external ones, and the vacuum representation is not too unrealistic.

The next simplifying assumption clearly cannot be justified, since it is inconsistent with the internal integrity of any simple cloud, that is that the internal field be uniform and the external field a dipole, which is, of course, not an internal magnetohydrodynamic equilibrium; however, we can still inquire under what conditions the virial would permit an equilibrium; and the problem is simple enough to analyze. The genuine problem is a great deal more formidable, both to formulate and to analyze. The present attempt is one small step.

If the internal current distribution is completely characterized by a dipole moment **m**, the virial equation reduces to

$$\int \left(3P + \frac{B_i^2}{8\pi}\right) \mathrm{d}^3 r = -2\mathbf{m} \cdot \mathbf{B}_e(0)\,, \tag{6}$$

where $\mathbf{B}_e$ is the external magnetic field. A similar expression was derived by Sweet (1958) for a nearly-uniform external field.

This statement indicates that a positive pressure is possibe (with the caveats given above) provided that the energy in the external field $E = -\mathbf{m} \cdot \mathbf{B}_e(0)$ is greater than half the internal magnetic energy – i.e.,

$$E > \int \frac{B_i^2}{16\pi}\, \mathrm{d}^3 r \equiv M_i\,; \tag{7}$$

and to get the probability of confinement in some distribution of clouds we ask for the probability that $E > M_i$.

To be specific, all clouds are considered as equivalent, i.e., they all have the same

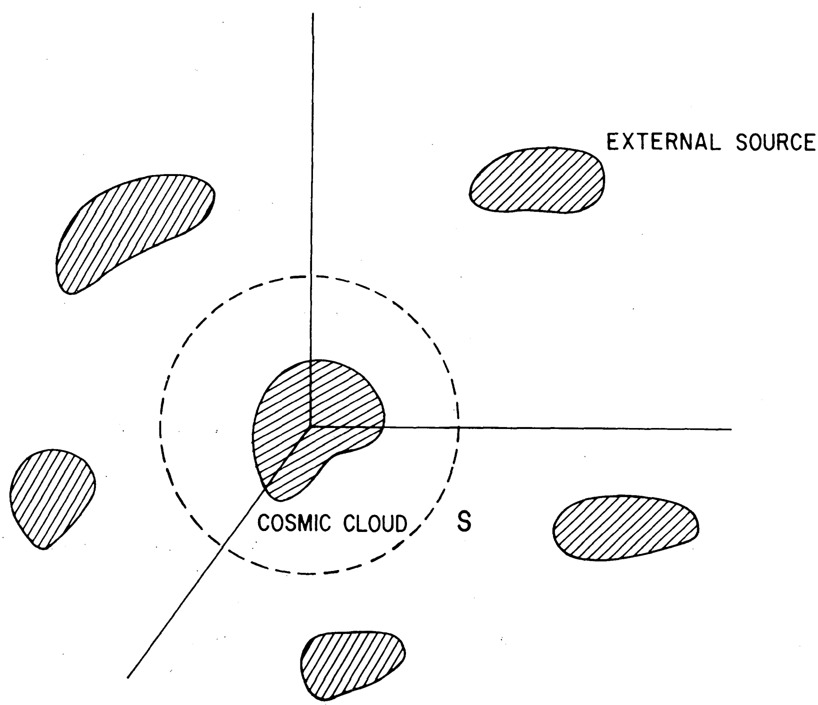


Fig. 1. An isolated cosmic cloud in the presence of magnetic fields produced by internal and external sources. S is the surface of a spherical region of integration used in deriving the equilibrium condition (27).

dipole moment, but are ordered almost at random; if there is an average field $\mathbf{B}_0$ it is much less than the internal field $\mathbf{B}_i$.

To evaluate $P_r(E > M_i)$ we form the probability $W(E)\,\mathrm{d}E$ that a chosen cloud will have energy between E and $E + \mathrm{d}E$, given a probability distribution for the assembly of the clouds – i.e.,

$$W(E) = \int \mathrm{d}^N X\, P(X_1, X_2, \dots, X_N)\, \delta[E_1(X_1, X_2, \dots, X_N) - E]\,, \tag{8}$$

where $P(X_1, X_2, \dots, X_N)$ is the probability of finding the assembly of N clouds in a configuration characterized by dynamical coordinates $(X_1, \dots, X_N)$. For illustrative purposes, we assume that $P(X_1, X_2, \dots, X_N)$ is given by the Gibbs distribution

$$P(X_1, \dots, X_N) \alpha\, e^{-\beta \mathscr{E}_t(X_1, X_2, \dots, X_N)}\,, \tag{9}$$

where $\mathscr{E}_t = \Sigma_{i=1}^N E_i(X_1, \dots, X_N)$ is the energy of the entire configuration of clouds and $E_1(X_1, \dots, X_N)$ is the energy of the chosen cloud in that configuration – i.e.,

$$E_i(X_1, \dots, X_N) = \mathscr{E}_0(X_i) + \sum_{j \neq i}^{N} \mathscr{E}(X_i, X_j)\,, \quad i = 1, 2, \dots, N\,, \tag{10}$$

$$\begin{aligned} \mathscr{E}_t(X_1, \dots, X_N) &= \sum_{i=1}^{N} \mathscr{E}_0(X_i) + \sum_{i=1}^{N} \sum_{j>i} \mathscr{E}(X_i, X_j) = \\ &= \mathscr{E}_0(X_1) + \sum_{j>1}^{N} \mathscr{E}(X_1, X_j) + \sum_{i=2}^{N} \mathscr{E}_0(X_i) + \\ &\quad + \sum_{i=2}^{N} \sum_{j>i} \mathscr{E}(X_i, X_j)\,; \end{aligned} \tag{11}$$

where $\mathscr{E}_0(X_i)$ is the energy of the ith dipole in any average field $\mathbf{B}_0$ and $\mathscr{E}(X_i, X_j)$ is the interaction energy between dipoles $\mathbf{m}_i$ and $\mathbf{m}_j$.

Assuming that the clouds are weakly correlated, we can neglect the term $\Sigma_{i=2}^N \Sigma_{j>i} \mathscr{E}(X_i, X_j)$ to obtain an expression for $W(E)$ in the form

$$W(E) = \frac{a/\pi}{E^2 + a^2} \exp\left\{-\beta\left[E + \frac{2a}{\pi} F(\beta\mathscr{E}_{\max})\right]\right\} \left\{1 - \frac{1}{3}\frac{E_0^2}{a^2}\frac{1 - 3E^2/a^2}{(1 + E^2/a^2)^2} - \Delta_1\right\}, \tag{12}$$

where

$$F(\beta\mathscr{E}_{\max}) \equiv \mathrm{Shi}(\beta\mathscr{E}_{\max}) - \frac{\cosh\beta\mathscr{E}_{\max}}{\beta\mathscr{E}_{\max}} = \int_0^{\beta\mathscr{E}_{\max}} \frac{\sinh x}{x}\,\mathrm{d}x - \frac{\cosh\beta\mathscr{E}_{\max}}{\beta\mathscr{E}_{\max}}\,, \tag{13}$$

$$\Delta_1 \equiv \frac{\beta^2 E_0^2}{6}\left\{1 - \frac{8}{15}\frac{E_0^2}{a^2}\left[\frac{1 - \frac{5}{2}E^2/a^2 - \frac{13}{4}E^4/a^4}{(1 + E^2/a^2)^3} + \frac{\beta a}{8}\frac{1 - 3E^2/a^2}{(1 + E^2/a^2)^2}\right]\right\}. \tag{14}$$

In the above expressions $a \equiv (2\pi/9\sqrt{3})nm^2$ (n is the number of clouds per unit volume) and $\mathscr{E}_{\max}$ denote the average and maximum dipole-dipole interaction energy, respectively, $\beta\mathscr{E}_{\max}$ and βa being finite, while $E_0 = mB_0$ is the (negative) energy of the dipole in any average field B_0, and is assumed to be small in comparison with a and β^{-1}.

If βE_0, E_0/a, and βa can be neglected, then $W(E)$ will reduce to a simple form

$$W(E) \approx \frac{1/\pi}{E^2 + a^2} . \tag{15}$$

Using this we obtain

$$P_r(E > M_i) \sim nr_c^3 , \tag{16}$$

where r_c is the cloud radius.

Having thus calculated the probability of occurrence of equilibrium, we now turn to the question of how the total magnetic field changes with distance in the region between a confined cloud and its nearest neighbours. In this region the magnetic field is a vacuum field. Its radial variation will be calculated as

$$B^2(r) = \frac{1}{4\pi} \int |B_e + B_i|^2 \, d\Omega = B_e^2(0) + \frac{2m^2}{r^6} + \\ + \sum_{l=2} \sum_{m=-l}^{+l} \frac{l(2l+1)}{4\pi} |\alpha_{lm}|^2 r^{2(l-1)} , \tag{17}$$

with

$$|\alpha_{lm}|^2 r^{2(l-1)} < \left(\frac{r}{D}\right)^{2(l+2)} \frac{m^2}{r^6} , \quad l = 2, 3, \ldots ; \tag{18}$$

where D is the average inter-cloud distance.

It is evident from (17) and (18) that in the vicinity of a magnetically confined cloud the magnitude of the total magnetic field decreases outward to a minimum, and then starts increasing as the nearest neighbours are approached.

3. General Conditions for Equilibrium when Internal and External Sources are Present

The condition for equilibrium is given by the virial equation

$$\int_V \left(3P + \frac{B^2}{8\pi}\right) d^3r = \int_S \mathbf{r} \cdot \left[\mathbf{I} \frac{B^2}{8\pi} - \frac{\mathbf{BB}}{4\pi}\right] \cdot d\mathbf{S} , \tag{19}$$

where we have assumed that the external pressure is negligible; i.e., $p = 0$ outside of the cloud. The integration in (19) is over the volume V and surface S of a sphere of radius R, containing the whole cloud, and such that no currents flow on or across the surface

S. This implies that on S the fields sare vacuum fields and, hence, may be derived from a harmonic potential.

Following a procedure used by Gauss (see Chapman and Bartels, 1962) in his analysis of the terrestrial magnetism, we now derive an expression for the virial theorem in the general case where currents of internal and external origin are present.

Let $\mathbf{B}_i$ denote the magnetic field due to currents in the cloud and $\mathbf{B}_e$ that which derives from external sources. Thus

$$\mathbf{B} = \mathbf{B}_e + \mathbf{B}_i , \tag{20}$$

where $\mathbf{B}_e$ and $\mathbf{B}_i$ both being vacuum fields derivable from harmonic potentials and may be written

$$\mathbf{B}_e = -\nabla\phi_e ; \qquad \phi_e = \sum_{l,m} \alpha_{lm} Y_{lm}(\theta, \phi) r^l , \tag{21}$$

for $r \leq R$, since the sources of $\mathbf{B}_e$ are outside R.

$$\mathbf{B}_i = \nabla\phi_i ; \qquad \phi_i = \sum_{l,m} \frac{q_{lm}}{r^{l+1}} Y_{lm}(\theta, \phi) \tag{22}$$

for the region outside the cloud, then r is the radial distance from the origin; $Y_{lm}(\theta, \phi)$ the spherical harmonic of order l, m and the coefficients α_{lm} and q_{lm} are constants determined by the current distribution in the external and internal sources, respectively.

Now, using Equations (20, (21), and (22), the surface term

$$K \equiv \oint \mathrm{d}\mathbf{s} \cdot \left[\frac{\mathbf{BB}}{4\pi} - \mathsf{I}\,\frac{B^2}{8\pi}\right]$$

is

$$K = \frac{1}{8\pi} \sum_{l,m} \left\{(l+1)\frac{|q_{lm}|^2}{R^{2l+1}} - l\,|\alpha_{lm}|^2 R^{2l+1} - 4l(l+1)\,\mathrm{Re}(\alpha_{lm} q_{lm}^*)\right\}; \tag{23}$$

and the magnetic energy $\eta \equiv \int_{r<R} (B^2/8\pi)\,\mathrm{d}^3 r$ is

$$\eta = \frac{1}{8\pi} \int\limits_{r<R} B_i^2\,\mathrm{d}^3 r + \frac{1}{8\pi} \sum_{l,m} \{l\,|\alpha_{lm}|^2 R^{2l+1} - 2(l+1)\,\mathrm{Re}(\alpha_{lm} q_{lm}^*)\} . \tag{24}$$

With the help of (23) and (24), the condition for equilibrium (19) can be expressed as

$$\int\limits_{r<R} \left(3P + \frac{B_i^2}{8\pi}\right)\mathrm{d}^3 r = -\frac{1}{8\pi} \sum_{l,m} (l+1)\frac{|q_{lm}|^2}{R^{2l+1}} + \\ + \frac{1}{4\pi} \sum_{l,m} (2l+1)(l+1)\,\mathrm{Re}(\alpha_{lm} q_{lm}^*) . \tag{25}$$

On the other hand, since $\mathbf{B}_i = -\nabla\phi_i$ and $p = 0$ outside of the cloud, then

$$\int_{r>R} \left(3P + \frac{B_i^2}{8\pi}\right) \mathrm{d}^3 r = \frac{1}{8\pi} \sum_{l,m} (l+1) \frac{|q_{lm}|^2}{R^{2l+1}} . \tag{26}$$

When we combine Equation (26) with (26), the condition for equilibrium becomes

$$3 \int P \,\mathrm{d}^3 r = - \int \frac{B_i^2}{8\pi} \mathrm{d}^3 r + \frac{1}{4\pi} \sum_{l,m} (2l+1)(l+1) \operatorname{Re}(\alpha_{lm} q_{lm}^*) , \tag{27}$$

a result independent of the size of the spherical region of integration provided all the cloud is inside.

The coefficients α_{lm} and q_{lm} are, respectively, given (cf Appendix A) by

$$-l\alpha_{lm} = \sqrt{\frac{2l-1}{2l+1}} \left\{ \sqrt{\frac{l-m}{l-m-1}}\, J^+_{l-1,m+1} - \sqrt{\frac{l+m}{l+m-1}}\, J^-_{l-1,m-1} \right\} , \tag{28}$$

$$(l+1) q_{lm} = \sqrt{\frac{2l+3}{2l+1}} \left\{ \sqrt{\frac{l-m+1}{l-m+2}}\, M^-_{l+1,m-1} - \sqrt{\frac{l+m+1}{l+m+2}}\, M^+_{l+1,m+1} \right\} , \tag{29}$$

$|m| \le l; \quad l = 1, 2, 3, \ldots ;$

where

$$M^{\pm}_{lm} \equiv \frac{\hat{x} \pm i\hat{y}}{2} \cdot \mathbf{M}_{lm} , \tag{30}$$

etc., and

$$\mathbf{M}_{lm} \equiv \frac{4\pi}{c(2l+1)} \int r^l Y^*_{lm}(\theta, \phi) \nabla \times \mathbf{J}_i(\mathbf{r}) \,\mathrm{d}^3 r , \tag{31}$$

$$\mathbf{J}_{lm} = \frac{4\pi}{c(2l+1)} \int \frac{Y^*_{lm}(\theta, \phi)}{r^{l+1}} \nabla \times \mathbf{J}_e(\mathbf{r}) \,\mathrm{d}^3 r ; \tag{32}$$

$\mathbf{J}_i$ and $\mathbf{J}_e$ being the internal (within the cloud) and external current densities, respectively.

As shown in Appendix A, $\mathbf{M}_{lm}$ and, therefore, q_{lm} are linear combinations of magnetic multipole moments due to the internal currents distribution $\mathbf{J}_i$; whereas $\mathbf{J}_{lm}$ and, therefore, α_{lm} are linear combinations of the derivatives of the components of $\mathbf{B}_e$ evaluated at the origin.

After substitution from (A16) and (A18) the virial Equation (27) reduces to a familiar form

$$\int \left(3P + \frac{B_i^2}{8\pi}\right) \mathrm{d}^3 r = 2\mathbf{m} \cdot \mathbf{B}_e(0) - 6\mathbf{m} \colon \nabla\mathbf{B}_e(0) + \cdots ; \tag{27'}$$

where $\mathbf{m}$ and $\boldsymbol{\mathsf{m}}$ are, respectively, the magnetic dipole and quadrupole associated with the internal distribution $\mathbf{J}_i$. In Equation (27′) the series on the right-hand side includes products of all higher moments and higher-order derivatives of $\mathbf{B}_e$.

We note that, according to Equation (27′) the sum of twice the thermal kinetic energy and the magnetic energy associated with the internal field is given in terms of the value of the external field and its derivatives at the center of the cloud.

4. Random Distribution of Isolated Cosmic Clouds

Let us consider a system consisting of a number of isolated cosmic clouds of average density n and average separation D. We assume that each cloud has a permanent magnetic dipole moment $\mathbf{m}$, and for illustrative purposes neglect the effects of all higher moments; a valid first approximation if the cloud separation D is large compared to the cloud radius r_c.

We focus attention on a particular cloud and choose the system of coordinates such that the origin is at the center of the cloud and the z-axis along the direction of a possible mean field $\mathbf{B}_0$ (Figure 2). In this case the equilibrium condition for any cloud in the

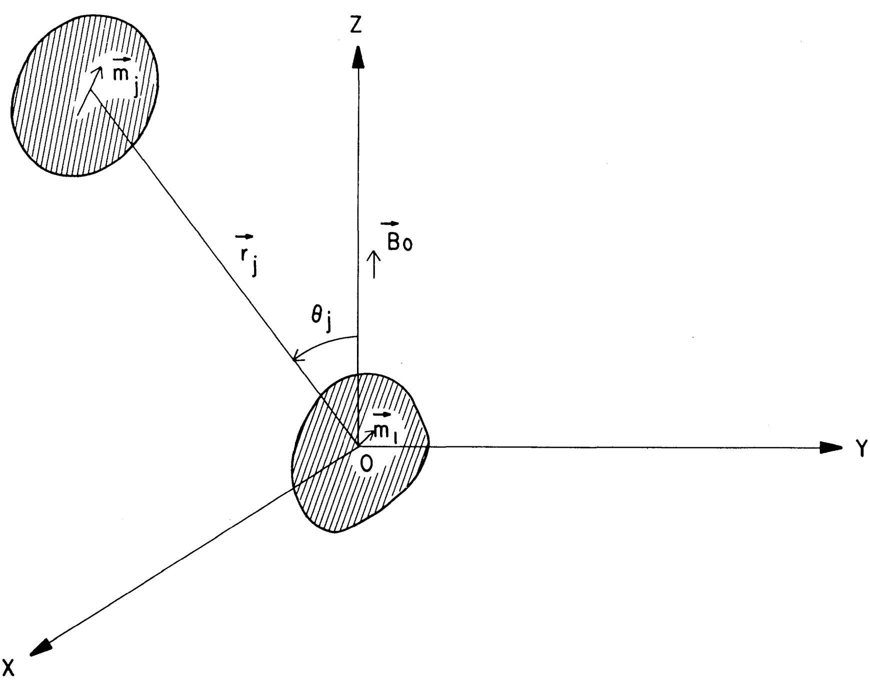


Fig. 2. The 'central cloud' and the jth cloud with dipoles $\mathbf{m}_1$ and $\mathbf{m}_j$. The coordinate system is chosen so that the origin is at the center of the central cloud and the z-axis is along the direction of the average field $\mathbf{B}_0$.

system becomes

$$\int \left(3P + \frac{B_i^2}{8\pi}\right) \mathrm{d}^3 r = -2\mathbf{m} \cdot \mathbf{B}_e(0) \equiv 2E\,, \tag{33}$$

where, P, B_i^2, and $\mathbf{m}$ refer to the particular cloud under consideration, and $\mathbf{B}_e(0)$ is the field produced by all the other remaining clouds at the center of the cloud in question. $\mathbf{B}_e(0)$ is, therefore, given by

$$\mathbf{B}_e(0) = \sum_j \frac{3\hat{r}_j \hat{r}_j \cdot \mathbf{m}_j - \mathbf{m}_j}{r_j^3}\,, \tag{34}$$

where $\hat{r}_j \equiv \mathbf{r}_j/r_j$; $\mathbf{r}_j$ being the vector from the center of the chosen cloud to the center of any other cloud labelled j and the sum runs over the remaining clouds.

4.1. Probability Distribution of E

From the virial Equation (33) we see that the effect of the magnetic field is compressive if twice the interaction energy, $-\mathbf{m} \cdot \mathbf{B}_e(0)$, is greater than the magnetic energy $\int (B_i^2/8\pi)\, \mathrm{d}^3 r$ associated with the internal field $\mathbf{B}_i$. For a random distribution of the clouds we can, therefore, ask for the probability $W(E)\, \mathrm{d}E$ that the interaction energy $E \equiv \mathbf{m} \cdot \mathbf{B}_e(0)$ has a value between E and $E + \mathrm{d}E$.

This probability distribution can clearly be written

$$W(E) = \int \mathrm{d}X_1, \ldots, \mathrm{d}X_N P(X_1, \ldots, X_N)\, \delta[E_1(X_1, \ldots, X_N) - E]\,, \tag{35}$$

where $P(X_1, \ldots, X_N)$ is the probability of finding the system of clouds in a configuration characterized by the dynamical coordinates $(X_1, \ldots, {}_{\mathrm{XN}})$, and is assumed to be given by the Gibbs distribution

$$P(X_1, \ldots, X_N) = \frac{1}{Z}\, e^{-\beta \mathscr{E}_t(X_1, \ldots, X_N)}\,, \tag{36}$$

where

$$Z = \int \mathrm{d}X_1, \ldots, \mathrm{d}X_N\, e^{-\beta \mathscr{E}_t(X_1, \ldots, X_N)} \tag{37}$$

is the partition function, $\mathscr{E}_t = \Sigma_{i=1}^N E_i(X_1, \ldots, X_N)$ is the interaction energy of the entire system of clouds. Hence,

$$E_i(X_1, \ldots, X_N) = \mathscr{E}_0(X_i) + \sum_{i=1}^{N} \mathscr{E}(X_i, X_j)\,, \tag{38}$$

$$\begin{aligned}
\mathscr{E}_t(X_1, \ldots, X_N) &= \sum_{i=1}^{N} \left[\mathscr{E}_0(X_i) + \sum_{j>i} \mathscr{E}(X_i, X_j) \right] = \\
&= \mathscr{E}_0(X_1) + \sum_{j>1}^{N} \mathscr{E}(X_i, X_j) + \sum_{i=2}^{N} \mathscr{E}_0(X_i) + \\
&\quad + \sum_{i=2} \sum_{j>i} \mathscr{E}(X_i, X_j)\,,
\end{aligned} \tag{40}$$

where

$$\mathscr{E}_0(X_i) \equiv -\mathbf{m}_i \cdot \mathbf{B}_0 , \quad i = 1, 2, \ldots, N \tag{41}$$

is the energy of the ith dipole in the mean field $\mathbf{B}_0$ and

$$\mathscr{E}(X_i, X_j) \equiv -\mathbf{m}_i \cdot \mathbf{B}_j , \quad i \neq j \tag{42}$$

is the interaction energy between dipoles $\mathbf{m}_i$ and $\mathbf{m}_j$. In particular, the interaction energy between the chosen cloud (labelled 1) and the field $\mathbf{B}_i$ produced by the ith cloud is

$$\mathscr{E}(X_1, X_i) = -\mathbf{m}_1 \cdot \mathbf{B}_i = -\mathbf{m}_1 \cdot \frac{3\hat{r}_i \hat{r}_i \cdot \mathbf{m}_i - \mathbf{m}_i}{r_i^3} , \quad i = 2, 3, \ldots, N, \tag{43}$$

where $\mathbf{r}_i$ is the vector connecting the center of the chosen cloud to that of the cloud labelled i.

By use of the Fourier representation of the Dirac function, the probability distribution Equation (35) becomes

$$W(E) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \mathrm{d}t A(t)\, e^{-iEt} , \tag{44}$$

where

$$A(t) \equiv \int \mathrm{d}X_1, \ldots, \mathrm{d}X_N P(X_1, \ldots, X_N)\, e^{itE_1(X_1, \ldots, X_N)} . \tag{45}$$

Introducing the Gibbs distribution given in Equation (36), we find this to become

$$\begin{aligned} A(t) &= \frac{1}{Z} \int \mathrm{d}X_1, \ldots, \mathrm{d}X_N\, e^{-(\beta - it)\mathscr{E}_0(X_1)} \prod_{i=2}^{N} e^{-(\beta - it)\mathscr{E}(X_1, X_i) - \beta\mathscr{E}_0(X_i)} = \\ &= \frac{1}{Z} \int \mathrm{d}X_1 \left\{ e^{-(\beta - it)\mathscr{E}_0(X_1)} \left[\int \mathrm{d}X\, e^{-(\beta - it)\mathscr{E}(X_1, X) - \beta\mathscr{E}_0(X)} \right]^{N-1} \right\} . \end{aligned} \tag{46}$$

Similarly, the partition function can be written as

$$Z = \int \mathrm{d}X_1 \{ e^{-\beta\mathscr{E}_0(X_1)} [\mathrm{d}X\, e^{-\beta\mathscr{E}(X_1, X) - \beta\mathscr{E}_0(X)}]^{N-1} \} . \tag{47}$$

The coordinates $\{X_1\}$ define the orientation $\hat{m}_1$ of the dipole of the chosen cloud, while $\{X\}$ include the position $\mathbf{r}$ of any other cloud in the system, as well as the orientation of its dipole moment. The range of r extends from a small sphere of radius r_c to a large one of radius R which contains the N clouds; hence, if n is the mean number of clouds per unit volume, then

$$N = n \frac{4\pi}{3} R^3 .$$

In writing Equations (46) and (47) we have assumed that the correlations between the clouds are weak and have, therefore, approximated the total energy (40) by

$$\mathscr{E}_t(X_1, \ldots, X_N) \approx \mathscr{E}_0(X_1) + \sum_{j>1}^{N} \mathscr{E}(X_1, X_j) + \sum_{i=2}^{N} \mathscr{E}_0(X_i). \tag{48}$$

Assuming the mean kinetic energy β^{-1} to be large compared to the energy of a dipole in the average field ($\beta E_0 \equiv \beta m B_0 \ll 1$), and the average field to be weak compared to the local field, i.e., $B_0 \ll m/D^3 \approx mn$, we can calculate the probability distribution $W(E)$ keeping terms up to 2nd order in the small quantities βE_0 and E_0/a. The details of the calculation are given in Appendix B; the final result is

$$W(E) = \frac{a/\pi}{E^2 + a^2} \exp\left\{-\beta\left[E + \frac{2a}{\pi} F(\beta\mathscr{E}_{\max})\right]\right\}\left\{1 - \frac{1}{3}\frac{E_0^2}{a^2}\frac{1 - 3E^2/a^2}{(1 + E^2/a^2)^2} - \Delta_1\right\}, \tag{49}$$

where

$$\Delta_1 \equiv \frac{\beta^2 E_0^2}{6}\left\{1 - \frac{8}{15}\frac{E_0^2}{a^2}\left[\frac{(1 - \frac{5}{2})E^2/a^2 - (\frac{13}{4})E^4/a^4}{(1 + E^2/a^2)^3} + \frac{\beta a}{8}\frac{1 - 3E^2/a^2}{(1 + E^2/a^2)^2}\right]\right\} \tag{50}$$

and

$$F(\beta\mathscr{E}_{\max}) \equiv \mathrm{Shi}(\beta\mathscr{E}_{\max}) - \frac{\cosh(\beta\mathscr{E}_{\max})}{\beta\mathscr{E}_{\max}} = \int_0^{\beta\mathscr{E}_{\max}} \frac{\sinh x}{x}\, dx - \frac{\cosh(\beta\mathscr{E}_{\max})}{\beta\mathscr{E}_{\max}};$$

$\mathscr{E}_{\max}$ being the maximum value of the dipole-dipole interaction energy is $\sim m^2/r_c^3$, and $\alpha \equiv (2\pi/9\sqrt{3})nm^2$ is the average dipole-dipole interaction energy. In (49) and (50), βE_0, $E_0)a$, and $a/\mathscr{E}_{\max}$ are all small quantities ($\ll 1$), while $\beta\mathscr{E}_{\max}$ and βa are finite.

