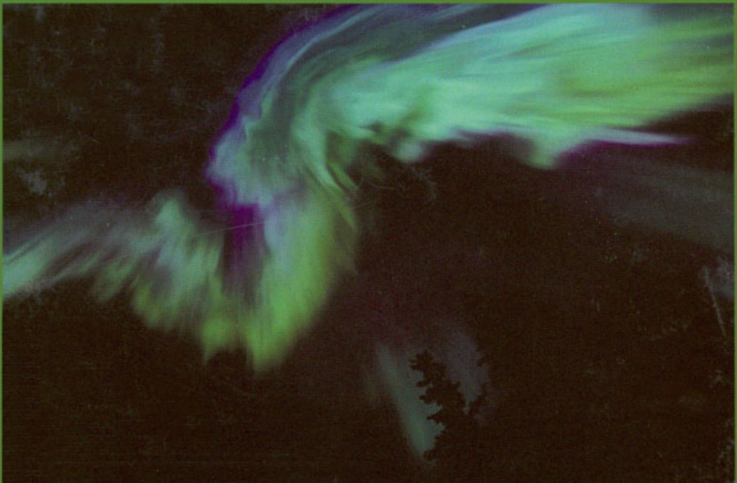
ASTROPHYSICS AND
SPACE SCIENCE LIBRARY

EXPLORING THE SECRETS OF THE AURORA

SYUN-ICHI AKASOFU



KLUWER ACADEMIC PUBLISHERS

EXPLORING THE SECRETS OF THE AURORA

ASTROPHYSICS AND SPACE SCIENCE LIBRARY

VOLUME 278

EXPLORING THE SECRETS OF THE AURORA

by

SYUN-ICHI AKASOFU
International Arctic Research Center,
University of Alaska, Fairbanks, U.S.A.

KLUWER ACADEMIC PUBLISHERS
NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

eBook ISBN: 0-306-47970-2
Print ISBN: 1-4020-0685-3

Created in the United States of America

Visit Kluwer Online at: http://kluweronline.com
and Kluwer's eBookstore at: http://ebooks.kluweronline.com

Table of Contents

Source of Figures .... **VII**
Preface .... **XVI**
Prologue .... **XIX**
Chapter 1 .... **1**
Search For The Unknown Quantity In The Solar Wind **1**
1.1 Solar Corpuscular Streams *1*
1.2 The Chapman-Ferraro Theory *6*
1.3 The Solar Wind *11*
1.4 Interplanetary Shock Waves *12*
1.5 The Modern Interpretation of the Chapman-Ferraro Theory *13*
1.6 The Main Phase of Geomagnetic Storms and the Ring Current *13*
1.7 Variety of the Development of Geomagnetic Storms *17*
1.8 Unknown Quantity *21*
1.9 The ε Parameter *24*
1.10 The Directly Driven and Unloading Components *35*
1.11 The Open Magnetosphere *36*
Chapter 2 .... **41**
Confronting Paradigms: Aurora Research During The Early Space Age **41**
2.1 My Earliest Association with the Aurora *41*
2.2 The Auroral Zone to the Auroral Oval *42*
2.3 Significance of the Auroral Oval *51*
2.4 Auroral Substorms: Fixed Pattern to Substorm Pattern *54*
2.5 Auroral Storms *68*
2.6 Auroral Rays *70*
2.7 Thickness of an Auroral Curtain *70*
2.8 Auroral Kilometric Radiation *73*
2.9 Auroral Observation at the South Pole *74*
2.10 Auroral Spectra as Tools for Detecting Extraterrestrial Life *74*
2.11 Emperor Showa and the Aurora *76*
2.12 An Exciting New Development *77*
Chapter 3 .... **89**
Realizing The Dream Of Our Pioneers: Polar Magnetic Substorms And The Associated Current System **89**
3.1 The Three-Dimensional Current System *90*
3.1.1 The Uniqueness Problem 90
3.1.2 Chapman's Equivalent Current System 90
3.1.3 Birkeland-Alfvén Model 94
3.2 Alaska Meridian Chain of Magnetic Observatories *97*
3.3 The IMS Meridian Chains of Observatories *99*
3.4 Boström's Current Loops *100*
3.5 Nikolsky's Spiral *109*
3.6 Substorm Onset *112*
3.7 Changes of Magnetic Energy in the Magnetotail *115*
3.8 Storm-Substorm Relationship *119*
3.9 Geomagnetic Indices *120*
3.10 Publication of Solar-Terrestrial Physics from Oxford University Press *121*
3.11 Summary of Chapters 1, 2, and 3 .... *124*