Using Equation (49) we can calculate the probability $P_r(E > \bar{E})$ that E is greater than a certain value $\bar{E}$. For $\bar{E} \geq 0$ we can write, neglecting both β and E_0,

$$P_r(E > \bar{E}) = \frac{a}{\pi}\int_{\bar{E}}^{\infty} \frac{dE}{E^2 + a^2} = \frac{1}{2}\left[1 - \frac{\mathrm{arctg}(\bar{E}/a)}{\pi/2}\right]. \tag{51}$$

For $\bar{E} = 0$, Equation (51) reduces to

$$P_r(E > 0) = \tfrac{1}{2}.$$

For

$$\bar{E} = M_i = \frac{1}{2}\int \frac{B_i^2}{8\pi}\, d^3 r \gg a,$$

we can write

$$\mathrm{arctg}\left(\frac{M_i}{a}\right) = \frac{\pi}{2} - \frac{a}{M_i}.$$

Hence,

$$P_r(E > \bar{E}) = \frac{1}{\pi}\frac{a}{M_i}\,, \tag{52}$$

where

$$a = \frac{2\pi}{9\sqrt{3}}\, nm^3 \tag{53}$$

and

$$M_i = \frac{1}{2}\int \frac{B_i^2}{8\pi}\,\mathrm{d}^3 r = \frac{1}{2}\frac{m^2}{r_c^3}\,. \tag{54}$$

Substituting Equations (54) and (53) into Equation (52), we obtain

$$P_r(E > M_i) = \frac{4}{9\sqrt{3}}\, np_c^3\,, \tag{55}$$

where n is the number of clouds per unit volume; r_c, radius of a cloud; nr_c^3 is, therefore, the fraction of space occupied by clouds.

4.2. Radial variation of the confining magnetic field

This can be determined by calculating the average value of $B^2 = |\mathbf{B}_e + \mathbf{B}_i|^2$ over the surface of a sphere of radius $r_c \le r < D$. As before, r_c is the radius of the confined cloud and D the average intercloud distance.

For $r_c \le r < D$, $\mathbf{B}_e$ and $\mathbf{B}_i$ are given by Equations (21) and (22).

The average value $B^2(r)$ of $|\mathbf{B}_e + \mathbf{B}_i|^2$ is calculated as

$$B^2(r) = \frac{1}{4\pi}\int |\mathbf{B}_e(\mathbf{r}) + \mathbf{B}_i(\mathbf{r})|^2\,\mathrm{d}\Omega\,. \tag{56}$$

We now insert Equations (21) and (22) into Equation (56), and use the orthogonality properties of the spherical harmonics; after a straightforward calculation we obtain

$$B^2(r) = \sum_{l=1}\sum_{m=-l}^{+l}\frac{(2l+1)}{4\pi}\left[l\,|\alpha_{lm}|^2\, r^{2(l-1)} + (l+1)\,\frac{|q_{lm}|^2}{r^{2(l+2)}}\right], \tag{57}$$

where q_{lm} and α_{lm} are given by Equations (A16) and (A18), respectively.

In the present case, where $\mathbf{B}_i$ and $\mathbf{B}_e$ are dipole fields, we have:

$$\sum_{m=-1}^{+1} |q_{lm}|^2 = \frac{4\pi}{3}\, m^2\,, \tag{58}$$

$$q_{lm} = 0\,, \quad l = 2, 3, \ldots\,, \tag{59}$$

$$\sum_{m=-1}^{+1} |\alpha_{lm}|^2 = \frac{4\pi}{3} B_e^2(0). \tag{60}$$

Insertion of Equations (58)–(60) into Equation (57) leads to

$$B^2(r) = B_e^2(0) + \frac{2m^2}{r^6} + \sum_{l=2} \sum_{m=-l}^{l} \frac{l(2l+1)}{4\pi} |\alpha_{lm}|^2 r^{2(l-1)}. \tag{61}$$

Since $\mathbf{B}_e$ is the sum of dipole fields, it can be easily shown from Equations (28) and (32) that

$$|\alpha_{lm}| \sim \left| \frac{\partial^{(l-1)} B_{e_j}(0)}{\partial x_i^{(l-1)}} \right| < \frac{1}{D^{(l-1)}} \frac{m}{D^3} ;$$

and, hence,

$$|\alpha_{lm}|^2 r^{2(l-1)} < \left(\frac{r}{D}\right)^{2(l-1)} \frac{m^2}{D^6} = \left(\frac{r}{D}\right)^{2(l+2)} \frac{m^2}{r^6}, \quad l = 2, 3, \ldots . \tag{62}$$

From Equations (61) and (62) we see that, for $r_c \leq r \ll D$, $B^2(r)$ decreases as $1/r^6$, reaches a minimum, and then increases as r approaches D, the distance to the nearest neighbour.

5. Conclusion

The condition for the equilibrium Equation (27) derived here in the general case where external sources are present shows that the magnetic field may, under certain circumstances, provide a triggering mechanism for cloud contraction. If, indeed, there are a number of isolated cosmic clouds in random motion, then we may expect that for a small but finite fraction of them (roughly equal to the fraction of space occupied by clouds) the magnetic field is compressive.

Of course, we have presented only a crude approximation to the actual interaction of large local clouds, and have left aside the vexing question of internal equilibrium. The conclusions are, therefore, far from profound.

Acknowledgements

M. Azar wishes to acknowledge support from the grant NASA NSG 7623. He also wishes to thank Professor H. Alfvén for many helpful discussions.

Appendix A. Evaluation of the Coefficients α_{lm} and q_{lm}

In terms of the vector potential

$$\mathbf{A}(\mathbf{r}) = \frac{1}{c} \int \frac{\mathbf{J}_i(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, d^3 r' + \frac{1}{c} \int \frac{\mathbf{J}_e(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, d^3 r' ,$$

the magnetic field is written as

$$\mathbf{B} = \nabla \times \mathbf{A} = \frac{1}{c} \int \frac{\nabla' \times \mathbf{J}_i(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, \mathrm{d}^3 r' + \frac{1}{c} \int \frac{\nabla' \times \mathbf{J}_e(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, \mathrm{d}^3 r' ;$$

$\mathbf{J}_i$ and $\mathbf{J}_e$ being the internal and external current densities, respectively. Using

$$\frac{1}{|\mathbf{r} - \mathbf{r}'|} = \sum_{lm} \frac{4\pi}{(2l+1)} \frac{r^l_<}{r^{l+1}_>} Y_{lm}(\theta, \phi) Y^*_{lm}(\theta', \phi'),$$

we obtain

$$\mathbf{B} = \sum_{lm} \frac{Y_{lm}(\theta, \phi)}{r^{l+1}} \mathbf{M}_{lm} + \sum_{lm} r^l Y_{lm}(\theta, \phi) \mathbf{J}_{lm} \tag{A1}$$

for the magnetic field in the region between the internal and external sources. The constant vectors $\mathbf{M}_{lm}$ and $\mathbf{J}_{lm}$ are given by

$$\mathbf{M}_{lm} \equiv \frac{4\pi}{c(2l+1)} \int r^l Y^*_{lm}(\theta, \phi) \mathbf{J}_i(\mathbf{r}) \, \mathrm{d}^3 r, \tag{A2}$$

$$\mathbf{J}_{lm} = \frac{4\pi}{c(2l+1)} \int \frac{Y^*_{lm}(\theta, \phi)}{r^{l+1}} \mathbf{J}_e(\mathbf{r}) \, \mathrm{d}^3 r. \tag{A3}$$

In the region outside of the cloud, the part of the magnetic field which is due to currents within the cloud is given by

$$\mathbf{B}_i = \sum_{lm} \frac{Y_{lm}(\theta, \phi)}{r^{l+1}} \mathbf{M}_{lm}. \tag{A4}$$

Comparison with (22) gives

$$-\sum_{lm} \nabla \left(\frac{Y_{lm}(\theta, \phi)}{r^{l+1}} \right) q_{lm} = \sum_{lm} \frac{Y_{lm}(\theta, \phi)}{r^{l+1}} \mathbf{M}_{lm}. \tag{A5}$$

Multiplying both sides by $Y^*_{l'm'}(\theta, \phi)\hat{r}(\hat{r} = \mathbf{r}/r)$ and then integrating over all solid angles, we get

$$\delta_{l, l'+1}(l'+1) q_{l'm'} = \sum_l \sum_{m=-l}^{+l} \left\{ \mathbf{M}_{lm} \cdot \int \hat{r} Y_{lm}(\theta, \phi) Y^*_{l'm'}(\theta, \phi) \, \mathrm{d}\Omega \right\}. \tag{A6}$$

The integral $\int \hat{r} Y_{lm}(\theta, \phi) Y^*_{l'm'}(\theta, \phi) \, \mathrm{d}\Omega$ can be evaluated by virtue of the orthogonality properties and recursions relations for the associated Legendre polynomials P^m_l. The end result is

$$\int Y_{lm}(\theta, \phi) Y^*_{l'm'}(\theta, \phi) \hat{r} \, d\Omega =$$

$$= \frac{\delta_{l, l'+1}}{\sqrt{(2l'+1)(2l'+3)}} \left\{ \sqrt{(l'-m+1)(l'-m'+2)} \, \delta_{m, m'-1} \frac{\hat{x} - i\hat{y}}{2} - \right.$$

$$- \sqrt{(l'+m'+1)(l'+m'+2)} \, \delta_{m, m'+1} \frac{\hat{x} + i\hat{y}}{2} +$$

$$\left. + \sqrt{(l'+m'+1)(l'-m'+1)} \, \delta_{m, m'} \hat{z} \right\} +$$

$$+ \frac{\delta_{l, l'11}}{\sqrt{(2l'+1)(2l'-1)}} \left\{ \sqrt{(l'-m'-1)(l'-m')} \, \delta_{m, m'+1} \frac{\hat{x} + i\hat{y}}{2} - \right.$$

$$\left. - \sqrt{(l'+m'-1)(l'+m')} \, \delta_{m, m'-1} \frac{\hat{x} - i\hat{y}}{2} + \sqrt{(l'+m')(l'-m')} \, \delta_{m, m'} \hat{z} \right\}. \tag{A7}$$

Therefore, using (A7) in (A6) we obtain

$$(l+1) q_{lm} = \frac{1}{\sqrt{(2l+1)(2l+3)}} \left\{ \sqrt{(l-m+2)(l-m+1)} \, M^-_{l+1, m-1} - \right.$$

$$- \sqrt{(l+m+2)(l+m+1)} \, M^+_{l+1, m+1} +$$

$$\left. + \sqrt{(l+m+1)(l-m+1)} \, M^3_{l+1, m} \right\}, \tag{A8}$$

$$0 = \sqrt{(l-m)(l-m-1)} \, M^+_{l-1, m+1} - \sqrt{(l+m)(l+m-1)} \, M^-_{l-1, m-1} +$$

$$+ \sqrt{(l-m)(l+m)} \, M^3_{l-1, m} \, . \tag{A9}$$

By combining Equations (A8) and (A9), the final result becomes

$$(l+1) q_{lm} = \sqrt{\frac{2l+3}{2l+1}} \left\{ \sqrt{\frac{l-m+1}{l-m+2}} \, M^-_{l+1, m-1} - \right.$$

$$\left. - \sqrt{\frac{l+m+1}{l+m+2}} \, M^+_{l+1, m+1} \right\}, \tag{A10}$$

where

$$M^{\pm}_{lm} \equiv \frac{\hat{x} \cdot \mathbf{M}_{lm} \pm i\hat{y} \cdot \mathbf{M}_{lm}}{2} \, , \quad M^3_{lm} \equiv \hat{z} \cdot \mathbf{M}_{lm} \, , \tag{A11}$$

$|m| \le l$; $l = 1, 2, 3, \ldots$. Following similar steps, we can express the coefficients α_{lm} in terms of $\mathbf{J}_{lm}$ as

$$-l\alpha_{lm} = \frac{1}{\sqrt{(2l+1)(2l-1)}} \{\sqrt{(l-m)(l-m-1)}\, J^{+}_{l-1,m+1} - \\ - \sqrt{(l+m)(l+m-1)}\, J^{-}_{l-1,m-1} + \sqrt{(l+m)(l-m)}\, J^{3}_{l-1,m}\}\,. \tag{A12}$$

The components J^{+}_{lm}, J^{-}_{lm}, and J^{3}_{lm} are related by

$$\sqrt{(l-m+2)(l-m+1)}\, J^{-}_{l+1,m-1} - \sqrt{(l+m+2)(l+m+1)}\, J^{+}_{l+1,m+1} + \\ + \sqrt{(l+m+1)(l-m+1)}\, J^{3}_{l+1,m} = 0 \tag{A13}$$

or

$$-l\alpha_{lm} = \sqrt{\frac{2l-1}{2l+1}} \left\{ \sqrt{\frac{l-m}{l-m-1}}\, J^{+}_{l-1,m+1} - \sqrt{\frac{l+m}{l+m-1}}\, J^{-}_{l-1,m-1} \right\}, \tag{A14}$$

$l = 1, 2, 3, \ldots$; $|m| \le l$, where

$$J^{\pm}_{lm} \equiv \frac{\hat{x}\cdot\mathbf{J}_{lm} \pm i\hat{y}\cdot\mathbf{J}_{lm}}{2}\,, \quad J^{3}_{lm} \equiv \hat{z}\cdot\mathbf{J}_{lm}\,;$$

$\mathbf{J}_{lm}$ being given by (A3). Let us note that

$$\alpha_{l,-m} = (-1)^m \alpha^{*}_{l,m}\,.$$

Similarly,

$$q_{l,-m} = (-1)^m q^{*}_{l,m}\,.$$

To see the physical meaning of $\mathbf{M}_{lm}$ and $\mathbf{J}_{lm}$ (and, therefore, q_{lm} and α_{lm}), we calculate explicitly the first few terms. It can easily be shown that

$$\begin{aligned} \mathbf{M}_{1m} &= 0\,, \\ \mathbf{M}_{20} &= \sqrt{\frac{4\pi}{5}}\, [3\hat{z}\hat{z}\cdot - \mathbf{I}]\cdot\mathbf{m}\,, \\ \mathbf{M}_{21} &= -\sqrt{\frac{6\pi}{5}}\, [(\hat{x} - i\hat{y})\hat{z} + \hat{z}(\hat{x} - i\hat{y}))\cdot\mathbf{m}\,, \\ \mathbf{M}_{22} &= \sqrt{\frac{6\pi}{5}}\, [(\hat{x} - i\hat{y})(\hat{x} - i\hat{y})]\cdot\mathbf{m}\,, \quad \text{etc}\,; \end{aligned} \tag{A15}$$

and from (A10), we have

$$q_{10} = \sqrt{\frac{4\pi}{3}}\ \hat{z}\cdot\mathbf{m}\,,$$

$$q_{11} = -\sqrt{\frac{2\pi}{3}}\ (\hat{x} - i\hat{y})\cdot\mathbf{m}\,,$$

$$q_{20} = 12\sqrt{\frac{\pi}{5}}\ \mathbf{z}\cdot\mathsf{m}\cdot\mathbf{z}\,, \tag{A16}$$

$$q_{21} = -2\sqrt{\frac{6\pi}{5}}\ [(\hat{x} - i\hat{y})\cdot\mathsf{m}\cdot\hat{z} + \hat{z}\cdot\mathsf{m}\cdot(\hat{x} - i\hat{y})]\,,$$

$$q_{22} = 2\sqrt{\frac{6\pi}{5}}\ (\hat{x} - i\hat{y})\cdot\mathsf{m}\cdot(\hat{x} - i\hat{y})\,,\quad \text{etc.}\,;$$

$\mathbf{m}$ and m being, respectively, the magnetic dipole and quadrupole moments due to the internal current density $\mathbf{J}_i$:

$$\mathbf{m} = \frac{1}{2c}\int \mathbf{r}\times\mathbf{J}_i(\mathbf{r})\,\mathrm{d}^3 r\,;\qquad \mathsf{m} = \frac{1}{6c}\int \mathbf{r}\times\mathbf{J}_i(\mathbf{r})\mathbf{r}\,\mathrm{d}^3 r\,.$$

Similarly, using Equation (A3), we can show that

$$\mathbf{J}_{00} = \sqrt{4\pi}\,\mathbf{B}_e(0)\,,$$

$$\mathbf{J}_{10} = \sqrt{\frac{4\pi}{3}}\ \hat{z}\cdot\nabla\mathbf{B}_e(0)\,, \tag{A17}$$

$$\mathbf{J}_{11} = -\sqrt{\frac{2\pi}{3}}\ (\hat{x} - i\hat{y})\cdot\nabla\mathbf{B}_e(0)\,,$$

etc.; and from (A14), we obtain

$$\alpha_{1,0} = -\sqrt{\frac{4\pi}{3}}\ \hat{z}\cdot\mathbf{B}_e(0)\,,$$

$$\alpha_{1,1} = -\alpha^*_{1,1} = \sqrt{\frac{2\pi}{3}}\ (\hat{x} - i\hat{y})\cdot\mathbf{B}_e(0)\,,$$

$$\alpha_{2,0} = -\sqrt{\frac{\pi}{5}}\ \nabla\cdot\{\hat{z}\hat{z}\cdot\mathbf{B}_e(0)\}\,, \tag{A18}$$

$$\alpha_{2,1} = -\alpha^*_{2,-1} = \sqrt{\frac{\pi}{30}}\ \nabla\cdot\{[(\hat{x} - i\hat{y})\hat{z} + \hat{z}(\hat{x} - i\hat{y})]\cdot\mathbf{B}_e(0)\}\,,$$

$$\alpha_{2,2} = \alpha^*_{2,-2} = \sqrt{\frac{\pi}{30}}\ \nabla\cdot\{[(\hat{x} - i\hat{y})(\hat{x} - i\hat{y})]\cdot\mathbf{B}_e(0)\}\,,\quad \text{etc.}\,;$$

where $\mathbf{B}_e$ is the magnetic field produced by the external sources. $\mathbf{B}_e(0)$, $(\partial/\partial x)\mathbf{B}_e(0)$, $(\partial/\partial y)\mathbf{B}_e(0)$, etc., denote the value of $\mathbf{B}_e$ and its various derivatives at the origin.

Appendix B. Evaluation of the Probability Distribution Function $W(E)$

From Equations (44), (46), and (47) we have

$$W(E) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \mathrm{d}t\, A(t)\, e^{-iEt}, \tag{44}$$

$$A(t) = \frac{1}{Z} \int \mathrm{d}X_1 \left\{ e^{-(\beta - it)\mathscr{E}_0(X_1)} \left[\int \mathrm{d}X\, e^{-(\beta - it)\mathscr{E}(X_1, X) - \beta\mathscr{E}_0(X)} \right]^{N-1} \right\}, \tag{46}$$

$$Z = \int \mathrm{d}X_1 \left\{ e^{-\beta\mathscr{E}_0(X_1)} \left[\int \mathrm{d}X\, e^{-\beta\mathscr{E}(X_1, X) - \beta\mathscr{E}_0(X)} \right]^{N-1} \right\}; \tag{47}$$

where

$$\begin{aligned} \mathrm{d}X_1 &= \mathrm{d}^2\Omega_1, \\ \mathrm{d}X &= r^2\, \mathrm{d}r\, \mathrm{d}^2 w\, \mathrm{d}^2\Omega, \quad r_c \le r \le R; \end{aligned} \tag{B1}$$

while

$$\mathscr{E}_0(X_i) = -\mathbf{m}_i \cdot \mathbf{B}_0 \equiv -E_0 \hat{m}_i \cdot \hat{B}_0, \tag{B2}$$

$$\mathscr{E}(X_1, X) = \frac{m^2}{r^3} f(\Omega_1; w, \Omega), \tag{B3}$$

$$f(\Omega_1, w, \Omega) \equiv \hat{m}_1 \cdot \hat{m}_i - 3\hat{m}_1 \cdot \hat{r}_i \hat{r}_i \cdot \hat{m}_i. \tag{B4}$$

Insertion of (47) in (46) leads to

$$\int \mathrm{d}X_1\, e^{-\beta\mathscr{E}_0(X_1)} \left[\int \mathrm{d}X\, e^{-\beta\mathscr{E}(X_1, X) - \beta\mathscr{E}_0(X)} \right]^{N-1} A(t) - e^{it\mathscr{E}_0(X_1)} \times$$

$$\times \frac{\left[\int \mathrm{d}X\, e^{-(\beta - it)\mathscr{E}(X_1, X) - \beta\mathscr{E}_0(X)} \right]^{N-1}}{\int \mathrm{d}X\, e^{-\beta\mathscr{E}(X_1, X) - \beta\mathscr{E}_0(X)}} = 0. \tag{B5}$$

We now take the limit of this equation as N (or R) $\to \infty$:

$$\lim_{N \to \infty} \int \mathrm{d}X\, e^{-\beta\mathscr{E}(X_1, X) - \beta\mathscr{E}_0(X)} =$$

$$= \lim_{N \to \infty} \int \mathrm{d}^2 w\, \mathrm{d}^2\Omega \left\{ e^{-\beta\mathscr{E}_0(\Omega)} \int_{r_c}^{R} \mathrm{d}r\, r^2 \exp\left[-\beta \frac{m^2}{r^3} f(\Omega_1; w, \Omega) \right] \right\}.$$

On the other hand, we observe that

$$\lim_{R\to\infty} \frac{\int_{r_c}^{R} \mathrm{d}r\, r^2 \exp\left[-\beta \frac{m^2}{r^3} f(\Omega_1, w, \Omega)\right]}{R^3} =$$

$$= \lim_{R\to\infty} \frac{\frac{\partial}{\partial R}\int_{r_c}^{R} \mathrm{d}r\, r^2 \exp\left[-\beta \frac{m^2}{r^3} f(\Omega; w, \Omega)\right]}{\frac{\partial}{\partial R} R^3} =$$

$$= \lim_{R\to\infty} \frac{R^2 \exp\left[-\beta \frac{m^2}{R^3} f(\Omega_1, w, \Omega)\right]}{3R^2} = \frac{1}{3}\,.$$

Thus

$$\lim_{N\to\infty} \int_{r_c}^{R} \mathrm{d}r\, r^2 \exp\left[-\beta \frac{m^2}{r^3} f(\Omega_1; w, \Omega)\right] = \frac{R^3}{3}$$

and

$$\lim_{N\to\infty} \int \mathrm{d}X\, e^{-\beta\mathscr{E}(X_1, X) - \beta\mathscr{E}_0(X)} = \lim_{N\to\infty} \left\{\frac{(4\pi)^2 R^3}{3}\, h(\beta E_0)\right\} =$$

$$= \lim_{N\to\infty} \left\{\frac{4\pi}{n}\, N h(\beta E_0)\right\} \simeq \lim_{N\to\infty} \left\{\frac{4\pi}{n}\, (N-1) h(\beta E_0)\right\}, \tag{B6}$$

where

$$h(\beta E_0) = \int \frac{\mathrm{d}^2\Omega}{4\pi}\, e^{-\beta\mathscr{E}_0(\Omega)} = \frac{\sinh \beta E_0}{\beta E_0}\,. \tag{B7}$$

Furthermore,

$$\lim_{N\to\infty} \frac{\left[\int \mathrm{d}x\, e^{-(\beta - it)\mathscr{E}(X_1, X) - \beta\mathscr{E}_0(X)}\right]^{N-1}}{\int \mathrm{d}x\, e^{-\beta\mathscr{E}(X_1, X) - \beta\mathscr{E}_0(X)}} =$$

$$= \lim_{N\to\infty} 1 - \frac{\left[\int \mathrm{d}X[1 - e^{it\mathscr{E}(X_1, X)}]\, e^{-\beta[\mathscr{E}_0(X) + \mathscr{E}(X_1, X)]}\right]^{N-1}}{\int \mathrm{d}x\, e^{-\beta[\mathscr{E}_0(X) + \mathscr{E}(X_1, X)]}} =$$

$$= \lim_{N\to\infty} \left[1 - \frac{1}{N-1}\, \frac{n}{4\pi h(\beta E_0)}\, C(t; \beta E_0, X_1)\right]^{N-1} = \exp\left[-\frac{n}{4\pi}\, \frac{C(t; \beta E_0, X_1)}{h(\beta E_0)}\right], \tag{B8}$$

where we have used Equations (B6) and (B7) and set

$$\int \mathrm{d}X\{[1 - e^{it\mathscr{E}(X_1, X)}]\, e^{-\beta[\mathscr{E}_0(X) + \mathscr{E}(X_1, X)]}\} \equiv C(t;\, \beta E_0; X_1)\,. \tag{B9}$$