Chapter 4: **133**
Is The Earth's Dipole Really Off-Centered And Inclined?: Planetary Magnetic Fields **133**
4.1 Introduction *133*
4.2 Rotation of the Solar Magnetic Dipolar Field on the Source Surface *134*
4.3 Large Inclination and Eccentricity of the Dipole-like Field of Uranus and Neptune *139*
4.4 Is the Earth's Dipole Actually Inclined with Respect to the Rotation Axis? *142*
4.5 Does the Main Dipole of the Geomagnetic Field Rotate during the Reversals? *143*
4.6 Heliospheric Current Sheet *145*
Chapter 5 .... **149**
Myth Of The Emerging Flux Tubes: Sunspots And Solar Flares **149**
5.1 Introduction *149*
5.2 Emerging Magnetic Flux Tubes *150*
5.3 Energy Source *153*
5.4 Sunspots *156*
5.5 Force-free Fields and Solar Flares *158*
5.6 Magnetic Reconnection *162*
Chapter 6 .... **165**
Synthesis Of The Four Major Disciplines: Predicting Geomagnetic Storms As A Space Weather Project **165**
6.1 Introduction *165*
6.2 Modeling the Background Solar Wind Flow *166*
6.3 Parameterizing Solar Events *170*
6.4 Modeling the Propagation of Shock Waves *171*
6.5 Detecting Shock Waves by IPS *173*
6.6 Estimating the Velocity, Density, and IMF at the Earth *177*
6.7 Characterizing Geomagnetic Storms *177*
6.8 Predicting Ionospheric Effects *179*
6.9 Effects on Power Transmission Lines and Oil/Gas Pipelines *180*
6.10 Missing Links *181*
6.11 High-speed Streams from Coronal Holes *185*
6.12 A New Example *186*
6.13 Summary *192*
Chapter 7 .... **197**
Beyond The Inner Heliosphere: The Magnetic Field Structure Of The Outer Heliosphere: A Three-Dimensional Model **197**
Epilogue **204**
Further Readings: Mostly Historical **213**
Name Index **217**
Index **223**

A Scenario on Substorm Onset (Note added on March 1, 2002) **227**

SOURCE OF FIGURES

p. XIX	Carrington, R.C., Mon. Not. Roy. Astronom.Soc., **20,** 1860.
p. XXI	Maunder, E.W., courtesy of Eather, R.H.
p.XXIII-XXV	Akasofu, S.-I., *Magnetospheric Substorms,* ed. by Kan, J.R., et al., AGU Monograph, 64, p. 3, Washington D.C. 1991.
p. XXVI	Akasofu, S.-I. and Chapman, S., *Solar-Terrestrial Physics,* p. 542, Oxford University Press, Oxford, 1972.
p. XXVII	Birkeland, K., courtesy of University of Oslo.
p. XXVII	Chapman, S., courtesy of Geophysical Institute.
p. XXIX	Alfvén, H., courtesy of Alfvén, H.
p. XXIX	Van Allen, J.A., courtesy of Van Allen, J.A.
p. XXX	Dungey, J.W., Phys. Rev. Lett, **6**, 47, 1961.
p. XXXII	Hakamada, K. and Akasofu, S.-I., Space Sci. Rev., **31**, 3, 1982.
p. XXXII	Akasofu, S.-I. and Fry, C.D., J. Geophys. Res., **91**, 13,679, 1986.
p. XXXIII	SOHO/(Instrument) consortium. SOHO is a project of international cooperation between ESA and NASA.
p. XXXIII	Akasofu, S.-I. and Hakamada, K., Geophys. Res. Lett., **10**, 1577, 1983.

CHAPTER 1

Figure 1.1 a	Courtesy of University of Oslo.
Figure 1.1b	Courtesy of Fukushima, N.
Figure 1.2a	Courtesy of University of Oslo.
Figure 1.2b	McCracken, K.G., Rao, U.R., and Shea, M.A., M.I.T. Tech. Rep., 77, 2.
Figure 1.3a,b	Chapman, S., and Bartels, J., *Geomagnetism,* Oxford University Press, 1940.

Figure 1.3c,d Chapman, S., and Ferraro, V.C.A., Terr. Magn., **36**, 77, 1931.

Figure 1.4 Roederer, M, see also Akasofu, S.-I. and Covey, D.N., Planet. Space ScL, **28**, 757, 1980.

Figure 1.5 Akasofu, S.-I.

Figure 1.6 Störmer, C, *The Polar Aurora,* Oxford University Press, Oxford, 1955 (quoted by Van Allen, J.A., in Proceedings of the General Assembly, Berkeley, 1961, p. 99, Academic Press, London, 1962).

Figure 1.7 Akasofu, S.-I., Cain, J.C., and Chapman, S., J. Geophys. Res., **66**, 4013, 1961.

Figure 1.8a Akasofu, S.-I. and Chapman, S., J. Geophys. Res., **68**, 125, 1963.

Figure 1.8b Akasofu, S.-I. and Chapman, S., *Solar-Terrestrial Physics,* Oxford University Press, Chapter 8, 1972.