Substituting (B6) and (B9) in Equation (B5), we obtain

$$\int \mathrm{d}X_1\, e^{-\beta\mathscr{E}_0(X_1)} \left\{ \lim_{N\to\infty} \left[\frac{4\pi(N-1)}{n}\, h(\beta E_0) \right]^{N-1} \right\} \times$$

$$\times \left\{ A(t) - e^{it\mathscr{E}_0(X_1)} \lim_{N\to\infty} \left[1 - \frac{1}{N-1}\, \frac{n}{4\pi}\, \frac{C(t;\, \beta E_0, X_1)}{h(\beta E_0)} \right]^{N-1} \right\} = 0$$

or

$$\int \mathrm{d}X_1\, e^{-\beta\mathscr{E}_0(X_1)} \left\{ A(t) - e^{it\mathscr{E}_0(X_1)} \exp\left[-\frac{n}{4\pi}\, \frac{C(t;\, \beta E_0, X_1)}{h(\beta E_0)} \right] \right\} = 0\,. \tag{B10}$$

Hence,

$$A(t) = \frac{1}{h(\beta E_0)} \int \frac{\mathrm{d}X_1}{4\pi} \left\{ e^{-(\beta - it)\mathscr{E}_0(X_1)} \exp\left[-\frac{n}{4\pi}\, \frac{C(t;\, \beta E_0, X_1)}{h(\beta E_0)} \right] \right\}. \tag{B11}$$

From (B9), we have

$$\begin{aligned} C(t;\, \beta E_0, X_1) &= \int \mathrm{d}X[1 - e^{it\mathscr{E}(X_1, X)}]\, e^{-\beta[\mathscr{E}(X_1, X) + \mathscr{E}_0(X)]} = \\ &= \int \mathrm{d}^2 w\, \mathrm{d}^2\Omega\, e^{-\beta\mathscr{E}_0(\Omega)} \int_{r_c}^{\infty} \mathrm{d}r\, r^2 (1 - e^{it\mathscr{E}(X_1, X)})\, e^{-\beta\mathscr{E}(X_1, X)}\,. \end{aligned} \tag{B12}$$

We now make a change of variable $r \to \mathscr{E} = (m^2/r^3) f(\Omega_1;\, w, \Omega)$

$$r^2\, \mathrm{d}r = r^2\, \frac{\mathrm{d}r}{\mathrm{d}\mathscr{E}}\, \mathrm{d}\mathscr{E} = \frac{1}{3}\, \frac{m^2}{\mathscr{E}^2}\, f(\Omega_1;\, w;\, \Omega)\, \mathrm{d}\mathscr{E},$$

where $\mathscr{E}$ runs over positive values and f has either sign. We can restrict the range of f to positive values and allow $\mathscr{E}$ to take either sign by writing $f \to \frac{1}{2}\,|f|$. Then

$$C(t,\, \beta E_0, X_1) = \frac{1}{3}\, m^2 \left\{ \int \mathrm{d}^2 w\, \mathrm{d}^2\Omega\, e^{-\beta\mathscr{E}_0(\Omega)}\, |f(\Omega_1;\, w, \Omega)| \right\} \times$$

$$\times \frac{1}{2} \int_{-\mathscr{E}_{\mathrm{max}}}^{\mathscr{E}_{\mathrm{max}}} \frac{1 - e^{it\mathscr{E}}}{\mathscr{E}^2}\, e^{-\beta\mathscr{E}}\, \mathrm{d}\mathscr{E} \equiv \tag{B13}$$

$$\equiv \frac{m^2}{3}\, I(\beta E_0;\, X_1) J(t;\, \beta\mathscr{E}_{\mathrm{max}})\,, \tag{B14}$$

where

$$I(\beta E_0; X_1) = \int d^2 w\, d^2\Omega\, |f(\Omega_1; w, \Omega)|\, e^{-\beta \mathscr{E}_0(\Omega)} \tag{B15}$$

and

$$J(t; \beta\mathscr{E}_{\max}) = \frac{1}{2} \int\limits_{-\mathscr{E}_{\max}}^{\mathscr{E}_{\max}} d\mathscr{E}\, \frac{1 - e^{it\mathscr{E}}}{\mathscr{E}^2}\, e^{-\beta\mathscr{E}} =$$

$$= \beta F(\beta\mathscr{E}_{\max}) - (\beta - it) F[(\beta - it)\mathscr{E}_{\max}]\,; \tag{B16}$$

while

$$F(x) \equiv \mathrm{Shi}(x) - \frac{\cosh x}{x}, \quad \mathrm{Shi}(x) = \int\limits_0^x \frac{\sinh y}{y}\, dy; \quad \mathscr{E}_{\max} \sim \frac{m^2}{r_c^3}. \tag{B17}$$

Insertion of (B16) and (B14) in (B11) leads to

$$A(t) = \frac{1}{h(\beta E_0)} \int \frac{dX_1}{4\pi} \exp\left\{ -(\beta - it)\mathscr{E}_0(X_1) - \right.$$
$$\left. - \frac{nm^2}{12\pi} \frac{I(\beta\mathscr{E}_{\max}; X_1)}{h(\beta E_0)} [\beta F(\beta\mathscr{E}_{\max}) - (\beta - it) F[(\beta - it)\mathscr{E}_{\max}]] \right\}. \tag{B18}$$

The probability distribution (44) may therefore be written in the form

$$W(E) = \frac{1}{2\pi} \frac{1}{h(\beta E_0)} \int \frac{dX_1}{4\pi} \left[\!\!\left[G(X_1) \times \right.\right.$$
$$\left.\left. \times \exp\left\{ -\beta \left[E + \frac{nm^2}{12\pi} \frac{I(\beta\mathscr{E}_{\max}, X_1)}{h(\beta E_0)} F(\beta\mathscr{E}_{\max}) \right] \right\} \right]\!\!\right], \tag{B19}$$

where

$$G(X_1) \equiv \int\limits_{-\infty + i\beta}^{+\infty + i\beta} dz \exp\left\{ -i[E - \mathscr{E}_0(X_1)]z - \frac{nm^2}{12\pi} \frac{I(\beta\mathscr{E}_{\max}, X_1)}{h(\beta E_0)} \times \right.$$
$$\left. \times \left[z S_i(\mathscr{E}_{\max} z) + \frac{\cos \mathscr{E}_{\max} z}{z} \right] \right\}. \tag{B20}$$

In writing (B20) we have made a change of variable $t \to z = t + i\beta$ and used

$$\mathrm{Si}(x) = -i\, \mathrm{Shi}(ix) \equiv \int\limits_0^x \frac{\sin y}{y}\, dy.$$

(a) EVALUATION OF $I(\beta E_0; X_1)$

To carry out the integration in (B15), we rotate the coordinate axes so that the z-axis is along the direction of $\mathbf{m}_1$, the dipole of the chosen cloud. Then

$$f = \hat{m}_1 \cdot \hat{m}_i - 3\hat{m}_1 \cdot \hat{r}_i \cdot \hat{m}_i = \tag{B21}$$

$$= (1 - 3u^2)v - 3u\sqrt{1 - u^2}\sqrt{1 - v^2}\cos(\varphi - \alpha) \equiv A - B\cos(\varphi - \alpha), \tag{B21}$$

$$-\beta\mathscr{E}_0(\Omega) = \beta E_0[vv_1 + \sqrt{1 - v^2}\sqrt{1 - v_1^2}\cos(\alpha - \alpha_1)] \equiv$$

$$\equiv \beta E_0[C + D\cos(\alpha - \alpha_1)], \tag{B22}$$

where

$$u = \hat{m}_1 \cdot \hat{r} = \cos\theta, \quad v = \hat{m}_1 \cdot \hat{m}_i = \cos\lambda, \quad v_1 = \hat{m}_1 \cdot \hat{B}_0 = \cos\lambda_1; \tag{B23}$$

(θ, φ), (λ, α), and (λ_1, α_1) define, respectively, the direction of the position vector $\mathbf{r}$, the ith dipole $\mathbf{m}_i$, and the mean field $\mathbf{B}_0$ in the new coordinate system. Substitution of (B21) and (B22) in (B15) yields

$$I(\beta E_0; X_1) = \int_{-1}^{+1} du \int_{-1}^{+1} dv \int_{-\theta}^{+2\pi} d\varphi \int_0^{2\pi} d\alpha\, |A - B\cos(\varphi - \alpha)| \times$$

$$\times e^{\beta E_0[C + D\cos(\alpha - \alpha_1)]} =$$

$$= \int_{-1}^{+1} dv\, |v| \int_0^{2\pi} d\alpha\, e^{\beta E_0[C + D\cos(\alpha - \alpha_1)]} \int_{-1}^{-1} du\, |1 - 3u^2| =$$

$$= \frac{8}{3\sqrt{3}} \int_{-1}^{+1} dv\, |v| \int_0^{2\pi} d\alpha\, e^{\beta E_0[C + D\cos(\alpha - \alpha_1)]}. \tag{B24}$$

For $\beta E_0 \ll 1$, we can expand $e^{\beta E_0[C + D\cos(\alpha - \alpha_1)]}$ as

$$e^{+\beta E_0[C + D\cos(\alpha - \alpha_1)]} = 1 + \beta E_0[C + D\cos(\alpha - \alpha_1)] +$$

$$+ \frac{\beta^2 E_0^2}{2}[C + D\cos(\alpha - \alpha_1)]^2 + \cdots.$$

Using this in (B24) and keeping terms up to second order in βE_0, we obtain

$$I(\beta E_0, X_1) \approx \frac{16\pi}{3\sqrt{3}}\left[1 + \frac{\beta^2 E_0^2}{8}(1 + v_1^2)\right], \tag{B25}$$

where $v_1 \equiv \hat{m}_1 \cdot \hat{B}_0$.

(b) EVALUATION OF $G(X_1)$

To evaluate $G(X_1)$ we consider the integral of

$$H(z) \equiv \exp\left\{-i[R - \mathscr{E}_0(X_1)]z - \frac{2a}{\pi}\left[z\,\mathrm{Si}(\mathscr{E}_{\max} z) + \right.\right.$$

$$\left.\left. + \frac{\cos \mathscr{E}_{\max} z}{\mathscr{E}_{\max}}\right] g(\beta E_0; X_1)\right\} \tag{B26}$$

along the closed contour shown in Figure 3

$$a \equiv \frac{2\pi}{9\sqrt{3}}\, nm^2 ; \tag{B27}$$

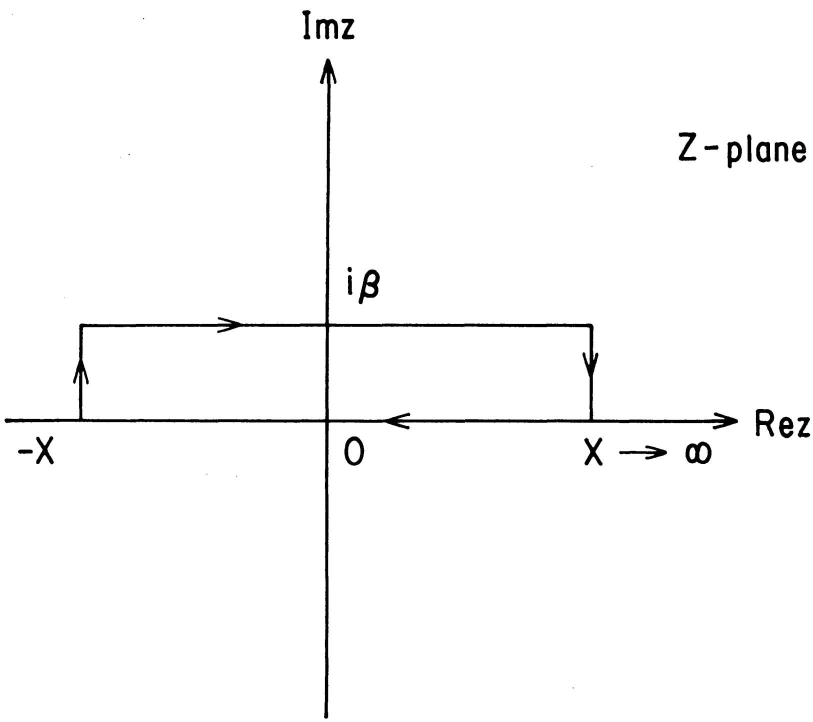


Fig. 3. Contour for the integral (B29).

being the (average) dipole-dipole interaction energy, and

$$g(\beta E_0; X_1) \equiv \frac{1 + \dfrac{\beta^2 E_0^2}{8}(1 + v_1^2)}{h(\beta E_0)} \approx 1 - \frac{\beta^2 E_0^2}{24}(1 - 3v_1^2) . \tag{B28}$$

Since $H(z)$ is an entire function of z, then

$$\lim_{X\to\infty} \oint H(z)\,\mathrm{d}z =$$

$$= \lim_{X\to\infty} \left\{ \int\limits_{-X+i\beta}^{X+i\beta} + \int\limits_{X+i\beta}^{X} + \int\limits_{X}^{-X} + \int\limits_{-X}^{-X+i\beta} H(z)\,\mathrm{d}z \right\} = 0\,. \tag{B29}$$

It is easy to show that, for large X,

$$|H(\pm X + i\beta)| \sim \frac{e^{-(2a/\pi)X}}{E_0 X}\,.$$

Hence,

$$\lim_{X\to\infty} \int\limits_{\pm X}^{\pm X+i\beta} H(z)\,\mathrm{d}z = 0\,. \tag{B30}$$

This result, together with Equation (B29), allow us to write

$$\int\limits_{-\infty+i\beta}^{+\infty+i\beta} H(z)\,\mathrm{d}z = \int\limits_{-\infty}^{+\infty} H(z)\,\mathrm{d}z\,. \tag{B31}$$

Then

$$G(X_1) = \int\limits_{-\infty}^{+\infty} \mathrm{d}z \exp\left\{ -i[E - \mathscr{E}_0(X_1)]z - \frac{2a}{\pi}\left[z\,\mathrm{Si}(\mathscr{E}_{\max} z) + \right.\right.$$

$$\left.\left. + \frac{\cos \mathscr{E}_{\max} z}{\mathscr{E}_{\max}} \right] g(\beta E_0; X_1) \right\}. \tag{B32}$$

Assuming $\mathscr{E}_{\max}/a \gg 1$, we can use the asymptotic series expansion of the sine integral. To zeroth-order in $\mathscr{E}_{\max}$, we have

$$z\,\mathrm{Si}(\mathscr{E}_{\max} z) + \frac{\cos \mathscr{E}_{\max} z}{z} \approx \frac{\pi}{2}\,|z|\,. \tag{B33}$$

In this case (B32) becomes

$$G(X_1) \approx \frac{2}{a} \frac{g(\beta E_0, X_1)}{g^2(\beta E_0; X_1) + \left[\dfrac{E - \mathscr{E}_0(X_1)}{a}\right]^2}\,. \tag{B34}$$

We can, therefore, express the probability distribution $W(E)$ in the form

$$W(E) \approx \frac{1}{2\pi a} \frac{e^{-\beta E}}{h(\beta E_0)} \int_{-1}^{+1} \mathrm{d}x \frac{g(\beta E_0; x)}{g^2(\beta E_0, x) + \left[\dfrac{E + E_0 x}{a}\right]^2} \times$$

$$\times \exp\left[- \beta a g(\beta E_0; x) \frac{2}{\pi} F(\beta \mathscr{E}_{\max})\right] , \tag{B35}$$

where, as before

$$E_0 = mB_0 ,$$

$$a = \frac{2\pi}{9\sqrt{3}} nm^2 ,$$

$$h(\beta E_0) = \frac{\sinh \beta E_0}{\beta E_0} \approx 1 - \frac{\beta^2 E_0^2}{6} , \quad \beta E_0 \ll 1 ,$$

and

$$g(\beta E_0; x) = 1 - \frac{\beta^2 E_0^2}{24} (1 - 3x^2) .$$

Assuming that the energy E_0 of the dipole in the mean field $\mathbf{B}_0$ to be small compared to the average interaction energy a, enables us to perform the integration in (B35). Keeping terms up to second order in E_0/a, we obtain

$$W(E) = \frac{\pi/a}{E^2 + a^2} \exp\left\{ - \beta \left[E + a \frac{2F}{\pi} (\beta \mathscr{E}_{\max}) \right] \right\} \times$$

$$\times \left\{ I_0(y) - \frac{E_0^2}{a^2} \frac{1 - 3E^2/a^2}{(1 + E^2/a^2)^2} I_2(y) + \Delta \right\} , \tag{B36}$$

where

$$\Delta \equiv \frac{\beta^2 E_0^2}{24} \frac{1 - E^2/a^2}{1 + E^2/a^2} (I_0 - 3I_2) -$$

$$- 4I_0 + \frac{E_0^2}{a^2} \quad 4 \frac{1 - 3E^2/a^2}{(1 + E^2/a^2)^2} I_2 - 3 \frac{1 - \dfrac{10}{3} \dfrac{E^2}{a^2} + \dfrac{11}{3} \dfrac{E^4}{a^4}}{(1 + E^2/a^2)^3} (I_2 - 3I_4) \quad , \tag{B37}$$

in which

$$I_n(y) \equiv \frac{e^{y^2/3}}{2} \int_{-1}^{+1} \mathrm{d}x\, e^{-y^2x^2} x^n . \tag{B38}$$

For $n = 0$, 2, and 4, we have

$$\begin{aligned}
I_0(y) &= \frac{\sqrt{\pi}}{2} \frac{e^{y^2/3}}{y} \operatorname{erf}(y), \\
I_2(y) &= \frac{e^{y^2/3}}{2} \left[\frac{\sqrt{\pi}}{2} \frac{\operatorname{erf}(y)}{y^3} - \frac{e^{-y^2}}{y^2} \right], \\
I_4(y) &= \frac{e^{y^2/3}}{2} \left[\frac{3\sqrt{\pi}}{4} \frac{\operatorname{erf}(y)}{y^5} - \frac{3}{2} \frac{e^{-y^2}}{y^4} - \frac{e^{-y^2}}{y^2} \right];
\end{aligned} \tag{B39}$$

where

$$y \equiv \left[\frac{F(\beta \mathscr{E}_{\max})}{4\pi} \beta a (\beta E_0)^2 \right]^{1/2} \tag{B40}$$

and

$$\operatorname{erf}(y) \equiv \int_0^y e^{-t^2} \,\mathrm{d}t .$$

For small values of y, we can further expand $I_n(y)$. To second order in y, E_0/a and $\beta E_0(\beta \mathscr{E}_{\max}$ and, hence, $F(\beta \mathscr{E}_{\max})$, being finite) we obtain

$$W(E) = \frac{a/\pi}{E^2 + a^2} \exp\left\{ -\beta \left[E + \frac{2a}{\pi} F(\beta \mathscr{E}_{\max}) \right] \right\} \left[1 - \frac{1}{3} \frac{E_0^2}{a^2} \frac{1 - 3E^2/a^2}{(1 + E^2/a^2)^2} - \Delta_1 \right],$$

where

$$\Delta_1 \equiv \frac{\beta^2 E_0^2}{6} \left\{ 1 - \frac{8}{15} \frac{E_0^2}{a^2} \left[\frac{1 - \frac{5}{2} \frac{E^2}{a^2} - \frac{13}{4} \frac{E^4}{a^4}}{(1 + E^2/a^2)^3} + \frac{\beta a}{8} \frac{1 - 3E^2/a^2}{(1 + E^2/a^2)^2} \right] \right\}.$$

References

Chandrasekhar, S.: 1943, *Rev. Mod. Phys.* **14**, 1.

Chandrasekhar, S. and Fermi, E.: 1953, *Astrophys. J.* **118**, 113.

Chapman, S. and Bartels, J.: 1962, *Geomagnetism*, Vol. II, Clarendon Press, Oxford, p. 639.

Herbst, W. and Assousa, G. E.: 1978, in T. Gehrels (ed.), *Protostars and Planets*, University of Arizona Press, Tucson, Arizona, p. 368.

Mestel, L.: 1972, in H. Y. Chiu and A. Muriel (eds.), *Stellar Evolution*, MIT Press, Cambridge, Massachusetts, p. 720.
Muschovias, T. C.: 1978, in T. Gehrels (ed.), *Protostars and Planets*, University of Arizona Press, Tucson, Arizona, p. 209.
Larson, R. B.: 1978, in T. Gehrels (ed.), *Protostars and Planets*, University of Arizona Press, Tucson, Arizona, p. 43.
Parker, E. N.: 1958, *Phys. Rev.* **109**, 1440.
Sweet, P. A.: 1958, in B. Lehnert (ed.), *Electromagnetic Phenomena in Cosmical Physics*, The University Press, Cambridge, p. 499.

MASS PICK-UP IN ASTRONOMICAL FLOWS*

T. W. HARTQUIST

Max-Planck Institute for Physics and Astrophysics, Institute for Extraterrestrial Physics, Garching, F.R.G.

and

J. E. DYSON

Department of Astronomy, The University, Manchester, England

(Received 17 September, 1987)

Abstract. Many astrophysical sources are clumpy. The flows in the diffuse media in these sources are affected by the mass picked up through the ablation of the embedded clumps. The diffuse media-clump interfaces can be observable sources themselves. Flows in clumpy stellar wind blown bubbles, supernova remnants, quasar winds, cooling gas accreting on to galaxies, and in pre-galactic bubbles blown by energy input from young galaxies, are discussed. Pick-up has important chemical as well as dynamical effects in star forming regions. The natures of the interfaces are considered.

1. Introduction

As is discussed in other papers in this issue, the Alfvén (1960) critical velocity mechanism for the ionization of gas has received wide application in studies of mass pick-up by the solar wind. In this review, we discuss mass loaded flows in a wide variety of astronomical sources outside of the solar system. Traditionally, theoreticians constructing gas dynamical models of astronomical sources have considered flows in smooth media only. Our goal is to illustrate that mass loading by the ablation of clumps often affects significantly the structures of the flows in the media in which they are embedded, and that the flow-clump interface regions can be observable interesting sources themselves.

In Section 2, we describe observations of shocked wind material in a Wolf–Rayet nebula, and suggest that they support the conjecture that slowly cooling mass-loaded flows are characterized in some statistical sense by a constant Mach number of about unity. In Section 3, we discuss simple models of the ablation of the clumps and observations which can be made to diagnose the nature of the interface between clumps and more diffuse flowing gas in several astronomical sources. Section 4 concerns the effects on supernova remnant evolution of mass loading by the dissipation of embedded clouds. The subject of Section 5 is the chemistry in regions in which low-mass star formation occurs; in such regions, molecular material condenses into clumps which are dissipated by the winds of the newly formed stars, but the ablated material condenses again to produce new clumps. Section 6 contains less detailed treatments of mass

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 615–627.

loading and clump ablation in other astrophysical sources including extragalactic flows and protostellar jets and material accreting on to large galaxies.

2. The Mean Value of the Mach Number in a Flow in RCW 58

RCW 58 is a nebula surrounding a Wolf–Rayet star which has a high speed $\approx 2000 \text{ km s}^{-1}$), high mass loss rate ($\dot{M} \approx 4 \times 10^{-5} M_\odot \text{ yr}^{-1}$) stellar wind which interacts with clumps of previously ejected slowly moving ($\lesssim 50 \text{ km s}^{-1}$) stellar material and with ambient interstellar gas. The central Wolf–Rayet star was used by Smith *et al.* (1984) as a background source against which optical and ultraviolet absorption features were detected. The velocities of the features were found to be correlated linearly with the ionization potentials of the species giving rise to their formation. The velocity of the C IV feature is about -150 km s^{-1} whilst the velocity of the Fe II feature is about -102 km s^{-1}.

If the temperature at which a species exists is taken to be proportional to its ionization potential (as is approximately true in a plasma in which the ionization structure has attained equilibrium), and if the flow region is isobaric and thin, a linear correlation between the ionization potentials of the species and the velocities of their features would be expected. Under such conditions, the temperature T and the magnitude of the velocity v, are related by $T = (\mu P/\mathbf{R}\rho_1 v_1)\ (v - v_0)$, where v_0 is a constant; μ, the mean mass in a.m.u.; P, the pressure; $\mathbf{R}$, the gas constant; and $\rho_1 v_1$, the mass flux through the region. Assuming that the absorbing C^{3+} is in gas which has a temperature of 1.2×10^5 K and the absorbing Fe^+ is in gas at 1×10^4 K, we can infer that $(\mu P/\mathbf{R}\rho_1 v_1) = -0.02\text{K s cm}^{-1}$. In cooled gas which has passed through a strong steady shock propagating into a uniform medium at rest, and behind which the mass and momentum fluxes are constant, $P = \rho_A v_S^2$ and $\rho_1 v_1 = \rho_A v_S$ where ρ_A is the pre-shock density and v_S is the shock speed. Taking $v_S = -2 \times 10^8 \text{ cm s}^{-1}$ (i.e., assuming that the shock is formed in the stellar wind), one obtains a value of roughly -1 K s cm^{-1} for the slope expected in such cooled post-shock gas.

The discrepancy between the expected slope of the correlation and the observationally determined slope led Smith *et al.* (1984) to suggest that the ablation of clumps embedded in the post-shock flow provides a source of mass which is injected into the flow. If the total mass ablation rate were to exceed the stellar mass loss rate by a factor of roughly fifty in the time during which the gas cools from its initial post-shock temperature to the temperature at which the detected ions exist, and if subsequent cooling is rapid compared to the rate at which additional mass is picked up, the inferred slope would be obtained. If the sources are at rest, mass loading of a steady adiabatic subsonic flow will lead to an increase in the Mach number (Hartquist *et al.*, 1986).

Hartquist *et al.* (1986) argued that the inferred slope would be expected if a sustained slowly cooling one-dimensional mass-loaded flow mimics, in some statistical sense, a steady flow of constant adiabatic Mach number of about unity. They showed that mass loading behind steady adiabatic shocks will result in a decrease of $P/\rho_1 v_1$ by a factor of $\frac{8}{25}$ from the initial postshock value of $(3\mu v_s/4\mathbf{R})$ as the post-shock Mach number increases to unity. In a mass loaded adiabatic flow behind a 2000 km s^{-1} shock, the temperature will be 2.5×10^7 K when the Mach number has risen close to unity. When the Mach number becomes unity, the flow will become unsteady; however, Hartquist *et al.* (1986) conjectured that on average, the Mach number remains around unity as the gas cools radiatively to about 1.5×10^5 K, the temperature at which the equilibrium cooling rate peaks (Gaetz and Salpeter, 1983). They assumed that the subsequent cooling time-scale is short relative to the mass injection time-scale. If these assumptions were valid, the slope of the temperature-velocity correlation would be $(\frac{3}{4})(\frac{8}{25})(1.5 \times 10^5/2.5 \times 10^7)^{1/2} = 0.019$ times that in cool post-shock gas which has not been mass loaded. The conjecture that the Mach number is unity on average throughout a substantial fraction of the post-shock mass loaded region is then in harmony with the observed ionization potential-velocity correlation.

3. The Natures of Interfaces

Section 2 concerned observational data which provide evidence that mass pick-up can affect a particular astronomical flow significantly. The interface regions between the clumps and flowing more diffuse gas themselves can be observable sources and their natures determine the injection rate and the composition of material picked up by the flow.

Considerable effort has been expended on the study of conductively driven mass loss (Penston and Brown, 1970; Graham and Langer, 1973; Castor *et al.*, 1975; Cowie and McKee, 1977; McKee and Cowie, 1977; Weaver *et al.*, 1977; Balbus and McKee, 1982; Draine and Giuliani, 1984; Balbus, 1986; Böhringer and Hartquist, 1987) from interstellar clouds into the hot phase of the interstellar medium. A conduction front model provides a good description of an interface between an unmagnetized cloud and a static, tenuous, unmagnetized surrounding medium. In the interstellar conductive interface models, the length scale over which the temperature varies sometimes becomes sufficiently short that the coefficient of conductivity is determined by particle-wave scattering processes as well as by particle-particle collisions; unfortunately, no reliable theory for calculating the coefficient of conductivity for such cases exists. If a cloud is sufficiently small or the external medium is sufficiently hot, radiative losses will not affect the structure of the conductive interface. If radiative losses are negligible, the evaporative flow is very subsonic everywhere and is steady, and if the coefficient of conductivity is determined by Coulomb collisions, then the conductively driven mass loss rate from a spherical cloud of radius R_0, embedded in a medium at a temperature T_∞ is approximately $\dot{M} \approx 2.75 \times 10^{19}$ $(T_\infty/10^6\ \mathrm{K})^{5/2}$ $(R_0/\mathrm{pc})\mathrm{g\ s}^{-1}$ (Cowie and McKee, 1977). Böhringer and Hartquist (1987) have constructed steady models of conductive interfaces which are modified by radiative losses and determined the mass loss rates; the non-equilibrium ionization structure is calculated for the interfaces in order to determine the radiative loss rates. The thermal stability of the interfaces which are modified by radiative losses has not been investigated; if they are unstable, steady models would be inappropriate.

In reality, tenuous astronomical media surrounding clouds are not static. The flows generally are turbulent and often are characterized by moderate to high Mach numbers. For the moment, we will continue to assume that the media are unmagnetized. In such circumstances, the mass loss rate is probably not determined by conductive processes, though they will play a rôle in the dissipation of small-scale structures which ultimately form during the hydrodynamically driven ablation of a cloud. Simkins (1971) has published photographs of the dispersal of drops embedded in a high Reynolds number, high Mach number, laboratory flow; when surface tension is unimportant, drops break up after expanding perpendicularly to the upstream flow direction and undergoing an umbrella-like deformation. The perpendicular expansion is due to the pressure decrease across the drop's surface due to the Bernoulli effect. Surface tension is never important in diffuse astrophysical plasmas and hydrodynamic ablation of any clumps which are not bound gravitationally and which are embedded in fast flowing diffuse media should occur following clump expansion perpendicular to the upstream flow direction. This occurs at roughly the clump internal sound speed if the flow, as observed in the clump rest frame, is supersonic, or at roughly the square of the Mach number of the flow times the clump internal sound speed if the flow is very subsonic (Hartquist *et al.*, 1986). During an expansion time, the development of various instabilities including the Kelvin–Helmholtz instability and the Rayleigh–Taylor instability should result in clump deformation and fragmentation. Numerical studies (Woodward, 1976; Nittmann *et al.*, 1982) in which the numerical viscosity was much higher than is ever appropriate for astrophysical flows and in which the evolution was not followed for such long periods, suggest that instabilities develop very rapidly compared to the expansion time-scales. Since clumps which are not bound gravitationally are ablated on a time-scale comparable to the greater of the sound crossing time and the sound crossing time divided by the square of the diffuse flow Mach number, it follows that ram pressure can accelerate a clump to at most about the speed of the slow shock or sound wave driven into it by the ram pressure before it is dissipated. Hence, mechanisms other than ram pressure acceleration (e.g., Hartquist and Dyson, 1987) must be responsible for the formation of clumps of high velocity cold material (e.g., high velocity interstellar water vapour masers).

While we have argued above that a clump which is not gravitationally bound will expand as a whole if it is embedded in a flowing medium, the actual momentum transfer between the flowing diffuse gas and the clump material may be confined to a boundary layer. The flows are almost certainly turbulent and a turbulent boundary layer should form. Typically, turbulent hydrodynamic boundary layers have thicknesses of the order of $L/\sqrt{R}$ where L is the length of the obstacle and R is the flow's Reynolds number. If the coefficient of molecular viscosity is used, $R \geq 10^6$ for many astronomical flows. However, the turbulence itself gives rise to an effective coefficient of viscosity which is estimated to be $\sim l_c\,\delta V$ where l_c is the correlation length and δV is the root-mean-square turbulent velocity. If one adopts $l_c\,\delta V$ as the coefficient of viscosity to calculate a 'turbulent' Reynolds number R_t, it becomes $R_t \approx VL/l_c\,\delta V$, where V is the upstream flow speed as measured in the obstacle's frame. The onset of turbulence often occurs for

Reynolds numbers of the order of 10^3. One would expect that R_t is not much less than 10^3 in a turbulent astrophysical medium; if it were, the turbulent viscosity would be sufficient to impede the development of the turbulence. One would also expect R_t not to exceed about 10^3; if it did, the medium should become more turbulent. Hence, we will assume that $R_t \sim 10^3$ in a boundary layer between flowing diffuse matter and a embedded clump. Taking the boundary layer thickness to be $\sim L/\sqrt{R}$ we find it to be about $0.03L$. Momentum transfer and acceleration of clump material may, therefore, be concentrated in this narrow layer.

If the clump and interclump media are magnetized and the flow is subsonic, the boundary layer may not become turbulent. Syrovatskii (see Section 53 of Landau and Lifschitz, 1960) analyzed the stability of the flow at a tangential discontinuity between two incompressible magnetized fluids of densities ρ_A and ρ_B. If the flow velocities and magnetic fields in the fluids on either side of the discontinuity are, respectively, $\mathbf{V}_A$ and $\mathbf{V}_B$, and $\mathbf{B}_A$ and $\mathbf{B}_B$, the flow in the boundary will be stable if

$$B_A^2 + B_B^2 > 4\pi\rho_{AB} V_{AB}^2$$

and

$$(\mathbf{B}_A \times \mathbf{B}_B)^2 \geq 2\pi\rho_A \{(\mathbf{B}_A \times \mathbf{V}_{AB})^2 + (\mathbf{B}_B \times \mathbf{V}_{AB})^2\} ;$$

where

$$\rho_{AB} = \rho_A \rho_B/(\rho_A + \rho_B) \quad \text{and} \quad \mathbf{V}_{AB} = \mathbf{V}_A - \mathbf{V}_B .$$

If turbulence is suppressed, as it will be in flows with low magnetosonic Mach numbers, the boundary layer may become very thin, but the pressure of the magnetic field will guarantee that momentum transfer from the flow in the diffuse medium will be distributed over the entire cloud. The presence of magnetic fields will not prevent clump expansion perpendicular to the upstream flow; however, the Mach number and the sound speed should be replaced by their appropriate magnetosonic counterparts in the estimates for the clump expansion time-scales. Magnetic Rayleigh–Taylor instability and, if the stability condition given above is not satisfied, magnetic Kelvin–Helmholtz instability, should be important for clump dissipation on time-scales shorter than, or comparable to, the expansion time-scale.

If the media are fully-ionized, the presence of a magnetic field will prevent mixing of material from the two media. However, in some astrophysical environments, including low-mass star-forming regions, where only the diffuse material will be highly ionized while the clump material may be weakly ionized, ambipolar diffusion across field lines may lead to mixing in the turbulent boundary layer between a highly-ionized medium and a weakly-ionized clump. We estimate the maximum ion number density which can be produced by mixing in a turbulent magnetic boundary layer. We assume that l_c is comparable to the thickness of the boundary layer. There, the ambipolar diffusion speed across turbulent 'eddies' is $V_D \approx W_B R_t^{1/2}/4\pi L\alpha n_i \rho$ where $W_B/4\pi$ is the perturbation field energy density (we assume that the turbulence is trans-Alfvénic or super-Alfvénic, otherwise W_B would be simply the square of the cloud field strength), α is the ion-neutral collision rate coefficient, n_i is the ion number density, and ρ is the mass density. We assume that $W_B/4\pi \approx (\delta V)^2$. Referring to our earlier argument that $VL/\delta V l_c \sim R_t$, one finds that $\delta V/V \sim R_t^{1/2}$. We require that the time for diffusion across an 'eddy' on the boundary layer be comparable to the time for the diffuse material to flow past the obstacle; $L/V \sim L/R_t^{1/2} V_D \sim L^2 \alpha n_i/V^2$. The maximum number density of ions in the turbulent boundary layer is then roughly

$$n_i \sim \frac{V}{\alpha L} \sim 1.4 \times 10^{-1}\,\mathrm{cm}^{-3} \left(\frac{V}{100\,\mathrm{km\,s}^{-1}}\right)\left(\frac{L}{0.1\,\mathrm{pc}}\right)^{-1}.$$

If the turbulent 'eddies' were about a factor of three smaller than the thickness of the boundary layer, then $\delta V/V \sim 3R_t^{1/2}$, W_B is nine times larger than we have assumed, and the maximum number density of ions would be 81 times the above estimate. Hence, it is difficult to estimate reliably the ion density in a turbulent magnetic boundary layer between a weakly ionized clump and a fully-ionized wind.

From the above discussion, one should conclude that the construction of a theory of the structures of interfaces between clumps and flowing diffuse media in astronomical sources is difficult. We now briefly consider attempts to use observational data to diagnose the structures of interfaces in three types of astronomical sources.

Much of the interest in models of conductive interfaces between material in different thermal phases was stimulated by the detection of O VI absorption features in the interstellar gas (Jenkins and Meloy, 1974; York, 1977; Jenkins, 1978). Böhringer and Hartquist (1987) have argued that the O^{5+} in conductive interfaces is in higher temperature gas than it would be if ionization equilibrium obtained and that the narrowness of many of the observed O VI features imply that they are not formed in conductive interfaces (see also Hartquist and Morfill, 1984). The O VI features may arise in the hot flowing gas itself and may not have any relationship to the boundary layers.

Smith *et al.* (1988) have studied the optical line emission from clumps embedded in the wind of the central star in RCW 58. They find that the Hα and [N II] emission features from the clumps are roughly 30 km s^{-1} broad. There is rather inconclusive evidence that the [O III] emission features are 5 to 10 km s^{-1} broader. The centroid velocities of the [O III] features are not shifted significantly relative to the centroid velocities of the (N II] features, and if the [O III] features are more broadened because the fraction of the [O III] emission arising in the turbulent boundary layers is greater than the fractions of Hα and [N II] emission coming from them, the [O III] regions are confined to the inner parts of the boundary layers where acceleration by momentum transfer from the wind has not occurred. We note that if $\delta V \approx V/R_t^{1/2}$ and $V \approx 2000$ km s^{-1} as is appropriate for the wind of a Wolf–Rayet central star, $\delta V = 30$ km s^{-1} if $R_t \approx 4500$.

Finally, B5 is a small (~ 2 pc), nearby (~ 100 pc) molecular cloud containing four compact infra-red sources which are probably associated with T-Tauri stars. The winds of the stars interact with clumps in the cloud. The interface regions in B5 may contain observable enhanced abundances of particular molecular species which could serve as useful interface diagnostics and may be observable sources of H_2 IR line emission. We discuss B5 more fully in Section 5.

We have already noted consequences of mass pick-up on the structure of the flow observed in RCW 58. We now describe other astrophysical flows where substantial modification occurs due to its process.

4. Mass-Loaded Supernovae

The local soft X-ray background has been interpreted in terms of the emission from a hot supernova heated phase of the interstellar medium (Cox and Anderson, 1982; Innes and Hartquist, 1984). Cox and Anderson (1982) performed a fully time-dependent calculation of the ionization structure and emission of an adiabatic supernova remnant expanding into a uniform medium and found that the observed B- and C-band emissions could be produced in such a remnant formed by the occurrence of a supernova 10^5 yr ago. The explosion energy is 5×10^{50} erg and the ambient density is 4×10^{-3} cm^{-3}. This particular model, however, cannot produce the more energetic M-band emission, though it is debatable whether or not Cox and Anderson (1982) explored sufficient parameter space to rule out similar models entirely. Innes and Hartquist (1984) noted that the emitted soft X-ray spectrum of a remnant hardens with time (because the hottest gas cools more slowly) and found that a local old ($\approx 4 \times 10^6$ yr) superbubble produced by the injection of 4×10^{52} erg into an ambient medium ($n \approx 4$ cm^{-3}) could produce the required M-band emission and would probably contain O^{+5} at a density similar to that observed.

McKee and Ostriker (1977), Chièze and Lazareff (1981), and Cowie *et al.* (1981) suggested that mass pick-up from embedded clumps can significantly affect the structure of supernova remnants. Dyson and Hartquist (1987) have compared the structures of adiabatic remnants dominated by mass pick-up from embedded clumps for the cases where the mass loss is driven by conduction (Chièze and Lazareff, 1981) and by hydrodynamic ablation (Hartquist *et al.*, 1986). They attempted to explain the origin of the soft X-ray background by postulating that the Sun sits in such a remnant whose structure has been modified by mass pick-up.

The density distributions in both conduction dominated and ablation dominated remnants are similar: the gas density is very low behind the shock, increases to a maximum at some interior radius, and decreases inwards towards the centre. In the standard Sedov solution, the gas density decreases monotonically inward. The temperature in an ablation dominated remnant increases inwards, as in the Sedov solution, but unlike that in the conduction dominated remnant in which the temperature decreases inwards.

Dyson and Hartquist (1987) assumed ionization equilibrium in their calculation of the soft X-ray emission, although the thermal and ionization history of mass injected into the interior of a supernova remnant is quite unknown. In the specific case of Mach number dependent ablation discussed by Hartquist *et al.* (1986), the ablated material may flow along a hot dense boundary layer until it reaches the rear of the clump. The density, flow time-scale and thermal structure might then ensure that the injected material is highly ionized before it becomes part of the general more tenuous flow in which the ionization time-scale is much larger. They found that the relative M, B, and C band count rates can be well fitted by an ablation-dominated remnant with an immediate post-shock temperature $T_0 \approx 8 \times 10^5$ K. However, the bulk of the emission comes from the region of the density maximum where the temperature is about a factor two higher.

Conduction-dominated remnants with $T_0 \approx 3.5 \times 10^6$ K also fit these relative count rates. The global parameters of the remnants themselves are all very similar. The models also give O^{+5} column densities consistent with observational data although these data may only give upper bounds to these densities.

Dyson and Hartquist (1987) concluded, therefore, that it was not possible to advance a unique model to describe the local remnant. Since it appears that it is not possible to make an unambiguous determination of the structure of the local bubble from the soft X-ray data, it follows by extension that similar data cannot unambiguously determine realistic remnant evolution and the global structure of the interstellar medium. Further spectral discriminants (e.g., UV and EUV) are required to distinguish between models in which mass injection occurs. This then returns to the question of the physical structures of interface regions.

5. Dynamics and Chemistry in Low-Mass Star-Forming Regions

T-Tauri stars, which are pre-Main-Sequence low-mass stars, have mass loss rates which may be as high as $10^{-7} M_\odot \, \mathrm{yr}^{-1}$ and wind terminal speeds of up to $400 \, \mathrm{km \, s}^{-1}$. Norman and Silk (1980) have argued that in regions of low-mass star formation, T-Tauri wind blown bubbles drive shells which break into fragments which are destroyed by the winds of subsequent T-Tauri stars. Star formation, therefore, occurs in clumpy media and the wind blown bubble flows must be affected by mass loading.

Charnley *et al.* (1987) have studied the non-equilibrium chemistry in star forming regions. They assumed that material in clumps is ablated by, and well-mixed with, stellar wind material. The accretion of molecules containing heavy elements on to the surfaces of grains in the clumps leads to a very low abundance of heavy elements in the gas phase until the ablated gas passes through a bow shock around another clump or through a shock just inside the outer shell of swept-up gas; sputtering in such shocks returns heavy elements to the gas phase. The presence of H^+ and of He^+ ions from the stellar wind ensures that post-shock molecules bearing heavy elements are dissociated and that the heavy elements are primarily in gas phase atomic species in the cooled post-shock gas. As time passes, gas phase ion-molecule chemistry removes the H^+ and He^+ and forms molecules containing heavy elements. On somewhat longer time-scales, accretion of heavy elements on to dust grains becomes important in the cooled gas. The shell fragments and accretion on to grains continues. Williams and Hartquist (1984) and Charnley *et al.* (1987) suggest that such chemical-dynamical recycling in star forming regions may be responsible for maintaining high gas phase abundances of heavy elements (even though accretion on to grains occurs on time-scales shorter than a cloud lifetime) and high gas phase abundances of atomic species.

Goldsmith *et al.* (1986) have argued that in the nearby star forming region B5, such chemical-dynamical mixing is occurring. Charnley *et al.* (1988a) are modelling the non-equilibrium chemistry in B5. They consider two cycling scenarios, one of which was suggested by Goldsmith *et al.* (1986) and repeats on a time-scale somewhat less than 10^6 yr, and one of which repeats on a time-scale about an order of magnitude longer. They have followed the chemical evolution through several cycles and determined that in either scenario a chemical limit cycle obtains; is thus, the chemical composition a useful diagnostic of the cycling time-scale.

Because B5 is so close, the structures of the wind-clump interfaces in it can, in principle, be studied by observing H_2 IR emission and millimetre and submillimetre emission from atomic and molecular species. Charnley *et al.* (1988b) are constructing models of the thermal and chemical structures in the interfaces. They assume that the interfaces are isobaric and that heating and mixing occur at uniform rates per unit volume in a thin boundary layer. Cooling is due to radiative losses and adiabatic expansion. A comparison of model predictions and observational results should provide insight into the nature of these interfaces.

6. Other Mass Loaded Astronomical Flows

Sections 2 and 3 contain discussions of diagnostic studies of the mass loaded flow and of the clump-flowing diffuse medium interfaces in RCW 58. In Sections 3 and 5, we described potential observations of B5 which could be used to investigate the interface structures in clumpy low mass star forming regions. Section 4 describes the global structure of supernova remnants dominated by mass pick-up. We conclude this review with brief descriptions of the possible importance of mass loading in several other astronomical environments.

Herbig–Haro (HH) objects are observed as localized patches of emission line nebulosity against dark molecular clouds. They have masses comparable to that of the Earth and typical sizes $\sim 10^3$ AU. They are generally found located within about 0.1 pc of low-luminosity newly-formed stars (or IR sources) and are often associated with extended bipolar molecular flows around these young stars (e.g., Lada, 1985). It is generally accepted that HH objects are shock excited and produced by the interaction of stellar winds with their surroundings (e.g., Dyson, 1987). HH objects are often distributed in a strikingly collinear fashion (e.g., HH 46–47; HH 7–11) suggesting that the flows which form them are highly collimated. If the flows are continuous jets, their interactions with the ambient cloud along the quasi-cylindrical interfaces should lead to the entrainment of colder, lowly-ionized material and possibly to mixing and enhanced cooling in the flowing gas. Norman and Silk (1979) and recently Hartigan *et al.* (1987) have discussed the possibility that the Herbig–Haro emission often, but not always, originates in ambient cloud material which has passed through bow shocks around supersonic 'bullets' formed in the vicinity of the stars. Hartigan *et al.* have shown that bow shock models are consistent with the optical emission line strengths and profiles of a number of Herbig–Haro objects; they did not include the possibility that lowly-ionized material in the bullet could mix with the shocked ambient cloud material. However, such models in their simplest form are not compatible with the data for the extremely low excitation objects HH 7–11 which have line widths of nearly 120 km s^{-1} but emission line strengths which are in ratios that are characteristic of 20–30 km s^{-1} shocks (e.g., Böhm, 1983). Mixing of gas across of the interface between the bullet and shocked ambient medium could lead to the emission from a relatively strong shock to be dominated by lines of lowly excited species. Dyson (1984) pointed out that a jet interacting with an ambient medium will produce a spectrum similar to that originating in gas which has passed through a bow shock; mixing of shocked decelerated jet gas and gas from the swept-up ambient medium could create the low excitation objects.

Meaburn and Dyson (1987) have discussed the jet-like feature HH 47B which connects HH 46 and 47A. The Hα line profile over this feature is broad whereas the associated [S II] line exhibits distinct splitting over the same velocity range. They attribute this to the entrainment and mixing in of ambient gas by high-speed ionized jet material at a boundary layer.

Shapiro and Field (1976), Bregman (1980), and Kahn (1981) have described thermally-driven fountain models for the coronal gas in the Galactic halo. Thermal instability in the fountains may give rise to the formation of high velocity clouds. Hartquist and Morfill (1986) have summarized the arguments based on observations for favouring models of the structure of the halo coronal gas in which cosmic-ray pressure greatly exceeds the thermal and kinematic pressures, but fountains could, in principle, be driven by cosmic-ray pressure. Clouds formed following the onset of thermal instability, the development of which may be modified but possibly not suppressed by the presence of cosmic rays, begin to fall back towards the galactic disk, and their motion relative to the upwards flowing gas should lead to their ablation. We can estimate the maximum velocity difference between upwards hot gas and cool clouds that can be attained before they are ablated. The magnitude of the velocity difference will be roughly gt_s where g is the galactic gravitational acceleration and t_s is the approximate survival time which we estimate to be $\sim \mathrm{P}_{\mathrm{cr}} R_c/\rho_{\mathrm{H}} g^2 t_s^2 C_c$ where P_{cr} is the cosmic-ray pressure, ρ_{H} is the density in the hot gas, R_c is the initial cloud radius, and C_c is the signal speed in the cloud. The maximum relative speed is then $\sim (gP_{\mathrm{cr}} R_c/\rho_{\mathrm{H}} C_c)^{1/3}$. Taking $P_{\mathrm{cr}}/\rho \sim 10^{15}$ cm^2 s^{-2}, $C_c \sim 10^6$ cm s^{-1}, $g \sim 10^{-8}$ cm s^{-1}, and $R_c \sim 100$ pc, one finds that large clouds can reach speeds of about 1.4×10^7 km s^{-1} relative to the coronal gas; such speeds are comparable to those of many observed high-velocity clouds which are at unknown distances and hence have unknown sizes.

The infall of radiatively cooling gas on to the central galaxies of galaxy clusters results in X-ray emission and the formation of optical filaments (e.g., Fabian *et al.*, 1984; Sarazin, 1986). The formation of optical filaments through the development of thermal instability is problematic if thermal conduction occurs at its classical rate (Magnoli *et al.*, 1987). However, Fabian *et al.*, have suggested that filament formation through thermal instability may be particularly important as an inflow accelerates and cools and approaches the sonic point; they speculated that the existence of effective mass sinks in an inflow may prevent it from reaching the sonic point.

The larger cluster-scale parts of such a flow may be affected by mass sources. The existence of iron X-ray emission features from cluster gas implies that intracluster gas in many sources has iron fractional abundances which are comparable to the solar values (e.g., Sarazin, 1986). The flow must pick-up mass which is either ejected by galaxies or is ablated or stripped from them.

Galactic winds can inject material into a cluster flow. In the case of, for example, starburst galaxies, winds can be driven by supernovae energy injection (Chevalier and Clegg, 1985). Observational evidence of such large scale flows comes from the optical data of McCarthy *et al.* (1987). Optical filaments are supposedly produced when the supersonic supernova-driven wind impacts on interstellar clouds. These wind models ignore the wind mass-loading from these clouds and the possibility that the optical emission may come from the wind-cloud interfaces.

Galaxies with active nuclei may have winds driven by the strong radiation source at their centres. Dyson (1983) proposed that the narrow emission lines originated in wind-cloud interactions. Again such winds will be mass loaded and interfaces will occur. Quasar winds mass loaded by supernovae and stellar winds but driven by momentum transfer from the central radiation field have been suggested by Perry and Dyson (1985) as the site of broad emission line cloud formation by Compton cooling behind shocks induced into the wind by the mass loading process itself. In addition to the complex topology of these mass loaded flows, there is the interesting question of abundance variations produced when very heavy element enriched ejecta is added to a less enriched overall flow.

Ostriker and Cowie (1981) pointed out that the intergalactic medium during and somewhat after the epoch of galaxy formation probably had a structure similar to that of the present galactic interstellar medium. The formation of energy sources including the galaxies and possibly pre-galactic stars, resulted in the production of blast waves, bounded by shells of swept-up material, which may overlap before cooling. Such an intergalactic medium will become clumpy just as is the interstellar medium. Ikeuchi and Ostriker (1986) have argued that though the gravitational instability of massive shells driven by very energetic sources will lead to the formation of galaxies, lower mass shells are gravitational stable but may fragment due to Rayleigh–Taylor instability and form the metal poor clouds which give rise to the 'forest' of QSO Lα absorption features for which no corresponding absorption features due to heavy elements are found. They suggested that these clouds are confined by the pressure of the hotter phase of the intergalactic medium and will survive to the present epoch if the time-scale for them to be evaporated by conductively-driven heating is longer than the Hubble time. We would argue that the survival of a Lα absorbing cloud also depends on the nature of its motion relative to the more diffuse intergalactic gas in its vicinity, but, in any case, even the intergalactic medium during the galaxy formation epoch is likely to contain flows in a diffuse medium which are modified by the pick-up of material from embedded clouds.

7. Conclusions

Mass pick-up can appreciably modify the global structure of a very wide variety of astrophysical flows. The interface zones between the flowing medium and the embedded clumps can themselves be observable sources. Since the mass pick-up occurs on a wide variety of scales, detailed diagnostic observational studies of nearby astronomical sources in which mass pick-up is likely to be occurring are highly desirable. We would here particularly stress B5 as a possible source for such diagnostic studies. The elucidation of the interface structures is theoretically very complex and little progress is likely without these complementary observational studies. Nevertheless we would suggest that the general subject of mass loaded flows should be and will become a major area of investigation in astronomy.

Acknowledgement

TWH holds an SERC Visiting Fellowship. E. Knobloch in 1980 introduced us to the argument (which he attributed to D. Lynden-Bell) that, in fully developed turbulent flow, the Reynolds number, calculated when one adopts an effective turbulent viscosity, is of order of 10^3.

References

Alfvén, H.: 1960, *Rev. Mod. Phys.* **32**, 310.
Balbus, S. A.: 1986, *Astrophys. J.* **304**, 787.
Balbus, S. A. and McKee, C. F.: 1982, *Astrophys. J.* **252**, 529.
Böhm, K. H.: 1983, *Rev. Mexicana Astron. Astrophys.* **7**, 55.
Böhringer, H. and Hartquist, T. W.: 1987, *Monthly Notices Roy. Astron. Soc.* **228**, 915.
Bregman, J. N.: 1980, *Astrophys. J.* **236**, 577.
Castor, J., McCray, R., and Weaver, R.: 1975, *Astrophys. J.* **200**, L107.
Charnley, S. B., Dyson, J. E., Hartquist, T. W., and Williams, D. A.: 1987, *Monthly Notices Roy. Astron. Soc.*, (in press).
Charnley, S. B., Dyson, J. E., Hartquist, T. W., and Williams, D. A.: 1988a, in preparation.
Charnley, S. B., Dyson, J. E., Hartquist, T. W., and Williams, D. A.: 1988b, in preparation.
Chevalier, R. and Clegg, A.: 1985, *Nature* **317**, 44.
Chièze, J. P. and Lazareff, B.: 1981, *Astron. Astrophys.* **95**, 194.
Cowie, L. L. and McKee, C. F.: 1977, *Astrophys. J.* **211**, 135.
Cowie, L. L., McKee, C. F., and Ostriker, J. P.: 1981, *Astrophys. J.* **247**, 908.
Cox, D. P. and Anderson, P. R.: 1982, *Astrophys. J.* **253**, 268.
Draine, B. T. and Giuliani, J. L., Jr.: 1984, *Astrophys. J.* **281**, 690.
Dyson, J. E.: 1983, in R. M. West (ed.), *Highlights of Astronomy*, International Astronomical Union.
Dyson, J. E.: 1984, in P. M. Gonhalekar (ed.), *Proc. Rutherford–Appleton Laboratory Workshop on 'Gas in the Interstellar Medium'*, RAL Publications, Chilton, England, p. 107.
Dyson, J. E.: 1987, 'Circumstellar Matter', in I. Appenzeller and C. Jordan (eds.), *IAU Symp.* **122**, 159.
Dyson, J. E. and Hartquist, T. W.: 1987, *Monthly Notices Roy. Astron. Soc.* **228**, 353.
Fabian, A. C., Nulsen, P. E. J., and Canizares, C. R.: 1984, *Nature* **310**, 733.
Gaetz, T. J. and Salpeter, E. E.: 1983, *Astrophys. J. Suppl.* **52**, 155.
Goldsmith, P. F., Langer, W. D., and Wilson, R. W.: 1986, *Astrophys. J.* **303**, L11.
Graham, R. and Langer, W. D.: 1973, *Astrophys. J.* **177**, 605.
Hartigan, P. Raymond, J., and Hartmann, L.: 1987, *Astrophys. J.* **316**, 323.
Hartquist, T. W. and Dyson, J. E.: 1987, *Monthly Notices Roy. Astron. Soc.* **228**, 957.

Hartquist, T. W. and Morfill, G. E.: 1984, *Astrophys. J.* **287**, 194.
Hartquist, T. W. and Morfill, G. E.: 1986, *Astrophys. J.* **311**, 518.
Hartquist, T. W., Dyson, J. E., Pettini, M., and Smith, L. J.: 1986, *Monthly Notices Roy. Astron. Soc.* **221**,715.
Ikeuchi, S. and Ostriker, J. P.: 1986, *Astrophys. J.* **301**, 522.
Innes, D. E. and Hartquist, T. W.: 1984, *Monthly Notices Roy. Astron. Soc.* **209**, 7.
Jenkins, E. B.: 1978, *Astrophys. J.* **219**, 845.
Jenkins, E. B. and Meloy, D. A.: 1974, *Astrophys. J.* **193**, L121.
Kahn, F. D. 1981, in F. D. Kahn (ed.), *Investigating the Universe*, D. Reidel Publ. Co., Dordrecht, Holland, p. 1.
Lada, C. J.: 1985, *Ann. Rev. Astron. Astrophys.* **23**, 276.
Landau, L. D. and Lifschitz, E. M.: 1960, *Electrodynamics of Continuous Media*, Pergamon Press, Oxford.
Magnali, A., Rosner, R., and Bodo, G.: 1987, *Astrophys. J.* (in press).
McCarthy, P., Heckman, T., and van Breugel, W.: 1987, *Astron. J.* **92**, 264.
McKee, C. F. and Ostriker, J. P.: 1977, *Astrophys. J.* **215**, 213.
Meaburn, J. and Dyson, J. E.: 1987, *Monthly Notices Roy. Astron. Soc.* **225**, 863.
Nittmann, J., Falle, S. A. E. G., and Gaskell, P. H.: 1982, *Monthly Notices Roy. Astron. Soc.* **201**, 833.
Norman, C. A. and Silk, J.: 1979, *Astrophys. J.* **228**, 197.
Norman, C. A. and Silk, J.: 1980, *Astrophys. J.* **238**, 138.
Ostriker, J. P. and Cowie, L. L.: 1981, *Astrophys. J.* **243**, L127.
Penston, M. V. and Brown, F. E.: 1970, *Monthly Notices Roy. Astron. Soc.* **150**, 373.
Perry, J. J. and Dyson, J. E.: 1985, *Monthly Notices Roy. Astron. Soc.* **213**, 665.
Sarazin, C. L.: 1986, *Rev. Mod. Phys.* **58**, 1.
Shapiro, P. R. and Field, G. B.: 1976, *Astrophys. J.* **205**, 762.
Simkins, P. G.: 1971, Nature: *Phys. Sci.* **233**, 31.
Smith, L. J., Pettini, M., Dyson, J. E., and Hartquist, T. W.: 1984, *Monthly Notices Roy. Astron. Soc.* **211**, 679.
Smith, L. J., Pettini, M., Dyson, J. E., and Hartquist, T. W.: 1988, *Monthly Notices Roy. Astron. Soc.*, (in press).
Weaver, R., McCray, R., Castor, J., Shapiro, P., and Moore, R.: 1977, *Astrophys. J.* **218**, 377.
Woodward, P. R.: 1976, *Astrophys. J.* **196**, 555.
York, D. G.: 1977, *Astrophys. J.* **213**, 43.

INFRARED EMISSION FROM INTERSTELLAR PLASMA*

SATIO HAYAKAWA

Department of Astrophysics, Nagoya University, Japan

(Received 13 July, 1987)

Abstract. Small dust grains in hot interstellar plasmas are heated by electrons and emit near-infrared radiation. Its flux is estimated to be larger than that from small dust grains heated by stellar radiation and could explain the high infrared flux observed at wavelengths of about 5 μm.

1. Introduction

Alfvén (see, e.g., Alfvén, 1950) introduced the concept of a plasma in which magnetic fields are frozen in. Such plasmas fill the Universe and are an important part of various astrophysical phenomena. In interstellar space a large fraction of volume is filled with hot plasmas which emit X-rays. Such plasmas have temperatures of a few hundreds of eV and densities of 10^{-2} cm^{-3} or lower, thus being responsible for the diffuse component of soft X-rays (Tanaka and Bleeker, 1977). In the inner region of the Galaxy there exist still hotter instellar plasmas with temperatures of several keV (Koyama *et al.*, 1986). These plasmas have been detected mainly through X-rays and also through measurements of the spectral lines of highly ionized ions.

In the present paper we investigate a different effect of such hot plasmas, that is, infrared emission of dust heated by hot particles. It has been generally accepted that far-infrared radiation (FIR) of wavelengths ~ 100 μm in the Galaxy is emitted from dust heated by starlight. A high flux of near-infrared radiation (NIR) of wavelengths $\lesssim 10$ μm has been interpreted as due to the emission from small dust grains (SDG) which are heated to considerably high temperatures by the absorption of a single ultraviolet photon (Cox *et al.*, 1986), but the NIR flux expected from SDGs is found to be much weaker than the FIR flux (Draine and Anderson, 1985).

It has been recently pointed out by Dwek (1986) that SDGs heated by hot electrons are effective in the emission of NIR. Assuming a grain size distribution of $\mathrm{d}n(a)/\mathrm{d}a \propto a^{-3.5}$ for $a_{\min} < a < a_{\max}$, where a is the grain radius, he has calculated the infrared spectra for several values of $a_{\min}$. The flux νI_ν at 5 μm obtained for $a_{\min} = 5$ Å is found to be about 1/1.5 of that of 100 μm. This is compared with the ratio of about 1.5 observed by Matsumoto *et al.* (1987, 1988).

In the present paper I recalculate the infrared flux with a somewhat different mechanism of the energy transfer from electrons to SDGs and give the absolute flux value. The calculated result is in approximate agreement with the observed one. This suggests that the observations of NIR and soft X-rays provide an important means of investigating the interstellar medium.

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 629–633.

2. Infrared Spectrum

The spectrum of infrared radiation emitted from interstellar dust has been calculated by Draine and Anderson (1985) with the following model. Interstellar dust is a mixture of graphite and silicate grains with a size distribution $dn(a)/da \propto a^{-3.5}$ for $a_b < a < a_{max}$ and $\propto a_b^{0.5} a^{-4}$ for $a_{min} < a < a_b$. Using the radiation field derived by Mathis *et al.* (1983) and the optical properties of dust grains obtained by Draine and Lee (1984), they calculated the spectra for (i) $a_{max} = 0.25$ μm, $a_b = 30$ Å, and $a_{min} = 3$ Å, and (ii) $a_{max} = 0.25$ μm and $a_{min} = a_b = 3, 10, 30$, and 100 Å, as reproduced in Figure 1. The spectrum for (i) gives a higher NIR flux because of a dominant contribution of SDGs but the flux at 5 μm is about $\frac{1}{30}$ times that at 100 μm.

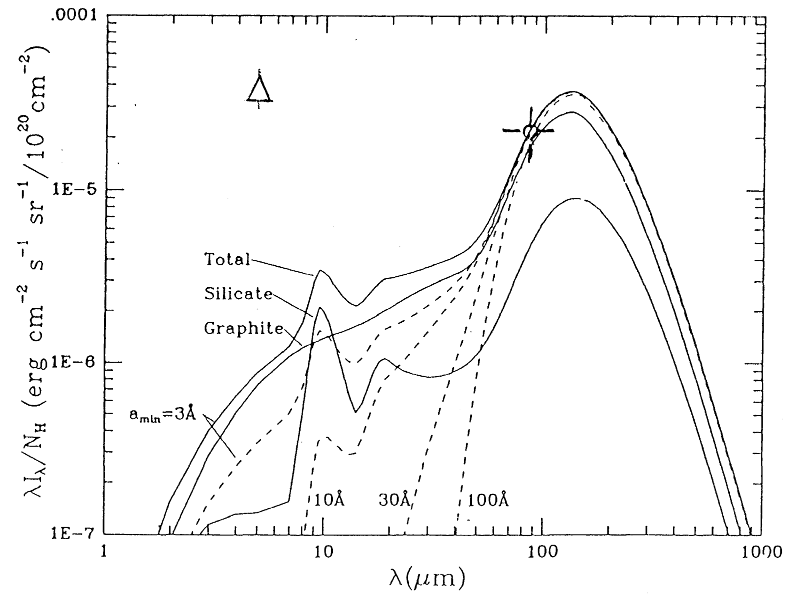


Fig. 1. Emission spectra for mixtures of graphite and silicate grains heated by the average interstellar field (adopted from Draine and Anderson, 1985). The solid curves represent those of $a_{min} = 3$ Å and $a_b = 30$ Å, and the dashed curves represent those for $a_{min} = a_b = 3$, 10, 30, and 100 Å. Symbols Δ and ○ represent the fluxes observed by Matsumoto *et al.* (1987, 1988) normalized to the model curve at $\lambda = 80$ μm, corresponding to $N_H = 8 \times 10^{20}$ cm^{-2}.

In comparison with this model spectrum we give the absolute fluxes observed at galactic latitudes of about 40° by Matsumoto *et al.* (1987, 1988) at about 100 and 5 μm in Figure 1 (open circle and triangle, respectively). The FIR flux varies with view direction depending on the hydrogen column density but is independent of the heliographic coordinate. The 5 μm flux is proportional to cosec b, where b is the galactic latitude, but does not appreciably depend on the heliographic coordinate. Hence, these fluxes are those of interstellar origin without appreciable contributions of the zodiacal, starlight, and extragalactic components. Note that these fluxes are the absolute values and are different from the excess fluxes in the galactic plane, the difference between the on-plane and off-plane fluxes, as summarized by Cox *et al.* (1986). The observed flux at 5 μm is higher by two orders of magnitude than the calculated one, if normalized at 80 μm. The calculated flux at 5 μm would increases roughly in proportion to $a_b^{0.5}$ if a_b were increased, but would still be much lower than the observed flux.

3. Heating of DSG by Radiation and Hot Electrons

In their model Draine and Anderson (1985) have assumed that dust grains absorb radiation with a cross-section $\pi a^2 Q(a, \lambda)$. According to Draine and Lee (1984), radiation in the wavelength range 1000–4000 Å is most effective for absorption, and the value of $Q(a, \lambda)$ averaged over the radiation spectrum is approximately represented by $Q(a) = 0.3(a/100\ \text{Å})$ for $a < 100$ Å. The radiation flux in this range is about $4 \times 10^{-3}\ \text{erg cm}^{-2}\ \text{s}^{-1}$. Hence, a grain absorbs radiation energy at a rate of $4 \times 10^{-18}(a/10\ \text{Å})^3\ \text{erg s}^{-1}$.

If dust grains are embedded in a plasma of electron temperature T and density n, electrons transfer their kinetic energies by the excitation and ionization of constituent atoms. For graphite grains the L-shell electrons of carbon atoms are responsible for the absorption of electron energies, since the K-shell electrons have too high a binding energy to be excited by electrons with a temperature of about 100 eV. The rate of energy transfer by an electron of velocity v to a carbon atom is given by

$$\left.\frac{\mathrm{d}E}{\mathrm{d}t}\right|_c = \frac{4\pi Z_L e^4}{mv^2}\, nv \left[\ln\left(\frac{mv^2}{I}\right) + 0.9 - \left(\frac{v}{c}\right)^2\right] \simeq$$

$$\simeq 1.2 \times 10^{-19}\left(\frac{n}{10^{-2}\ \text{cm}^{-3}}\right)\left(\frac{100\ \text{eV}}{T}\right)^{1/2}\ \text{erg s}^{-1}\,, \tag{1}$$

where Z_L is the number of electrons in the L shell, $I \simeq 50$ eV is the average ionization potential and other quantities have their usual meaning.

Since an SDG contains a small number of atoms (say $\sim 10^2$) most atoms are located on the grain surface and are exposed to impinging electrons. Hence, the rate of energy transfer to a grain is approximately given by a sum of the contributions of individual atoms. Hence, a graphite grain will absorb energy from electrons at a rate of

$$\left.\frac{\mathrm{d}E}{\mathrm{d}t}\right|_g \simeq \left.\frac{\mathrm{d}E}{\mathrm{d}t}\right|_c N_c \simeq 5 \times 10^{-17}\left(\frac{a}{10\ \text{Å}}\right)^3\left(\frac{n}{10^{-2}\ \text{cm}^{-3}}\right)\left(\frac{100\ \text{eV}}{T}\right)^{1/2}\ \text{erg s}^{-1}\,, \tag{2}$$

where $N_c = (4\pi/3)a^3\rho_g/m_c \simeq 4 \times 10^2(a/10\ \text{Å})^3$ is the number of carbon atoms in a graphite grain, $\rho_g \simeq 2\ \text{g cm}^{-3}$ is the density of graphite and m_c the mass of a carbon atom. The energy absorption rate in Equation (2) is more than ten times greater than the absorption rate of radiation energy estimated by Draine and Anderson (1985). Since NIR is emitted mainly from graphite grains with $a \lesssim 5$ Å, according to Draine and Anderson (1985), the NIR thus flux emitted from SDGs is much stronger than that calculated for radiation-heated SDGs. If the contributions of hydrogen atoms adsorbed by the grain and of the secondary excitation by electrons left after the primary ionizing collision are taken into account, the rate of energy absorption may be still greater.

It should be noted that the electron heating contributes little to the FIR flux. Since large dust grains are mainly responsible for FIR emission, and since the energy flux of radiation over a wider wavelength range is responsible for the absorption with $Q \sim 1$, the radiation heating is more efficient than the electron heating. Hence, the FIR flux estimated on the basis of radiation heating is little modified.

4. Discussion

I have shown that hot electrons are more effective than ultraviolet radiation for heating SDGs to emit NIR and may explain the high NIR flux observed. However, the result should not be taken quantitatively, since the microscopic structure of SDG is not properly taken into account. A grain of $a \sim$ several Å contains only tens of atoms, and its shape may be planar. The absorption of energy takes place mainly by electronic transitions, for which the effect of atomic binding may be of minor importance. The transfer of energy from the excited electron to internal modes depends on the grain structure, and the Debye model has to be modified so as to account for the emission bands as observed. Nevertheless, the essential result of the present paper remains to be valid, unless one intends to obtain the result accurate within a factor of two or so.

One might ask if SDGs could survive in a hot plasma. The lifetime of graphite grains against sputtering is estimated by Barlow and Silk (1977) to be

$$t_g \simeq 2 \times 10^6 (a/10\ \text{Å})\,(10^{-2}\ \text{cm}^{-3}/n)\,(100\ \text{eV}/T)\ \text{yr}\,. \tag{3}$$

This is longer than the lifetime of a hot plasma bubble which is presumably produced by a supernova blast wave.

The present model can be tested by the correlation between the NIR and soft X-ray fluxes. Since the NIR flux is proportional to $nn(\text{dust})R$, whereas the soft X-ray flux is proportional to n^2R, where R is the linear dimension of a hot plasma along the line-of-sight, the correlation between these two is considered to be somewhat weak but to be observable. It should be noted that soft X-rays are relatively strong in the region scanned by the NIR observation (Matsumoto *et al.*, 1987). In this region nonthermal radio emission is relatively weak, and this has been interpreted as due to the effect of a hot plasma to sweep out magnetic fields (Hayakawa, 1979).

Acknowledgement

It is a great pleasure to dedicate the present paper to Professor Hannes Alfvén who introduced me to astrophysics through his excellent books, stimulating papers on the origin of cosmic rays, and personal discussions.

References

Alfvén, H.: 1950, *Cosmical Electrodynamics*, Clarendon Press, Oxford.

Barlow, M. and Silk, J.: 1977, *Astrophys. J.* **211**, L83.

Cox, P., Krugel, E., and Mezger, P. G.: 1986, *Astron. Astrophys.* **155**, 380.

Draine, B. T. and Anderson, N.: 1985, *Astrophys. J.* **292**, 494.

Draine, B. T. and Lee, H. M.: 1984, *Astrophys. J.* **258**, 89.

Dwek, E.: 1986, *Astrophys. J.* **302**, 363.

Hayakawa, S.: 1979, *Publ. Astron. Soc. Japan* **31**, 621.

Koyama, K., Makishima, K., Tanaka, Y., and Tsunemi, H.: 1986, *Publ. Astron. Soc. Japan* **38**, 121.

Mathis, J. S., Mezger, P. G., and Panagna, N.: 1983, *Astron. Astrophys.* **128**, 212.

Matsumoto, T., Akiba, M., and Murakami, H.: 1987, *Astrophys. J.* (submitted).

Matsumoto, T., Hayakawa, S., Matsuo, H., Murakami, H., Lange, A. E., and Richards, P. L.: 1988, *Astrophys. J.*

Tanaka, Y. and Bleeker, J. A. M.: 1977, *Space Sci. Rev.* **20**, 815.

ON THE FLUCTUATIONS OF THE SURFACE BRIGHTNESS OF GALAXIES*

V. A. AMBARTSUMIAN

Byurakan Astrophysical Observatory, Armenian Academy of Sciences, Yerevan, Armenian SSR, U.S.S.R.

(Received 9 September, 1987)

Abstract. We consider the problem of correlation between the surface brightnesses of two nearby points for an external galaxy under certain simplifying assumptions. A probability density function for these intensities are derived, as are its moments and correlation coefficient.

1. Introduction

About fifty years have elapsed since the time when the idea was put forward that the absorbing matter in our Galaxy consists of a large number of discrete absorbing clouds of different optical thickness (see Ambartsumian and Gordeladze, 1938). The clouds of large optical thickness we often observe as the 'dark nebulae' or 'dark spots' in the Milky Way, The clouds of the optical thickness $\tau < 0.5$ are causing local decreases in the number of stars of certain apparent magnitudes per square min of arc. Thus clouds of both kind determine the differences of surface brightness of the Milky Way which conditionally bear the name of 'fluctuations of surface brightness'.

The construction of more or less realistic theory of such fluctuations encounters great difficulties. Therefore, some very simplified mathematical models have been studied. Thus the model was considered in which the stars are distributed in the infinite space with such uniformity that we can substitute them by a continuous luminous medium having a constant coefficient of emission η, which is independent of coordinates, while the absorbing clouds are distributed at random but uniformly in the sense that the probability of the presence of the centre of a cloud in a volume element is equal to this volume multiplied by a constant coefficient. In such a model the absorbing properties, the shapes and the sizes of clouds are to be described by different parameters and the statistical distribution of those parameters must be given. It is supposed that this distribution is independent of the coordinates of the centre of the cloud. For example we can have a set of spherical clouds of a given radius (or of different radii) that are distributed at random (according to Poisson law) in the space. However, the sizes of clouds have to be negligibly small as compared with the distances between them.

Under these assumptions an equation for the probability distribution function for the values of brightness at a point of Milky Way has been derived and the mean values of brightness and of the square deviation have been found (Ambartsumian, 1944).

Comparing the obtained formal expressions with the observations it was possible to

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 635–638.

estimate the approximate values of some parameters describing the statistics of absorbing clouds.

However, the problem of correlation between the surface brightnesses of two mutually nearby points of Milky Way, some finite angular distance α apart has not been solved. It appears that it requires some complicated calculations.

In this paper we consider the correlation problem for a somewhat different model, where we deal with an external spiral system (external galaxy), when the rays directed to the observer from two points of the system are almost parallel in so far as we remain within the system or in its neighbourhood. We shall assume also that the optical depth traversed by each of the rays before they reach the boundary directed to the observer is infinite. It follows that the results obtained from consideration of such a model are applicable mainly to the cases when the observer is situated near the equatorial plane of the external galaxy. This means that he sees the spiral system edge on. The examples of such systems are NGC 4594 (M104), NGC 4244, and NGC 4631.

Though in a system described above the volume filled by absorbing clouds and emitting matter (stars) is only a half space, the problem formulated is equivalent to the problem of correlation between intensities for two parallel rays (distance r apart) in the homogeneous medium of the same kind, but filling the entire space. But of course, we are interested only in the correlation between intensities in points of intersection of our two rays with a plane perpendicular to them.

2. The Equation for the Probability Distribution

Evidently we can apply here the same *invariance principle* which was used by us (see Ambartsumian, 1944) in the case of one single ray. Thus our problem is to find the probability density function $u(x_1, x_2; r)$ of the intensities x_1 and x_2, in the corresponding points of our two rays.

According to invariance principle when we displace both points by an amount ΔS in the same direction along the corresponding rays, the two-dimensional distribution function of the two quantities x_1 and x_2 must remain unchanged. This means that the sum of all possible changes equals zero. From this and passing to the limit $\Delta S \to 0$ we obtain – similarly as in the paper just quoted, the equation

$$u(x_1, x_2; r) + \frac{\eta}{\nu(r)}\left(\frac{\partial u}{\partial x_1} + \frac{\partial u}{\partial x_2}\right) = \int\int u\left(\frac{x_1}{q_1}, \frac{x_2}{q_2}; r\right) \frac{d^2\phi_r(q_1, q_2)}{q_1 q_2}, \quad (1)$$

where $\nu(r)$ is defined by the condition that $\nu(r)\Delta S$ is the probability of intersection of at least one of two rays on the range ΔS by some absorbing cloud, while $\phi_r(q_1, q_2)$ is the conditional probability that in this case the intensity of the first ray will be multiplied by a transmission coefficient that is $\leqq q_1$ while the intensity of the second ray by a coefficient $\leqq q_2$. Thus we assume that generally the transmission coefficients of a cloud when passing it at different points are different.

We assume that the clouds may differ in their shapes and transparencies, but they form a homogeneous set, which can be described by one transparency function $\phi_r(q_1, q_2)$ as defined above. It is clear that, in our model,

$$\phi_r(q_1, q_2) = \phi_r(q_2, q_1) \tag{2}$$

describes the optical-statistical properties of the whole set of absorbing clouds.

For the following it is convenient to introduce the dimensionless intensities

$$y_1 = \frac{\nu x_1}{\eta}\,; \qquad y_2 = \frac{\nu x_2}{\eta}\,. \tag{3}$$

Then for their distribution function we shall have the equation

$$u + \frac{\partial u}{\partial y_1} + \frac{\partial u}{\partial y_2} = \int\!\!\int u\left(\frac{y_1}{q_1}, \frac{y_2}{q_2}; r\right) \frac{\mathrm{d}^2\phi_r(q_1, q_2)}{q_1 q_2}\,, \tag{4}$$

from which we can easily obtain the expressions for the expectations of products $y_1^k y_2^l$.

3. The Moments and the Correlation Coefficient

If we multiply (4) by $y_1^k y_2^l$ and integrate over all possible values of y_1 and y_2 we obtain the relations from which we can find the values of corresponding mathematical *expectations* (moments). The most interesting fact is that the moments $\overline{y_1^k y_2^l}$ derived in this way are expressed in simple ways through the different moments of products of the type $q_1^n q_2^m$. Owing the symmetry condition (2) we always have

$$\overline{y_1^k y_2^l} = \overline{y_1^l y_2^k}\,; \qquad \overline{q_1^m q_2^n} = \overline{q_1^n q_2^m}\,. \tag{5}$$

Since, in particular

$$\overline{y_1^k} = \overline{y_2^k}\,; \qquad \overline{q_1^m} = \overline{q_2^m}$$

it is convenient to write simply $\overline{y^k}$ instead of $\overline{y_1^k}$ and of $\overline{y_2^k}$ as well as $\overline{q^k}$ instead of $\overline{q_1^k}$ and $\overline{q_2^k}$.

As the result of calculation we obtain

$$\bar{y} = \frac{1}{1 - q}\,; \qquad \overline{y^2} = \frac{2}{(1 - \bar{q})(1 - \overline{q^2})}\,;$$
$$\overline{y_1 y_2} = \frac{2}{(1 - \bar{q})(1 - \overline{q_1 q_2})}\,. \tag{6}$$

Thus for the ratio of moments we have

$$\rho = \frac{\overline{y_1 y_2}}{\overline{y^2}} = \frac{1 - \overline{q^2}}{1 - \overline{q_1 q_2}}\,, \tag{7}$$

and for the correlation coefficient

$$K = \frac{\overline{(1 - q_1)(1 - q_2)}}{\overline{(1 - q)}^2} \frac{1 - \overline{q^2}}{1 - \overline{q_1 q_2}} . \tag{8}$$

4. Example

Let us consider for simplicity an artificial example. We suppose that all clouds are thin square plates of equal size and of equal transparency coefficient q_0, when the rays are perpendicular to them. Let l is the side of the squares. The corresponding sides of squares are parallel each other and the plane of each square is perpendicular to the direction of the ray. One of the sides of squares is parallel to the line connecting the points of rays at which the intensities we consider. The centre of the squares are distributed in space at random, according to Poisson law.

If a given cloud intersects one of our rays the probability that it intersects also the second ray will be $(l - r)/(l + r)$, while the probability of intersecting of only one of the rays is $2r/(l + r)$. In this case we have

$$\overline{q_1} = \overline{q_2} = \frac{l}{l + r} q_0 ; \qquad \overline{q_1^2} = \overline{q_2^2} = \frac{l}{l + r} q_0^2 ; \qquad \overline{q_1 q_2} = \frac{l - r}{l + r} q_0^2 \tag{9}$$

and according (7) we obtain

$$\rho = \frac{l(1 - q_0^2) + r}{l(1 - q_0^2) + r(1 + q_0^2)} . \tag{10}$$

If for example $r = \frac{1}{2}l$, $q_0 = \frac{1}{2}$ the correlation coefficient, $K = \frac{25}{33}$. Thus the correlation in this case is sufficiently strong.

References

Ambartsumian, V. A. and Gordeladze, Sh. G.: 1938, *Bull. Abast. Obs.* **2**, 37.
Ambartsumian, V. A.: 1944, *Dokl. Akad. Sci. USSR* **44**, 244.

ON THE DYNAMICS OF THE METAGALAXY*

B. E. LAURENT

Institute of Theoretical Physics, University of Stockholm, Sweden

B. BONNEVIER

Plasma Physics and Fusion Research, The Royal Institute of Technology, Stockholm, Sweden

and

P. CARLQVIST

Department of Plasma Physics, The Royal Institute of Technology, Stockholm, Sweden

(Received 30 September, 1987)

Abstract. The dynamical aspects of the metagalactic model are studied. In particular, the consequences of high galaxy and quasar red shifts are considered. A number of explosion mechanisms are investigated and limits are given for the velocities they can produce. In case of constant density in the local rest frame limits are also given for the parameter Ω_0, for our position in the metagalaxy, and for the initial mass.

1. Introduction

In a number of papers Hannes Alfvén has discussed alternatives to Big Bang cosmologies. In the first of these he considered, together with O. Klein, the metagalactic model (Alfvén and Klein, 1962). The main assumption behind this model is that all the galaxies observed today form a huge cloud, the metagalaxy, surrounded by vacuum. An explosion is assumed to have taken place which has caused the observed Hubble expansion. It has been stressed by Alfvén as well as by Klein that this explosion could be a *genuine physical process* in contrast to the start of the Big Bang. Klein also pointed out that this could offer a physical explanation of the famous Eddington relations (Klein, 1953).

A number of different aspects of the metagalactic model has been discussed over the years – pro et contra. The most noteable is probably Klein's and Alfvén's ingenious studies of matter-antimatter as the origin of the metagalactic explosion (Alfvén and Klein, 1962). Another important aspect is the background microwave radiation. Its high degree of isotropy is not so easy to understand from the metagalactic point of view. On the other hand every cosmology seems to have difficulties on this point especially now when grand scale inhomogeneities in the matter distribution have been discovered (de Lapparent *et al.*, 1986).

Here we focus our interest on the dynamics of the metagalaxy. A simple Newtonian model was presented in the original paper by Alfvén and Klein (1962) and this was further studied by one of us (Bonnevier, 1964). Alfvén has put forward a number of ideas concerning the dynamical behaviour of the metagalaxy. For instance he has suggested

* Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 639–658.

a model in which the pressure within the metagalaxy is unimportant and where the main dynamical effect of the explosion is to lower the total gravitating mass through heavy radiative loss. Another of his suggestions is that 'hyperelastic' collisions between lumps of matter have taken place, i.e., that the kinetic energy is increased in each collision between a lump and an anti-lump due to annihilation processes in the contact surface. He has put forward the 'fire-work' model with a series of explosions at different hierarchical levels. Throughout he has stressed the importance of inhomogeneities.

This paper is to a large extent a renewed discussion of some of Alfvén's ideas. Among other things we incorporate them into relativistic models.

Newtonian models do not suffice. Certain important restrictions on the metagalactic mass and the maximum velocities which appear in general-relativistic models get lost in the Newtonian approximation. An example is found by Laurent and Söderholm (1969).

2. The Observational Background and the Energetics of the Metagalaxy

Before we can discuss various mechanisms of the expansion of the metagalaxy we must define the observational framework to which any metagalactic model of expansion has to be adapted. The most relevant observations in this context concern the Hubble parameter, the mean mass density, and the maximum velocities (red shifts) found in the metagalaxy. Both the Hubble parameter and the mean mass density are extremely difficult to determine and are, hence, subject to considerable sources of error. The value of the measured Hubble parameter has steadily decreased through the decades and is at present in the interval $H_0 = 50\text{–}100\ \mathrm{km\ s^{-1}\ Mpc^{-1}} = 1.6 \times 10^{-18}\text{–}3.2 \times 10^{-18}\ \mathrm{s^{-1}}$.

Moreover, H_0 seems not to be a constant but may vary with the direction and distance (Rubin *et al.*, 1976; Burstein *et al.*, 1986).

As regards the present mean mass density of the metagalaxy ρ_0 it may be interpreted in terms of the parameter

$$\Omega_0 = \frac{8\pi G}{3H_0^2}\,\rho_0 \,. \tag{2.1}$$

Gott *et al.* (1974) have given a range of values of Ω_0 rather than ρ_0. They find that if only masses of galaxies that are directly observed are taken into account then $\Omega_0 = 1.3 \times 10^{-3}$. If, on the other hand, the dynamics of groups and clusters of galaxies are considered, so that unseen matter is included, they obtain $\Omega_0 = 5 \times 10^{-2}$. Using the above intervals of values of H_0 and Ω_0 we get from Equation (2.1) a mean mass density of $\rho_0 \approx 6 \times 10^{-30}\text{–}9 \times 10^{-28}\ \mathrm{kg\ m^{-3}}$. It is to be noted that the smaller of these values is in fairly good agreement with the value $\rho_0 \approx 10^{-30}\ \mathrm{kg\ m^{-3}}$, obtained from an extrapolation of de Vaucouleurs's empirical law of the mass density of hierarchical systems (de Vaucouleurs, 1970; Alfvén, 1981; pp. 126).

Of great importance for the metagalactic theory are the highest observed red shifts, which are about $z = 3$ for galaxies and $z = 4$ for quasars.

In contrast to big-bang hypotheses the metagalactic hypotheses puts an upper limit to Ω_0. Assuming that the metagalaxy is homogeneous a measured red shift of $z = 4$ shows that $\Omega_0 \leq 0.04$. A discussion of this is found in the Appendix where we abstain from treating the mechanism which started the expansion and regard only the limits which must be set on a metagalaxy which is homogeneous out to its border. The Appendix also discusses, using general relativity, the ratio between rest mass m and total mass M of the metagalaxy. As has been pointed out by Alfvén this question can with reasonable precision be discussed within special relativity if we regard the presently observed state of the metagalaxy.

The following is an outline of the discussion: let (t, r, ϑ, ϕ) be polar coordinates for an inertial frame. An explosion is supposed to have taken place at $r = t = 0$. The galaxies are assumed to move according to the law

$$\mathrm{d}r/\mathrm{d}t = r/t \,. \tag{2.2}$$

This means for a particular galaxy $r = \text{const.} \times t$ or

$$\left.\begin{aligned} r &= c\tau f_1(\chi)\,, \\ t &= \tau f_2(\chi)\,, \end{aligned}\right\} \tag{2.3}$$

where χ is supposed to label the galaxies and τ is any parameter along the galaxy world lines. We may in particular choose it to be the proper time, i.e., $\mathrm{d}t^2 - (1/c^2)\,\mathrm{d}r^2 = \mathrm{d}\tau^2$ when $\chi = \text{const}$. We are, moreover, free to choose the labelling and we do it so that $f_1(\chi) = \sinh\chi$. All this gives

$$\left.\begin{aligned} r &= c\tau \sinh\chi\,, \\ \mathrm{t} &= \tau \cosh\chi\,. \end{aligned}\right\} \tag{2.4}$$

The line element is

$$\begin{aligned} \mathrm{d}s^2 &= c^2\,\mathrm{d}t^2 - \mathrm{d}r^2 - r^2\,(\mathrm{d}\vartheta^2 + \sin^2\vartheta\,\mathrm{d}\phi^2) = \\ &= c^2\left\{\mathrm{d}\tau^2 - \tau^2\,\mathrm{d}\chi^2 - \tau^2\sinh^2\chi\,(\mathrm{d}\vartheta^2 + \sin^2\vartheta\,\mathrm{d}\phi^2)\right\}. \end{aligned} \tag{2.5}$$

The space part of this line element shows that the situation is *homogeneous in space* (cf. the Appendix). Therefore, Equation (2.2) can be used with any matter particle at the origin.

In particular (2.5) tells us that the element of length in the χ-direction is $c\tau\,\mathrm{d}\chi$.

Choosing $\rho(\chi, \tau)$ to be the mass density which is measured in the rest frame of matter at (χ, τ) we find the total rest mass at constant $\tau = \tau_1$ to be

$$m = 4\pi c^3 \int_0^{\chi_1} \rho\tau_1^3 \sinh^2\chi\,\mathrm{d}\chi\,. \tag{2.6}$$

Assuming $\sigma(r, t)$ to be the rest mass density in the (r, t) frame we can calculate m as

$$m = 4\pi \int_0^{r_1} \sigma r^2\,\mathrm{d}r\,; \tag{2.7}$$

now at constant $t = t_1$ which we assume coincides with τ_1 at the origin. If we choose r_1 according to (2.4),

$$r_1 = ct_1 \tanh \chi_1 \,. \tag{2.8}$$

Equations (2.6) and (2.7) must give the same value of m. This gives

$$\sigma(r, t) = \frac{\rho(\chi, t)}{(1 - (r/ct)^2)^2} \,, \tag{2.9}$$

where χ is to be expressed in r/t according to (2.4). From these equations we obtain

$$\cosh \chi = (1 - (r/ct)^2)^{-1/2} \tag{2.10}$$

and

$$z + 1 = \sqrt{\frac{1 + r/ct}{1 - r/ct}} = e^{\chi} \,. \tag{2.11}$$

The total energy Mc^2 is obtained by replacing ρ in (2.6) by $\rho(1 - (r/ct)^2)^{-1/2} = \rho \cosh \chi$. This gives

$$M = \frac{4\pi}{3} \rho r^3 \,. \tag{2.12}$$

If ρ is constant, Equations (2.6), (2.11), and (2.12) agree with Equations (A.17), (A.16), and (A.6) in the Appendix. Thus Table II in the Appendix may be used here.

It is obvious that m in (2.6) cannot depend on τ_1. Hence, $\rho(\chi, \tau) = \tau^{-3} P(\chi)$, where P is a function of χ only. In contrast to the general-relativistic theory the present theory allows us to choose $P(\chi)$ freely.

As an example we can choose the density distribution $P(\chi) = P_1 \exp(-\chi^2/\chi_1^2)$. It is interesting to compare the results for two different distributions. We have calculated numerically the value of m/M for $z = 4$ and find for $\chi_1 = \infty$ that $m/M = 0.50$ and for $\chi_1 = \chi_B$ (where χ_B is the χ-value of the border) $m/M = 0.53$. Obviously the difference is quite small.

The highest observed red shifts ($z = 4$) correspond to velocities about $dr/dt \approx 0.92c$ (see Equation (2.11)). If we use the observed value of the Hubble parameter $H_0 \equiv t_0^{-1} = 1.6 \times 10^{-18}$–$3.2 \times 10^{-18}$ s^{-1}, Equation (2.2) gives the lower limit of the radius for the metagalaxy of $r_{\min} \approx 8.4 \times 10^{25}$–$17 \times 10^{25}$ m.

Now using the observed value of the mass density $\rho_0 \approx 6 \times 10^{-30}$–$9 \times 10^{-28}$ kg m^{-3}, Equation (2.12) gives us the following lower limit for the metagalactic mass $M_{\min} \approx 6 \times 10^{49}$–$5 \times 10^{51}$ kg. The corresponding Schwarzschild radius is $2GM_{\min}/c^2 \approx 9 \times 10^{22}$–$7 \times 10^{24}$ m.

3. Observer Position in the Metagalaxy

Up to this point we have argued as if we, the observers, were placed at the centre of the metagalaxy. Now this seems an unlikely position and it is interesting to consider an off centre position. Such a position requires an adjustment of our observed maximum red shifts to the corresponding red shifts at the centre. This may have dramatic consequences.

Let us discuss this problem within the special relativistic model presented in Section 2. Assume that the position of the centre is at a value χ_c counted from us, and that the border in the opposite direction is at a value χ_b. The discussion in the Appendix immediately shows that then the border is at the χ-value

$$\chi_B = \chi_c + \chi_b \tag{3.1}$$

counted from the centre.

We want to study the ratio

$$k \equiv \frac{m(\chi_1 = \chi_B)}{m(\chi_1 = \chi_c)} \tag{3.2}$$

(see Equation (2.6)). The value of k is a measure of the representative role of our position.

Putting as above $\rho = \text{const.}$, we obtain

$$k\,(\sinh 2\chi_c - 2\chi_c) = \sinh 2\chi_B - 2\chi_B\,. \tag{3.3}$$

Taking the high z-values observed for galaxies and quasars into account we conclude that z_b is at least 4. Assuming first that $\chi_c \gg 1$ so that Equation (3.3) may be approximated by

$$k\,e^{2\chi_c} = e^{2\chi_B}\,, \tag{3.4}$$

we find using (3.1) that

$$k = e^{2\chi_b} = (z_b + 1)^2\,. \tag{3.5}$$

Choosing $z_b = 4$ and studying (3.1) and (3.3) numerically we find, in fact, that $k > 25$ and that k tends to 25 for high χ_c values in accordance with Equation (3.5) (see Table III in Appendix).

This shows that if the metagalaxy is homogeneous out to the border the measured high z-values force us to conclude that we ourselves are within a very small portion of the metagalaxy close to the centre.

If we are unwilling to accept this we should probably have to assume that the density varies within the metagalaxy. Dynamical considerations point in the same direction.

4. Expansion Due to Radiation Pressure

In the original metagalactic model the explosion produces radiation which then pushes the light, charged particles outwards. The light and heavy particles are assumed to be coupled through magnetic fields and the result is the Hubble expansion. It is important for the model that the ratio between the energy density and the pressure is low to keep the gravitational forces down. For this reason the photon gas is more effective than any other gas.

There exists a simple argument indicating that the outward velocities obtained in such models can never be close to the velocity of light: for the explosion to be effective it must accomplish most of the acceleration of the matter in a short span of time and so the radius of the cloud changes little while the acceleration occurs (Laurent and Söderholm, 1969). During this time the radiation should be at least momentarily confined (i.e., the mean-free path of the photons should be less than the extension of the system) to get a pressure as high as possible. One can compare this to radiation closed up in a box with reflecting walls (see Figure 1), containing isotropic radiation.

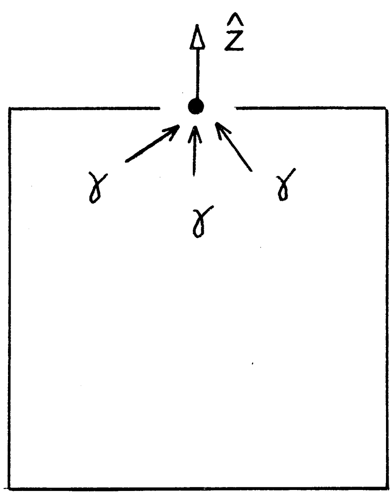


Fig. 1. A box with reflecting walls containing isotropic radiation is depicted. There is a hole in one of the walls. A particle being just inside the hole is influenced by the radiation pressure tending to accelerate the particle out of the hole. It can be shown that the particle cannot attain a velocity larger than $c/2$ due to aberration effects.

Suppose that there is a small hole in the wall and a charged particle in the hole on its way out of the box. Suppose also that the scattering cross section of photons on the particle is independent of the energy and direction of the photon *in the rest frame of the particle*. This means that the momentum delivered to the particle from one particular photon is proportional to k_z' – the z-component of the photon momentum in the particle rest frame. The particle is supposed to move very little compared to the dimensions of the box while it obtains momentum from the photons. The total momentum delivered to the particle is then proportional to

$$\int f(\omega) k_z' \, \mathrm{d}^3 k \,, \tag{4.1}$$

where $f(\omega)$ is the energy distribution of photons in the box and the integration of directions is performed over a half sphere. This gives

$$\int_0^\infty \int_0^{\pi/2} \int_0^{2\pi} f(\omega) \frac{k_z - \dfrac{V}{c}\,\omega}{\sqrt{1 - V^2/c^2}} \frac{\omega^2}{c^2} \sin\vartheta \frac{\mathrm{d}\omega}{c} \,\mathrm{d}\vartheta\,\mathrm{d}\phi\,, \tag{4.2}$$

where V is the particle velocity.

The integral contains a factor

$$\int_0^{\pi/2} \left(\cos\vartheta - \frac{V}{c}\right) \sin\vartheta\,\mathrm{d}\vartheta = \frac{1}{2} - \frac{V}{c}\,. \tag{4.3}$$

Thus the force on the particle vanishes for $V = 1/2c$. The particle cannot be accelerated to a velocity greater than half the velocity of light. The physical explanation of this phenomenon is that, due to aberration, the particle 'sees' some of the photons being directed in towards the box.

A detailed calculation on a big gas cloud containing particles and antiparticles supports this conclusion (Laurent and Söderholm, 1969). Numerical integrations giving the motion of such a cloud have not shown any outward velocities greater than 0.4 times the velocity of light.

5. Expansion without Pressure

Alfvén has suggested a possible way to overcome the velocity problem. Imagine a big cloud of particles and antiparticles, which also contains matter and antimatter lumps. The history of this cloud is assumed to run through three stages: (1) A free fall inwards for both particles and lumps under the action solely of the cloud's own gravitation. (2) A sudden explosion which converts the particle gas to radiation. The radiation then disperses quickly leaving the lumps unaffected behind. (3) The lumps do not collide but spread and fly outwards in a free fall.

On their way out the lumps experience less gravitation than on their way in. Is it possible that they could acquire high velocities this way?

5.1. Newtonian treatment

In order to answer this question we first construct a special Newtonian model of the free fall stages. Let us assume that the gravitation field is spherically-symmetric. The radial motion of each particle (and lump) is given by a curve $r = r(t)$. Let us also assume that such curves for different particles do not cross. A certain particle (or lump) then experiences a constant gravitational mass M throughout its motion which, therefore, is a Keplerian motion. The conservation of energy thus gives

$$\frac{1}{2}\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 + \frac{1}{2}\frac{l^2}{r^2} = \frac{GM}{r} + \mathscr{E}\,, \tag{5.1}$$

(see, e.g., Foster and Nightingale, 1979, p. 115). Here l and $\mathscr{E}$ are constants of the motion. If μ is the particle (or lump) mass, $\mu\mathscr{E}$ is the energy and μl is the angular momentum. As usual G is the constant of gravitation.

Equation (5.1) is invariant under the scale transformation $r \to Kr$, $t \to t$, $l \to K^2 l$, $M \to K^3 M$, $\mathscr{E} \to K^2 \mathscr{E}$, where K is a constant. Thus the matter may be arranged so that such re-scalings connect the motions of all the matter shells with one another. This implies that the mass density is independent of r and that velocities, tangential as well as radial, are linear in r. To obtain a spherically-symmetric gravitational field the tangential velocities must be isotropically distributed. Hence, particles have different relative velocities in different matter shells and the model is not homogeneous.

5.2. General-Relativistic Treatment

To obtain velocities close to the velocity of light the outermost lumps have to come close to the Schwarzschild limit. For this reason the Newtonian approximation does not suffice.

It is not very difficult to find out how the *outermost matter shell* behaves also when it is close to the Schwarzschild limit, for this shell moves all the time in a Schwarzschild vacuum field. Its equation of motion is

$$\frac{1}{2}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 + \frac{1}{2}\frac{l^2}{r^2} = \frac{GM}{r} + \frac{GM}{r^3}\frac{l^2}{c^2} + \mathscr{E}\,. \tag{5.2}$$

This can be found, e.g., in Foster and Nightingale (1979, pp. 106–107). Comparing with (5.1) we see that the time t has been replaced by the proper time τ and there is one more term on the right-hand side. The meaning of r is the Schwarzschild radial coordinate. The quantities M, $\mathscr{E}$, and l are constant in the motion.

We note that Equation (5.2) does not possess the scale invariance of the Newtonian equation (5.1). This is, however, of no concern to us now as we are considering only the outermost matter shell.

We shall assume that the cloud starts from rest when falling in so that $\mathscr{E} = 0$. In this case $\mathrm{d}\tau/\mathrm{d}t = 1 - 2GM/(c^2 r)$, where t is the Schwarzschild time (see Foster and Nightingale, 1979; p. 106). It is then easy to check that the outermost shell cannot overtake an inner shell. To that end we write (5.2) in the form

$$\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 \left(1 - \frac{2GM}{c^2 r}\right)^{-2} = \frac{2GM}{r} - \frac{l^2}{r^2}\left(1 - \frac{2GM}{c^2 r}\right). \tag{5.3}$$

Assume that the outermost shell had caught up an inner shell. Then r (and M) would have the same value for both, while l is greater for the former one. Then (5.3) shows that the originally outer shell has a lower velocity $\mathrm{d}r/\mathrm{d}t$ than the inner one and this contradicts the assumption.

If the cloud does not collapse, the velocity $\mathrm{d}r/\mathrm{d}\tau$ must vanish at some time. The radius $r_{\min}$ of the shell at that occasion satisfies the equation

$$\frac{2GM}{c^2}\left(\frac{1}{r_{\min}}\right)^2 - \frac{1}{r_{\min}} + \frac{2GM}{l^2} = 0\,. \tag{5.4}$$

The condition for real roots is

$$l \geq 4 \frac{GM}{c} . \tag{5.5}$$

Choosing the equality here, the cloud contracts as much as it can without collapsing. We then obtain from (5.4)

$$r_{\min} \geq 4 \frac{GM}{c^2} . \tag{5.6}$$

Thus the cloud cannot contract to a radius below twice the Schwarzschild radius without collapsing.

Suppose now that when the cloud is in its most contracted state a part q_r, of the original mass suddenly 'vanishes' through an explosion and loss of radiation. It is reasonable to assume that the lumps in the outermost shell are left with unchanged angular momentum and unchanged radius $r_{\min}$. Inserting $\mathrm{d}r/\mathrm{d}\tau = 0$, $M \to (1 - q_r)M$, $l = 4GM/c$, and $r = 4GM/c^2$ into (5.2) we obtain

$$\mathscr{E} = \tfrac{1}{2} q_r c^2 . \tag{5.7}$$

Now, using this and putting $r \to \infty$ (in 5.2) we obtain

$$\frac{v_f^2}{1 - v_f^2/c^2} \equiv \left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2_{\text{final}} = q_r c^2 , \tag{5.8}$$

where v_f is the final velocity.

Hence,

$$\frac{v_f}{c} = \sqrt{\frac{q_r}{1 + q_r}} . \tag{5.9}$$

We note that there is an upper limit for the final velocity

$$v_f \leq \frac{1}{\sqrt{2}} c \approx 0.7c . \tag{5.10}$$

In the limit $q = 1$ so that the entire mass vanishes and this is hardly a possible event. To see why, let us calculate the velocity w of an outer shell lump in the most contracted state. According to Foster and Nightingale (1979, p. 106)

$$r^2 \frac{\mathrm{d}\phi}{\mathrm{d}\tau} = l , \tag{5.11}$$

so that

$$w \equiv r_{\min} \frac{\mathrm{d}\phi}{\mathrm{d}t} = \frac{l}{r_{\min}} \left(1 - \frac{2GM}{c^2 r}\right) = \frac{1}{2} c . \tag{5.12}$$

We shall compare this 'coordinate' velocity with the corresponding velocity for light which travels radially. This is obtained by putting $d\tau^2 = 0$ in the Schwarzschild line element (see Foster and Nightingale, 1979; p. 89)

$$0 = \left(1 - \frac{2GM}{c^2 r}\right) dt^2 - \left(1 - \frac{2GM}{c^2 r}\right)^{-1} dr^2 , \tag{5.13}$$

which gives

$$\left(\frac{dr}{dt}\right)_{\text{light}} = \frac{1}{2} c . \tag{5.14}$$

This shows that lumps and light are equally fast. The sudden approximation is, therefore, actually not a very good one and the only reliable result is that v_f must lie well below $0.7c$. Even so this may be considered an improvement compared to the upper limit $0.5c$ which we obtained in the model with radiation pressure.

It should finally be remarked that both these models require quite special conditions to function well. The pressure free model requires that the angular momentum is close to a certain value $l \approx 4(GM/c)$. It also requires that the explosion takes place at a time very close to the time of the turning.

6. Expansion Due to Magnetic Pressure

We next consider a metagalactic model which includes an expansion mechanism relying on the presence of a magnetic field. In this model the metagalactic plasma is penetrated by a magnetic field which is assumed to be frozen-in to the plasma. For the sake of simplicity it is supposed that the field has no direction of preference on a large scale.

When gravity contracts the initial metagalactic cloud of radius r, the frozen-in magnetic field increases approximately as

$$B \propto r^{-2} . \tag{6.1}$$

This means that the magnetic energy per unit volume grows as $B^2/2\mu_0 \propto r^{-4}$, while the total magnetic energy in the metagalaxy increases as

$$W_B \propto r^{-1} . \tag{6.2}$$

Hence, it is clear that part of the gravitational potential energy released during the contraction is transferred to magnetic energy while the rest goes to kinetic energy of the metagalactic plasma (cf. Section 5). If the metagalactic magnetic field is sufficiently strong a large fraction of the potential energy may be transformed into magnetic energy.

At a certain time a considerable part of the initial metagalactic mass is annihilated. Most of the annihilated mass disappears quickly out of the system as radiation. For a large enough magnetic field the outwardly directed magnetic pressure force may dominate over the inwardly directed gravitational force the sources of which according to the Einstein theory are particle masses, magnetic energy, and also the magnetic pressure itself. The result will be an expansion of the metagalaxy.

We may make an estimate of the magnetic field required. In order for the metagalactic matter to expand with velocities up to values corresponding to $z \approx 4$ it is necessary that at the time of the turning of the metagalaxy the total magnetic energy is comparable to the rest mass energy of the remaining metagalaxy mc^2 (cf. Section 2 and Table II) which implies

$$\frac{B_t^2}{2\mu_0} \approx \rho_t c^2 , \tag{6.3}$$

where index t indicates the values at the turning point. If we put $\rho_0 = \kappa^3 \rho_t$ we find from relation (6.1) the present magnetic field to be $B_0 \approx \kappa^2 B_t$. Hence, we obtain from Equation (6.3)

$$B_0 \approx (2\mu_0 c^2 \kappa \rho_0)^{1/2} . \tag{6.4}$$

Inserting $\rho_0 = 6 \times 10^{-30}$–$9 \times 10^{-28}$ kg m^{-3} and $\kappa = 10^{-1}$ we find that a magnetic field of $B_0 \approx 4 \times 10^{-10}$–$4 \times 10^{-9}$ T is required to cause the Hubble expansion in one step. This magnetic field, which is comparable to the galactic field, is quite large. However, if the expansion takes place in subsystems of smaller mass and dimension a weaker field is required.

The condition for the magnetic field to be frozen-in to the metagalactic plasma during a time $\mathscr{T}_f$ is that the resistivity of the plasma (including the effects of double layers and other anomalous mechanisms) should be less than $\mu_0 l_B^2 / \mathscr{T}_f$ where l_B denotes a characteristic length scale of the magnetic field (see, e.g., Cowling, 1957). With for instance $\mathscr{T}_f = 10^{18}$ s $= 3 \times 10^{10}$ yr and $l_B = 10^{23}$ m $\approx 10^7$ light years we find the resistivity to be lower than 10^{22} Ω which is a very high value.

7. Expansion Due to Charged Particle Pressure

We shall now investigate a model of expansion of the metagalaxy which relies upon charged-particle pressure and which can be said to complement the model based on radiation pressure described in Section 4. The model is extreme in the sense that it is transparent to all kinds of radiation. Hence, there is no radiation pressure acting on the matter. The expansion of the metagalaxy is instead caused by the pressure of very energetic electrons and positrons formed during the annihilation phase.

When a proton and an antiproton annihilate a fraction $q_e \approx 0.17$ of the initial rest mass energy is transferred to in the mean 3.2 electrons and positrons (see, e.g., Ekspong *et al.*, 1966). The rest of the energy goes to neutrinos (≈ 0.50) and to γ-rays (≈ 0.33). The mean energy of each of the electrons and positrons is of the order of 100 MeV.

It is assumed that a weak magnetic field B pervades the metagalaxy. As the electron-positron gas tends to expand due to pressure forces also heavier particles like protons and antiprotons are dragged along by the common action of electric and magnetic fields. If the pressure (energy density) of the electron-positron gas is large enough the velocity of expansion may amount to a considerable fraction of c.

As was demonstrated in Section 2 the kinetic energy of the metagalaxy is approximately equal to the rest mass energy for reasonable models comprising maximum z_B-values of four. A minimum requirement (neglecting gravitation) for the expansion model to work is, therefore, that the electron-positron gas initially possesses an energy of this magnitude. If we denote the total mass annihilated in the metagalaxy by M_{A} it must be valid

$$q_e M_{\mathrm{A}} \gtrsim m \,. \tag{7.1}$$

With $q_e = 0.17$ we find $M_{\mathrm{A}} \geq 6m$. This means that if the model is to work the initial mass of the metagalaxy should have been at least seven times as large as the present rest mass. The Schwarzschild radius of the initial mass should, therefore, have a minimum value of seven times the value given in Section 2. A demand for the model is that the turning from a contraction to an expansion of the metagalaxy must have occurred in a volume with a radius not smaller than the Schwarzschild radius of the initial mass.

The time-scale in which the charged particles lose their kinetic energy is obviously of decisive importance. It seems reasonable to require that the energetic electron-positron gas should not lose a substantial fraction of its energy as (1) synchrotron radiation, (2) bremsstrahlung, (3) Compton radiation, and (4) annihilation radiation within a time shorter than the characteristic time it takes for the metagalaxy to turn from contraction to expansion

$$\mathscr{T}_t \approx r_t/c \approx 10^8\text{–}10^9\ \mathrm{yr} \approx 3 \times 10^{15}\text{–}3 \times 10^{16}\ \mathrm{s}\,.$$

(1) The time constant for energy loss of an electron of energy W_e due to synchrotron radiation is

$$\mathscr{T}_s \approx \frac{5}{B^2}\,\frac{1}{1 + W_e/W_{e0}}\,, \tag{7.2}$$

where $W_{e0} = 0.5$ MeV is the rest mass energy of the electron (see, e.g., Alfvén and Fälthammar, 1963; p. 69). For $W_e = 2 \times 10^2\, W_{e0}$ and $\mathscr{T}_s \geq \mathscr{T}_t$ we get $B \leq 10^{-9}\text{–}3 \times 10^{-9}$ T.

(2) According to, e.g., Korchak (1967) the spectral radiant power of bremsstrahlung from a relativistic electron is

$$P_b \approx 4.61 \times 10^{-56} n_i Z^2 L\,, \tag{7.3}$$

where n_i denotes the ion density, Z is the atomic number, while L is a function of the electron energy and the frequency of the radiation. Integrating Equation (7.3) over the frequency for $Z = 1$ we obtain the total power radiated by a 100 MeV electron, $P_b \approx 8 \times 10^{-33} n_i$. Hence, the time-constant for energy release due to bremsstrahlung is

$$\mathscr{T}_b \approx \frac{W_e}{P_b} \approx 2 \times 10^{21} n_i^{-1}\,. \tag{7.4}$$

If $\mathscr{T}_b$ is to be larger than $\mathscr{T}_t$ the following must be valid $n_i \leq 10^5$–10^6 m^{-3}.

(3) When an energetic electron or positron scatters a γ-quantum (~ 180 MeV) from the annihilation process its energy is influenced rather little. However, if the electrons and positrons interact with a radiation of much smaller photon energy the particles may suffer energy loss as a result of the inverse Compton effect. The spectral power of Compton radiation from a relativistic electron in a radiation field of temperature T and photon density n_{ph} is

$$P_C \approx 1.65 \times 10^{-53} n_{ph} F(\nu/\nu_c) \quad (\mathrm{W\,Hz^{-1}}), \tag{7.5}$$

provided $W_e \gg \omega \gg kT$ where ω is the photon energy (see, e.g., Korchak, 1967). Here $F(\nu/\nu_C)$ is a function of the frequency while $\nu_C = (4kT/h)(W_e/W_{e0})^2$ represents the frequency for which the radiation has a maximum. An integration over the frequency yields the total power of the 100 MeV electron $P_C \approx 3 \times 10^{-38} n_{ph} T$. The resulting time-constant for energy decrease is

$$\mathscr{T}_C \approx \frac{W_e}{P_C} \approx 5 \times 10^{26} (n_{ph} T)^{-1}. \tag{7.6}$$

The condition $\mathscr{T}_C > \mathscr{T}_t$ implies that $n_{ph} T < 2 \times 10^{10}$–$2 \times 10^{11}$. This is equivalent to saying that the energy density of the radiation should be $W_{ph} \approx \approx n_{ph} kT \leq 10^{-13}$–$3 \times 10^{-12}$ J m^{-3} corresponding to an equivalent mass density of the photons of $\rho_{ph} \leq 3 \times 10^{-30}$–$3 \times 10^{-29}$ kg m^{-3}.

(4) Finally we consider the time-constant of the electron-positron annihilation. Following Alfvén (1981, p. 99) we find that the cross-section for annihilation of 100 MeV electrons and positrons is about 10^3 times the classical cross-section of the electron. Furthermore, the cross-section for annihilation of protons and antiprotons with energies smaller than ~ 1 GeV is at most only ~ 5 times the classical cross-section of the electron. This means that the time-constant for electron-positron annihilation $\mathscr{T}_{e^-,e^+}$ is at least ~ 200 times the time-constant for proton–antiproton annihilation $\mathscr{T}_{p^+,p^-}$. Since $\mathscr{T}_{p^+,p^-}$ is supposed to be of the same order as $\mathscr{T}_t$ we should expect $\mathscr{T}_{e^-,e^+} \gg \mathscr{T}_t$.

8. Multi-Stage Acceleration

In the previous sections we have discussed single processes where the acceleration of metagalactic matter occurs in one step. Now we shall consider another possibility where the acceleration takes place in several steps. Such a scheme may lead to much higher velocities than in the case of a single-stage acceleration just in the same way as a multi-stage rocket may reach a much higher velocity than a rocket containing only one stage.

Multi-stage acceleration of the metagalaxy may have occurred in several ways. One possibility is that the metagalaxy was first bursted into a number of fairly large pieces due to annihilation. These pieces, which constituted the first-order fragments, were then triggered to burst into smaller pieces – the second-order fragments. The process may have been repeated several times. This scheme is in several ways similar to the fireworks model of the metagalactic evolution suggested by Alfvén (1983).

The maximum velocity in the system after the explosion may be obtained by adding all the velocity increments relativistically. If we for the sake of simplicity assume that the velocity increment of each stage ΔV is constant the maximum velocity of the Nth order fragments is

$$V_N = \frac{\Delta V + V_{(N-1)}}{1 + \Delta V\, V_{(N-1)}}\,, \tag{8.1}$$

where $V_0 = 0$. Figure 2 shows the red shift z corresponding to the maximum velocity that can be reached after N stages for a few different values of the velocity increment $\Delta V/c$. The velocity distribution of the matter is given by a relativistic random walk of N steps in velocity space.

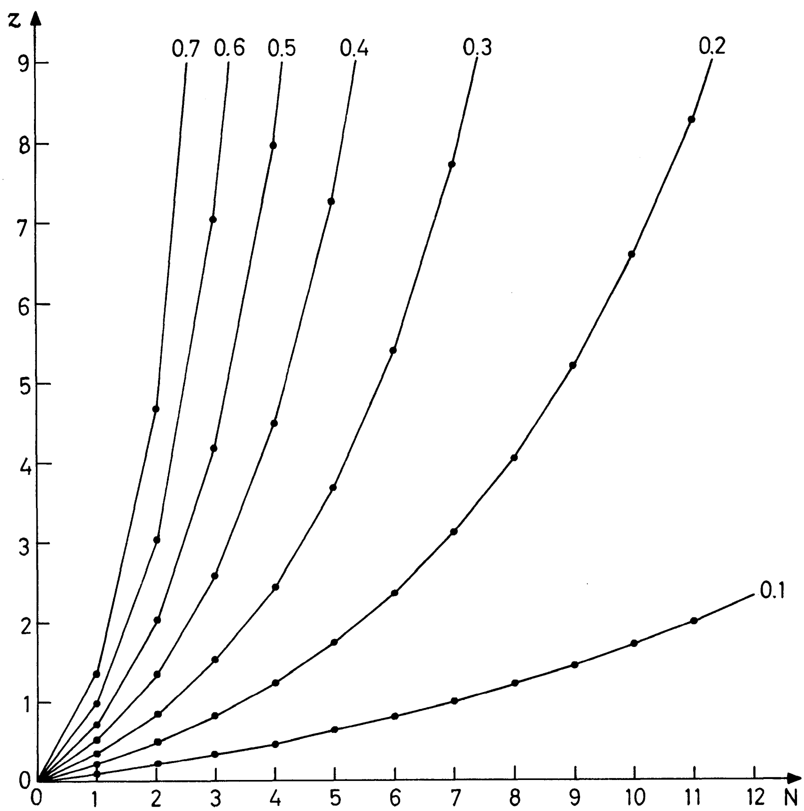


Fig. 2. The red shift z corresponding to the maximum velocity in a three-dimensional, relativistic random walk shown as a function of the number of steps N. The velocity increment of each step is supposed to be constant and equal to 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, and 0.7 times the velocity of light c.

More complicated models than the one above are also conceivable. For instance the velocity increment of each stage need not be constant but may be given by a distribution function.

Another possibility is that the annihilation in the metagalaxy occurred successively in shells (similar to those in an onion) starting from the inner ones. Also in this case an action according to the principle of the multi-stage rocket may have been reached.

As has been pointed out by Alfvén (1983) the fireworks model may lead to a hierarchical structure of the metagalaxy. Such a structure has on observational grounds been advocated by de Vaucouleurs (1970) a long time ago. The hierarchical picture is strengthened by the more recent observations of voids in intergalactic space (Kirshner *et al.*, 1981; de Lapparent *et al.*, 1986).

It is tempting to interpret the foam-like structure observed as a result of annihilation explosions.

9. Conclusions

In the present paper we have studied the dynamics behind the Hubble expansion within the framework of the metagalactic model and without introducing new laws of physics. Many of the ideas presented here originate from or have been inspired by Hannes Alfvén.

Observations show that galaxies and quasars escape from us with velocities corresponding to red shifts up to at least $z = 3.2–4.0$. In order to explain how such large velocities could once have been achieved we have studied in some detail a number of *separate expansion mechanisms* based on the annihilation of matter and antimatter. The mechanisms rely on various pressure forces but also on free fall combined with a time-varying gravitational field.

In one of the mechanisms the driving force behind the expansion is the radiation pressure. It has been shown that in this case there exists an upper limit to the final velocity of $c/2$. Similarly the free-fall mechanism seems not to be capable of accelerating the metagalactic matter to velocities larger than $\sim (1/\sqrt{2})c$.

As regards the two remaining mechanisms treated, which are based on the pressure of magnetic field and electrons-positrons, there does not seem to exist a similar fundamental upper limit to the velocity. Here instead the present metagalactic magnetic field and the relative amount of annihilated mass put constraints on the velocity of expansion.

We have also considered *collective processes of acceleration* which rely on the principle of the multi-stage rocket. Here one or more of the separate mechanisms may contribute in repetition. The velocities resulting may be much larger than the velocities of the separate mechanisms.

The distribution of density in the metagalaxy is of great importance for many of the problems we are studying. In principle we may divide the possible density distributions into two cases. On the one hand the density may be fairly constant and homogeneous (as measured in the local rest frame). On the other hand it may vary considerably over the metagalaxy (it may for instance decrease outwards). In the first case the high red shifts observed for galaxies and quasars can be considered representative for the metagalaxy. In the second case the high red shift objects may merely be regarded as a few sparks ejected especially far out by metagalactic fireworks.

The high expansion velocities or z_B values of the border of the metagalaxy, consistent with the high red shifts observed, imply that the total mass M (including the kinetic energy) is considerably larger than the sum of the rest masses m (see Table II). For $z_B = 4$ we have $M/m \approx 2$ in case of a fairly constant density distribution.

Even if the metagalaxy passes through a state close to the Schwarzschild limit a purely gravitational mechanism does not seem capable of giving more than a marginal contribution to the kinetic energy (see Section 5). For this reason we may apply a simple energy balance to find a lower limit for the initial mass before the turning of the metagalaxy. For the mechanism based on the electron-positron pressure the efficiency with which annihilation energy is transformed into kinetic expansion energy is 0.17 (see Section 7). Then $M/m \approx 2$ leads to the conclusion that the initial mass is at least seven times as great as the final sum of rest masses. The Schwarzschild radius of the initial mass must of course not exceed the present radius of the metagalaxy.

It may be considered unlikely that we are placed very close to the centre of the metagalaxy. Nonetheless red shifts of about $z = 4$ have been observed in approximately two opposite directions. In case of constant density this forces us to conclude that we are within the 4% of the total volume which is closest to the centre. Actually we must be much closer than this, unless the centre-border red shift z_B is very high (see Table III). For instance, if we require that we are in the outer 99% of the total volume z_B must be greater than nine.

Unlike big-bang the homogeneous metagalactic model tells us the value of the well-known quantity Ω_0 once z_B is known (see Table I). In the flat big-bang case Ω_0 is unity. In the metagalactic case we obtain $\Omega_0 \leq 0.06$ using $z_B \geq 4$. The quantity Ω_0 can be calculated from the mass density and the Hubble parameter and it turns out that a value $\Omega_0 \approx 0.06$ is not unreasonable.

It seems to us that the homogeneous case, for which the highest observed red shifts are representative for the metagalaxy, leads to serious difficulties. A highly inhomogeneous model has, in fact, been advocated by Hannes Alfvén for a long time (see, e.g., Section 8). This would mean that the representative red shifts are much smaller (say $z_B \approx 1$). We could then be situated rather far from the centre and the expansion velocities needed could more easily be reached. It should be investigated if the inhomogeneity required is in agreement with observations.

Appendix

A homogeneous expanding cloud of pressure free perfect gas can be treated as a part of a Friedmann model fitted to an outside vacuum space (Klein, 1961). At the spherical boundary between the inner and outer solutions the matter density has a discontinuity. The Friedmann line element can be written (Landau and Lifshitz, 1975, pp. 366–367)

$$ds^2 = a^2(\eta)\{d\eta^2 - d\chi^2 - \sinh^2\chi\,(d\vartheta^2 + \sin^2\vartheta\,d\phi^2)\}\,, \tag{A.1}$$

where

$$a = a_1(\cosh\eta - 1)\,; \tag{A.2}$$

where a_1 is a constant given by

$$a_1 = \frac{4\pi G}{3c^2}\,\rho a^3\,, \tag{A.3}$$

and ρ is the mass density in the local rest frame of the matter. The path of every gas particle fulfills $d\chi = d\vartheta = d\phi = 0$.

The proper time τ of a gas particle is according to (A.1) given by $c\,d\tau = a(\eta)\,d\eta$. The Hubble parameter H is given by

$$H \equiv \frac{1}{a}\frac{da}{d\tau} = c\,\frac{a'}{a^2} = \frac{c}{a_1}\,\frac{\sinh\eta}{(\cosh\eta - 1)^2}\,. \tag{A.4}$$

From this and Equation (A.3) we obtain

$$\alpha \equiv \cosh\frac{\eta}{2} = \mathrm{H}\sqrt{\frac{3}{8\pi G\rho}} \equiv \Omega^{-1/2}\,. \tag{A.5}$$

The quantity Ω was defined in Section 2. In the rest frame for matter the time component of the energy-momentum tensor is $T_0^0 = \rho c^2$, while all other components vanish. If $T_0'^0$ iss the corresponding component in the Schwarzschild coordinate frame it follows immediately that $T_0'^0 = \rho c^2$.

Equation (100.23) in Landau and Lifshitz (1975) then gives

$$M = \frac{4\pi}{3}\,\rho r^3\,, \tag{A.6}$$

where M is the Schwarzschild mass within radius

$$r \equiv a\sinh\chi\,. \tag{A.7}$$

In case that r equals the Schwarzschild radius corresponding to (A.6) we have $2MG/c^2 = r$. Inserting (A.6) and (A.7) as well as (A.2) we obtain $2\chi = \eta$.

All matter in the metagalaxy must be within the Schwarzschild sphere. Hence, we must require

$$\chi < \tfrac{1}{2}\eta\,. \tag{A.8}$$

A light signal which travels towards an observer (chosen as the origin) follows $d\vartheta = d\phi = 0, d\chi = -d\eta$. Suppose it started at an event (χ, η) and arrives at $(0, \eta_0)$. Then

$$\eta_0 = \eta + \chi\,. \tag{A.9}$$

Use this together with (A.8) and we obtain

$$\chi < \tfrac{1}{3}\eta_0\,. \tag{A.10}$$

Using (A.9) to eliminate χ we find

$$\eta > \tfrac{2}{3}\eta_0\,. \tag{A.11}$$

The red shift of the light signal is given by

$$z + 1 = \frac{a(\eta_0)}{a(\eta)} < \frac{a(\eta_0)}{a(\frac{2}{3}\eta_0)} = \frac{\cosh \eta_0 - 1}{\cosh \frac{2}{3}\eta_0 - 1}\,. \tag{A.12}$$

From this and (A.5) we obtain

$$z < \frac{4(\alpha^2 - 1)}{(\beta^{2/3} - \beta^{-2/3})^2} - 1\,, \tag{A.13}$$

where

$$\beta \equiv \alpha + \sqrt{\alpha^2 - 1}\,. \tag{A.14}$$

In Table I we have given α, the corresponding values of Ω and the right-hand side of (A.13). As the maximum measured red shifts are approximately $z = 4$ we see from the table that $\alpha \geq 5$ and $\Omega \leq 0.04$.

TABLE I

α	1	2	3	4	5	6	7	8	9	10
Ω	1.0	0.25	0.11	0.06	0.04	0.03	0.02	0.015	0.012	0.01
z_{max}	1.25	2.03	2.73	3.37	3.97	4.54	5.09	5.61	6.11	6.60

Quite generally the red shift of the light from some galaxy at χ is according to (A.9), given by

$$z + 1 = \frac{a(\eta_0)}{a(\eta_0 - \chi)} = \frac{\cosh \eta_0 - 1}{\cosh(\eta_0 - \chi) - 1}\,. \tag{A.15}$$

This expression can be rewritten in the form

$$\frac{1}{\sqrt{z + 1}} = \sinh \frac{\chi}{2}\left(\coth \frac{\chi}{2} - \coth \frac{\eta_0}{2}\right). \tag{A.16}$$

Now we assume that $\alpha \geq 5$. Then using $\chi \leq \eta_0/3$ we note that we make an error of at most 5% on the right-hand side of (A.16) if we replace $\coth \eta_0/2$ with unity. Doing this we obtain*

* This is the special-relativistic equation (see Section 2).

$$z + 1 \approx e^{\chi}. \tag{A.17}$$

The volume within χ is

$$\mathscr{V} = 4\pi a^3 \int_0^{\chi} \mathrm{d}\chi \sinh^2 \chi = 4\pi a^3 [\tfrac{1}{4} \sinh 2\chi - \tfrac{1}{2}\chi]. \tag{A.18}$$

The rest mass is obtained from this through multiplication with ρ. The part of the total mass M in (A.6) which is rest mass is thus

$$\frac{m}{M} = \frac{3}{2} \frac{\sinh \chi \cosh \chi - \chi}{\sinh^3 \chi}. \tag{A.19}$$

Inserting (A.17) we obtain

$$\frac{m}{M} = \frac{3(z+1)}{z^3(z+2)^3} [(z+1)^4 - 1 - 4(z+1)^2 \ln(z+1)]. \tag{A.20}$$

In Table II m/M is given as a function of z. As we can now see a red shift $z = 4$ we conclude from the table that at least half of the total original rest mass must have converted to kinetic energy.

TABLE II

z	0	1	2	3	4	5	6	7	8	9
m/M	1.0	0.87	0.71	0.59	0.50	0.44	0.38	0.34	0.31	0.28

The quantity

$$\zeta \equiv \frac{2MG}{rc^2},$$

tells us how far the metagalaxy is from the Schwarzschild limit.

If $\zeta \ll 1$ we can treat space-time as flat and the red shift as pure Doppler shift. Using (A.2), (A.3), and (A.6) we obtain

$$\zeta = \frac{\cosh 2\chi - 1}{\cosh \eta - 1} = \frac{\cosh 2\chi - 1}{\cosh(\eta_0 - \chi) - 1}. \tag{A.21}$$

This can be written in terms of z and Ω_0 (the present day Ω) by using (A.16) and (A.15). The condition $\zeta \ll 1$ means that $2\chi \ll \eta$. Comparing this to Equation (A.8) we realize that Table I may be used here. If, e.g., the maximum value of z is $z = 4$ special relativity can be used if $\Omega_0 \ll 0.04$. This would presumably only mean that time had passed far beyond the stage when the explosion could be seen. At the time of the explosion the metagalaxy may very well have been close to the Schwarzschild limit. For a discussion of Table III see Section 3.

TABLE III

z_c	0.5	1	2	4	8	16	32
z_B	6.5	9	14	24	44	84	164
k	260	93	48	33	28	26	25.3
$1/k$	0.004	0.011	0.021	0.030	0.036	0.038	0.039

References

Alfvén, H.: 1981, *Cosmic Plasma,* D. Reidel Publ. Co., Dordrecht, Holland.

Alfvén, H.: 1983, *Astrophys. Space Sci.* **89**, 313.

Alfvén, H. and Fälthammar, C.-G.: 1963, *Cosmical Electrodynamics*, 2nd ed., Clarendon Press, Oxford.

Alfvén, H. and Klein, O.: 1962, *Arkiv Fysik* **23**, 187.

Bonnevier, B.: 1964, *Arkiv Fysik* **27**, 305.

Burstein, D., Davies, R. L., Dressler, A., Farber, S. M., Lynden-Bell, D., Terlevich, R., and Wegner, G.: 1986, in B. F. Madore and R. B. Tully (eds.), *Galaxy Distances and Deviations from the Universal Expansion,* D. Reidel Publ. Co., Dordrecht, Holland, p. 123.

Cowling, T. G.: 1957, *Magnetohydrodynamics*, Interscience Publ., New York.

de Lapparent, V., Geller, M. J., and Huchra, J. P.: 1986, *Astrophys. J.* **302**, L1.

de Vaucouleurs, G.: 1970, *Science* **167**, 1203.

Ekspong, A. G., Yamdagni, N. K., and Bonnevier, B.: 1966, *Phys. Rev. Letters* **16**, 664.

Foster, J. and Nightingale, J. D.: 1979, *A Short Course in General Relativity,* Longman, London.

Gott, III, R. J., Gunn, J. E., Schramm, D. N., and Tinsley, B. M.: 1974, *Astrophys. J.* **194**, 543.

Kirschner, R. P., Oemler, A., Jr., Schechter, P. L., and Shectman, S. A.: 1981, *Astrophys. J.* **248**, L57.

Klein, O.: 1953, in *Les processus nucléaires dans les astres,* Société Royale des Sciences de Liège, p. 42.

Klein, O.: 1961, in *Werner Heisenberg und die Physik unserer Zeit,* Friedr. Vieweg und Sohn, Braunschweig.

Korchak, A. A.: 1967, *Soviet Astron.* **11**, 258.

Landau, L. D. and Lifshitz, E. M.: 1975, *The Classical Theory of Fields,* Pergamon, Oxford.

Laurent, B. E. and Söderholm, L.: 1969, *Astron. Astrophys.* **3**, 197.

Rubin, V., Ford, W. K., Thonnard, N., and Roberts, M. S.: 1976, *Astron J.* **81**, 687.

TEARING DOWN DISCIPLINARY BARRIERS*

JUAN G. ROEDERER
Geophysical Institute, University of Alaska, Fairbanks, Alaska, U.S.A.

(Received 24 November, 1987)

"He who understands nothing but chemistry does not truly understand chemistry either."

Georg Christoph Lichtenberg (1742–1799)

Abstract. Profesor Hannes Alfvén's life-long battle against scientific narrow-mindedness and parochial approaches to the solution of scientific problems is well known and deeply appreciated by this author. In this article the new interdisciplinary trends in science are critically examined and the psychological impacts of crumbling disciplinary barriers on the participating scientists are analyzed. Several examples of interdisciplinary research programs are discussed and some thoughts on the structural reform of scientific organizations, agencies, and universities needed to face these trends are given.

1. Introduction

In recent years the solutions of scientific problems in practically all branches of science have been demanding from the participating researchers an increased specialization on one hand and the pursuit of an increasingly interdisciplinary approach on the other. This paradoxical development of increasing convergence and specificity of the basic scientific questions with a concomitantly increasing global scope and multiplicity of the intervening disciplines is beginning to have a considerable impact on the conduct of research programs, on the structure of research institutions, on the internal organization of funding agencies and professional scientific societies, and on university curricula and training methods. In addition, it is beginning to have effects on the scientific community *per se*, effects that range from purely academic to eminently psychological.

In this article, I will discuss the new interdisciplinary trends in science, analyze the psychological impact of crumbling disciplinary barriers on the scientific community, and venture a few comments on what I believe might be necessary in the near future regarding structural reform of scientific organizations, agencies, and universities to face these trends.

Let me begin with a discussion of what I really mean by 'interdisciplinary trends'. In thinking about this subject, I discovered three different kinds of trends, three different causes for disciplinary unification, and three different types of psychological effects on the participating scientists.

* Reprinted with permission from EOS, American Geophysical Union.
† Paper dedicated to Professor Hannes Alfvén on the occasion of his 80th birthday, 30 May 1988.

Astrophysics and Space Science **144** (1988) 659–667.

2. Interdisciplinary Trends

2.1. The case of mutually interacting systems

In any branch of science, but most especially in the natural sciences, systems are considered that in first approximation can be studied in isolation. Interactions with other surrounding systems are taken to be one-way actions rather than interactions. They are given external input effects from the outside, and/or resulting output effects on the outside, or no effects at all. For instance, Mach's approach to Newtonian mechanics begins with a system of two interacting mass points in complete isolation from the rest of the Universe. A traditional meteorologist deals with the atmosphere as a system subject to given outside forcing agents and outside sources and sinks of energy and mass. A psychophysicist deals with the transformation of external physical stimuli into neural signals and associated sensations, and the traditional neuropsychologist deals with brain function as elicited by neural input signals from the environment and the body and with behavioural reactions as the main 'macroscopic' output. Quite generally, what is one discipline's output is another discipline's boundary condition or input.

However, the real Universe is more complicated, and we are obliged to consider interactions between natural systems. It is when we focus on the mechanisms of interaction *per se* that the first kind of interdisciplinary trend is likely to occur. It is most conspicuous whenever we consider several mutually interacting systems as components of one single, global 'supersystem'.

2.2. The case of a single system governed by processes pertaining to different disciplines

In the consideration of the intrinsic behaviour of a given system, a single discipline often dominates the study of the principal processes governing the system's behaviour. Traditional cell biology used to deal mainly with biochemical properties of and interactions between the components of the cell. Traditional oceanography used to be partitioned into more or less independent studies of either the physical, or the chemical, or the biological components of oceanic waters. Traditional seismology focused on geological structures and their geological characteristics.

Again, the real Universe is more complicated. In almost any system that the human mind defines and selects as a target of study, processes occur that pertain to a whole spectrum of disciplines. Physics and molecular biology have invaded the cell; the study of the ocean today requires a unifying, interdisciplinary approach; and solid state physics, mineral physics, and Hamilton–Jacobi wave mechanics are invading seismology. This represents 'interdisciplinarification' of the second kind.

2.3. The case of one phenomenon or process common to several different systems

Plasma processes occur in the upper atmosphere, in interplanetary and galactic space, on stars, in the laboratory, and in nuclear detonations. Nonlinear dynamics governs phenomena that occur in the atmosphere, the oceans, the brain, ecosystems, and the social behaviour of human masses. Thus a riot, a sudden memory recall, and the release of a Gulf Stream eddy all have one aspect of nonlinear dynamics in common. Scientific constituencies appertaining to quite different disciplinary areas are thus being unified through the common thread of one subdiscipline, process, or algorithm. The communication and interaction promoted by such common thread among diverse constituencies represents the third kind of interdisiplinary trend.

3. Unifying Factors

What factors contribute to the trends toward interdisciplinary modes of science? The first factor, all too obvious, is related to the following. Science, in order to progress, must proceed in a reductionist mode, at first simplifying and approximating until a first-order model of the real system under study is constructed that is quantitatively 'tractable' and testable. Only after this has been accomplished can the approximations and the models be refined – and it is in this quantitative refinement that disciplinary proliferation invariably occurs.

The second factor contributing to interdisciplinary unification is given by the role played by two disciplines (in addition to the ubiquitous field of mathematics) that are pervading with ever-increasing arrogance all other scientific disciplines. They are physics and 'informatics' (or 'information theory' or 'information science'). This is beginning to have a profound effect on the development of research, particularly in the life sciences.

The third unifying factor is given by new universal techniques and methods that are of common application to many different disciplines. The use of digital computers, data processing, numerical modeling, and computer simulation is beginning to bind many disparate areas together, and so is the use of orbital platforms, tomographic techniques, radionuclide tracing, lasers, nonlinear dynamics, and catastrophic theory, etc. Progress in any given discipline is increasingly controlled by the dexterity, skill, and ingenuity with which such 'universal' techniques and methods are being applied.

4. Psychological Effects

I now come to the psychological impact that the trends toward interdisciplinary modes of research may have on individual members of the scientific community. I see three main effects, each related to some basic instincts of human beings (and other animals): territorial dominance, greed, and fear of the unknown.

Scientists from different disciplines have different customs; they speak different scientific dialects or lingos; they use different approaches in research– even their ways of rationalizing may be different. Some traditional disciplines are eminently descriptive, others are more quantitative; some are relevance-oriented, others are of purely intellectual value. When a scientist from one discipline starts delving into another, some members of the community of this other discipline may develop a feeling of invasion of privacy by this 'alien' scientist.

Some scientists profess disciplinary parochialism: while they do recognize the existence of interdisciplinary links with their subject of research, they tend to resist interdisciplinary inquiries into their own territory on grounds that such inquiries would only involve second-order processes, negligible energies, small perturbations – or they may argue that no interactive processes could *a priori* be imagined as potential causes of interdisciplinary linkage. In many instances, such parochialism is founded on the fear that intrusion from other disciplines would compete unfairly for limited financial resources and thus diminish their own opportunities for research.

Finally, a third behavioural manifestation is that of methodological conservatism on the basis of the rather natural resistance to learning 'new things' at a mature age. This syndrome is particularly acute in scientific communities brought up in school systems like ours in the U.S.A., where, for instance, one can become a biologist without ever having been seriously exposed to physics, and *vice versa*.

5. Examples of 'Interdisciplinarification'

A few examples (taken from my own experience) are in order.

When the systematic study of energy, momentum, and mass transfer from the Sun to the Earth and the effects of their variability on the terrestrial environment became

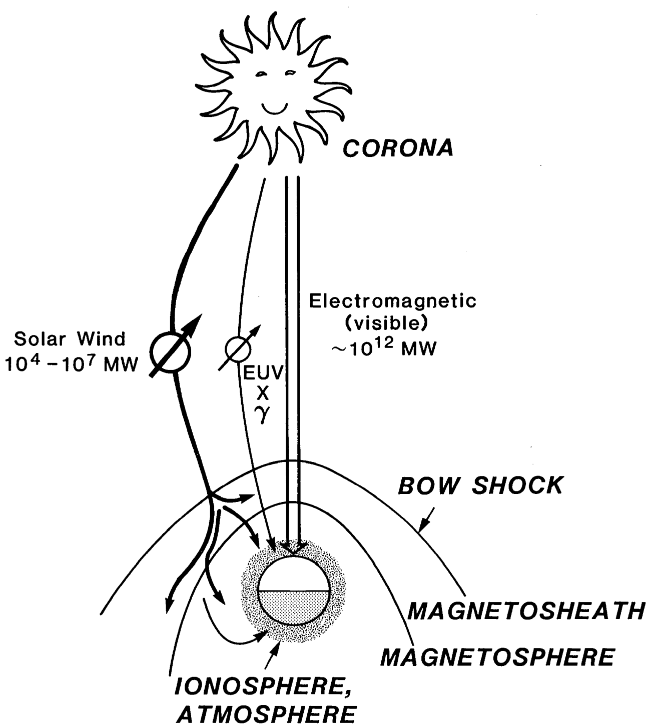


Fig. 1. Example of interdisciplinary research: study of principal components and energy channels of the solar-terrestrial system (from Roederer, 1980).

a discipline in its own right – solar-terrestrial physics, or STP – it quickly became evident that the various intervening regions and energy channels (Figure 1) had to be considered as parts of one 'supersystem' of mutually interacting components (e.g., Roederer, 1980). Astrophysicists, space physicists, and ionospheric and atmospheric scientists had to talk to each other and work together. To a large extent, plasma physics is a unifying element common to all parts of this supersystem, and so are the usage of space platforms and the techniques of data handling. The emergence of STP led to important changes in the International Council of Scientific Unions (ICSU) (e.g., the creation of the Scientific Committee on STP; the reorganization of both the International Association of Geomagnetism and Aeronomy and Union Radio Scientifique Internationale). Occasional friction between solar-terrestrial physicists and atmospheric scientists persists to this day whenever it comes to some common territory in what is vaguely called the 'upper atmosphere'. In addition, the evidence provided by STP that the medium physically tied to planet Earth extends outwards tens and hundreds of thousands of kilometers, thus expanding the concept of the geosphere far beyond the aerodynamically navigable atmosphere, is still subliminally (and sometimes not so subliminally) ignored by many geophysicsts.

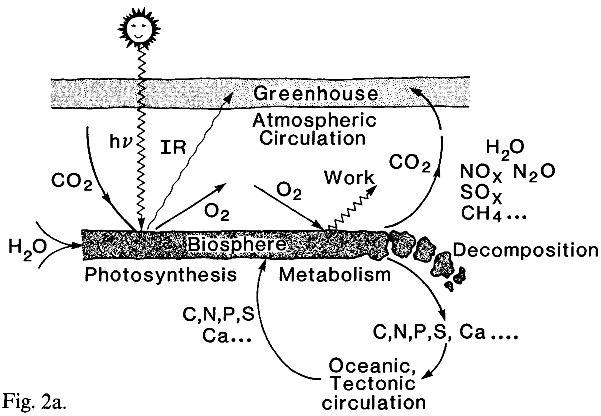


Fig. 2a.

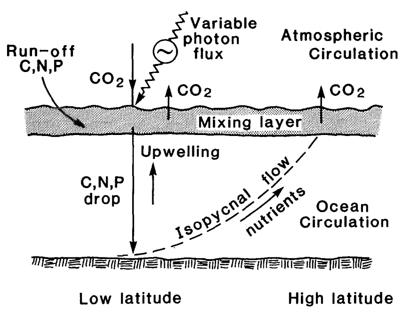


Fig. 2b.

Fig. 2. Examples of interdisciplinary research: interacting systems in biogeochemical cycles (from Roederer, 1985). (a) Basic biological processes controlling global cycles. (b) The 'biological pump', in which the biomass in the upper ocean layer participates in the global cycles of C, N, and P, and controls the strength of oceanic CO_2 sinks and sources, thereby transducing long-term variations of solar flux into variations of global climate.

The next example, involving parts that interact even more strongly, is provided by the major biogeochemical cycles. The quantitative understanding of the processes of interaction between biosphere and geosphere (Figures 2(a) and 2(b)) requires not only interdisciplinary studies but research on spatial scales ranging from molecular to global, time-scales ranging from geologic to atomic, energy scales ranging from major cosmic power sources to small instability-triggering perturbations, and mass scales ranging from that of the ocean to the trace gases emitted by ants and ruminants. In other words, in the study of biogeochemical cycles there is not only a fusion of disciplines but also a fusion of scales or dimensions. In addition, nonlinear processes abound, and the requirements for a strong mathematical physics background are increasing rapidly. For instance, in quantitative ecosystem studies the traditionally descriptive ecologists of yesteryear must yield to scientists familiar with Thom's catastrophic theory and non-linear dynamics.

At this very time, ICSU is planning a major international program for the 1990's, called the International Geosphere-Biosphere Program (IGBP). It has been proposed as a cross-disciplinary effort to obtain a truly comprehensive, quantitative understanding of the complicated terrestrial 'machine', the functions of and interactions between its parts, and the major geophysical and biogeochemical cycles by which it is driven (Malone and Roederer, 1985). The IGBP would consist of an aggregate of sharply focused research programs that share in common a global view of the earth system that emphasizes the connectedness of all intervening parts. These programs ultimately should be designed to assess trends and to anticipate natural and anthropogenic global change over a 50–100 year time-scale.

The major regions relevant to the study are sketched in Figure 3. This figure depicts what we may call the 'core system', consisting of the biosphere, lower atmosphere,

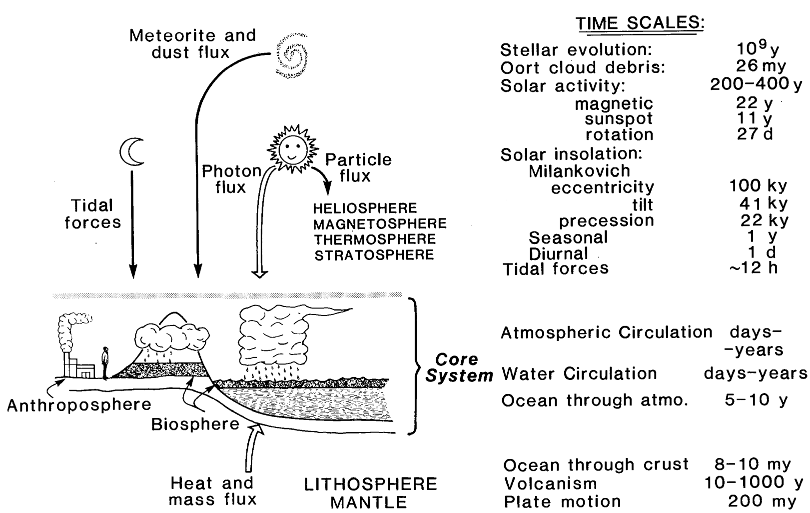


Fig. 3. Example of interdisciplinary research: the study on global change, or International Geosphere-Biosphere Program (IGBP). Shown are the relevant regions and time-scales (from Roederer, 1985).

hydrosphere (including cryosphere), and soil. In this core system of the terrestrial machine, the atmosphere plays the role of the working fluid, the oceans that of fundamental energy reservoirs, and the biosphere that of a global regulator. Some relevant time-scales are indicated. The core system is sandwiched between the 'outer' geosphere, or Earth-space, and the 'inner' geosphere, or lithosphere. For a study of the core system that only focuses on the intrinsic time-scales involved (say, up to a few decades), it would be enough to consider the outer and inner layers and related major energy sources as given boundary conditions. However, to achieve a quantitative understanding of the global terrestrial system and its full-scale, long-term behaviour, its instabilities, its potential for multi-equilibria states, and the potential impact of natural and anthropogenic hazards on human system on Earth and in orbit, the boundaries of the core system and all related time and energy scales must be included in the study of what we may call the total geosphere-biosphere system.

The discussions within the U.S. Academy of Sciences and ICSU as to which of these approaches to promote afforded a unique opportunity for a 'global behavioural response experiment', with the participating scientists (this writer included) as the experimental subjects and with disciplinary parochialism, methodological conservatism, and defense of home turf as the parameters to be measured.

The last example to be given here deals with the pervasive role of physics and the unification provided by universal modeling techniques. The achievement of a quantitative understanding of brain function, in particular of neural information processing, storage, and recall, will no doubt become one of the greatest challenges to science during the next decades. Very little is known about this subject today. During the dynamic phase of processing, neural information is encoded in the form of electric signals, i.e., the analog post-synaptic potentials and the digital action potentials (there is also a concurrent, much slower, chemical information system that plays a major role in controlling the affective states of brain function). The problems are that there are approximately 10^{10} neurons in the human brain, that the detailed anatomy of the neural system is not yet fully known, that the synaptic interconnections between neurons change with 'use' (thus encoding long-term memory information), that there is no spatio-temporal continuity in the distribution of neural activity, and that the entire system works in a parallel processing mode with distributed memory storage involving the brain as a whole.

Physicists and some of the more enlightened neuropsychologists are beginning to study and model brain function in analogy to some specific physical systems. For instance, Fresnel holography represents an excellent model offering insights into how the associative recall mechanism (remembering something by association) might work. Figure 4 schematically shows how information on two simultaneously presented objects is stored (in the case of the brain, e.g., the face and name of a person, respectively) and how, later, the presence of one object (presented in the right position under the right conditions) elicits the image of the other (e.g., the name of a person elicits the image of a face, or *vice versa*). The lack of comprehension of the scientific meaning, the limitations, and the potentialities of such modeling by many traditional neurologists and

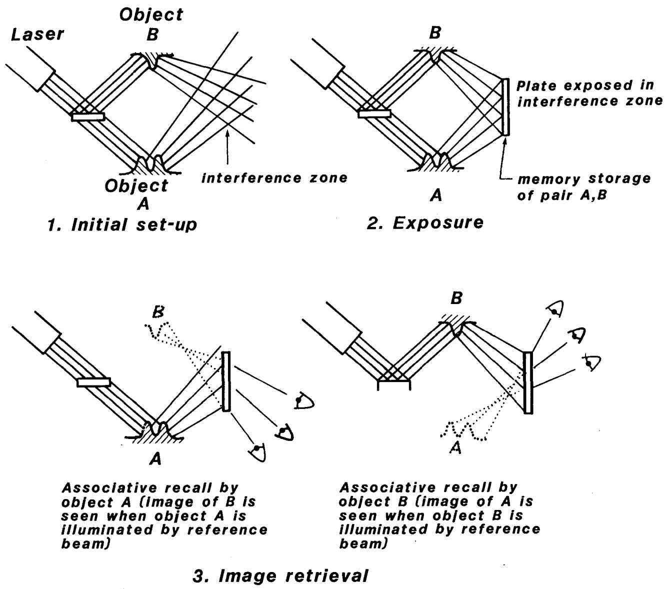


Fig. 4. Example of interdisciplinary research: modeling of brain function with Fresnel holography. The figure depicts basic processes that afford analog modeling of the processes of associative memory recall (from Roederer, 1987).

psychologists is staggering. (I have heard one exclaim 'how ridiculous, the brain does not operate with laser beams inside'!) Yet computer or analog (electronic) models of simple neural circuitry are being developed that accomplish quantitatively – albeit in a very primitive approximation – many basic associative recall functions (e.g., Kohonen, 1984).

6. Needs for Institutional Reform

What must we do, now that we are confronted with the irreversible course of increasing interdisciplinary trends in science? We are ill prepared for it. Our students are ill prepared for it. Our governmental organizations are ill prepared for it. The international scientific establishment is ill prepared for it.

When we take a look at our university and secondary school curricula, we see a dismal picture. We have courses like 'Physics for Sculptors', 'Chemistry for Homemakers', or 'Biology for Mountain Climbers'. Yet what we really need are de-'Mickey Moused', back-to-basic courses of physics, chemistry, and biology for everybody. The key point is that in each course, we must take pains to incorporate the links to other disciplines and build in examples of interdisciplinary applications. We have to rewrite syllabi of recitation classes and develop new laboratory classes, and we have to rewrite entire textbooks. On the other hand, there is a need for academic institutions to modify their faculty and research staff evaluation criteria to reflect the different expectations of those who are hired to work in the interdisciplinary mode.

Funding agencies must develop a matrix structure in the multi-dimensional space of traditional disciplines so that responsibilities for interdisciplinary problems can be clearly assumed and effectively shared. There is a need for agencies to modify their proposal evaluation criteria in order to give interdisciplinary research projects a fair treatment and to promote a more vigorous development of cross-disciplinary activities.

Finally, an international scientific organization such as ICSU must take a broad look at its own structure and subdivisions (in which interdisciplinary bodies already are beginning to proliferate, to the chagrin of unions that are oriented toward more traditional disciplines), face the reality of current trends, declare itself in a state of revolution, and strive toward recrystallization in a new state that exhibits less ceremonial traditionalism and more interdisciplinary activism.

References

Kohonen, T.: 1984, *Self-Organization and Associative Memory*, Springer-Verlag, New York.

Malone, T. F. and Roederer, J. G. (eds.), 1985, *Global Change*, Cambridge University Press, Cambridge.

Roederer, J. G.: 1980, 'The Solar Wind-Magnetosphere-Ionosphere System: An Overview', in C. S. Deehr and J. A. Holtet (eds.), *Exploration of the Polar Upper Atmosphere*, D. Reidel Publ. Co., Dordrecht, Holland.

Roederer, J. G.: 1985, 'Global Change: The IGY of the 1990's', in E. Bierly (ed.), *Report to the Community: Where are We Going in the Next Decade*, Am. Meteorol. Soc., Boston, Mass. (in press).

Roederer, J. G.: 1987, 'Neuropsychological Processes Relevant to the Perception of Music', in R. Spintge and R. Droh (eds.), *Music in Medicine*, Springer-Verlag, Berlin.