Figure 1.9 Akasofu, S.-I., Space Sci. Rev., **28**, 121, 1981.

Figure 1.10 Akasofu, S.-I., Hones, E.W., Jr., Bame, S.J., Asbridge, J.R., and Lui, A.T.Y., J. Geophys. Res., **78**, 7257, 1973.

Figure 1.11 Akasofu, S.-I., Planet. Space Sci., **12**, 801, 1964 (see also S.-I. Akasofu and S. Chapman, *Solar-Terrestrial Physics,* Chapter 8, Oxford University Press, Oxford, 1972.

Figure 1.12 Akasofu, S.-I., *Physics of Magnetospheric Substorms,* Chapter 5, D. Reidel Pub. Co., Holland, 1977.

Figure 1.13a,b Akasofu, S.-I., EOS, **66**, 465, 1985.

Figure 1.13c Akasofu, S-I., EOS, **70**, 529, 1989.

Figure 1.13d Akasofu, S.-I., *Dynamics of the Magnetosphere,* p. 447, D. Reidel Pub. Co., Dordrecht, Holland, 1980.

Figure 1.14 Akasofu, S.-I., Space Sci. Rev., **28**, 121, 1981.

Figure 1.15. Sun, W., Xu, W.-Y., and Akasofu, S.-I., J. Geophys. Res., **103**, 11,695, 1998.

Figure 1.16 Vampola, A.L., J. Geophys. Res., **76**, 36, 1971.

Figure 1.17 Akasofu, S.-I., Lin, W.C., and Van Allen, J.A., J. Geophys. Res., **68**, 5327, 1963.

CHAPTER 2

Figure 2.1 Courtesy of Yale University.

Figure 2.2 Loomis, E., Amer. J. Sci. and Arts, **30**, 89, 1860.

Figure 2.3 Akasofu, S.-I., *Aurora and Airglow,* ed. by McCormack, B.M., p. 267, Reinhold Pub. Co., New York, 1967.

Figure 2.4 Carlheim-Gyllensköld, N., Observations faites an Cap Thordsen, Spitzberg, vol. II, Stockholm, 1879.

Figure 2.5 Courtesy of University Tromsø.

Figure 2.6a,b Courtesy of Geophysical Institute, University of Alaska.

Figure 2.7a Akasofu, S.-I., Space Sci. Rev., **16**, 617, 1974.

Figure 2.7b Akasofu, S.-I.

Figure 2.8a Akasofu, S.-I.

Figure 2.8b Snyder, A.L., Ph.D. Thesis, University of Alaska, 1972.

Figure 2.8c Snyder, A.L. and Akasofu, S.-I., J. Geophys. Res., **77**, 3419, 1972.

Figure 2.9a Courtesy of Gassmann, G.

Figure 2.9b Buchau, J., Whalen, J.A., and Akasofu, S.-I., J. Geophys. Res., **75**, 7147, 1970.

Figure 2.10 Kamide, Y. and Akasofu, S.-I., J. Geophys. Res., **81**, 3999, 1976.

Figure 2.11 Iijima, T. and Potemra, T.A., J. Geophys. Res. **83**, 599, 1978.

Figure 2.12a Courtesy of Anger, C.

Figure 2.12b Courtesy of Anger, C.

Figure 2.13 Courtesy of University of Alaska.

Figure 2.14	Akasofu, S.-I.
Figure 2.15	Akasofu, S.-I.
Figure 2.16a	Akasofu, S.-I., Space Sci. Rev., **16**, 617, 1974.
Figure 2.16b	Akasofu, S.-I., J. Geophys. Res., **68**, 1667, 1963.
Figure 2.17	Akasofu, S.-I., Planet. Space Sci., **12**, 273, 1964.
Figure 2.18a	Ames Research Center, NASA (see also Akasofu, S.-I., Planet. Space Sci., **16**, 365, 1968).
Figure 2.18b	Courtesy of Geophysical Institute, University of Alaska.
Figure 2.18c	Courtesy of Geophysical Institute, University of Alaska.
Figure 2.19	Courtesy of Geophysical Institute, University of Alaska.
Figure 2.20	Anger, C.D., Lui, A.T.Y. and Akasofu, S.-I., J. Geophys. Res., **78**, 3020, 1973.
Figure 2.21	Akasofu, S.-I., Planet. Space Sci., **24**, 1349, 1975.
Figure 2.22a	Craven, J.D., Kamide, Y., Frank, L.A., Akasofu, S.-I., and Sugiura, M., *Magnetospheric Currents,* ed. by Potemra, T.A., Geophysical Monograph, 28, AGU, Washington D.C., 1983.
Figure 2.22b.	Courtesy of Parks, G.
Figure 2.23	Akasofu, S.-I., Space Sci. Rev., **19**, 169, 1976.
Figure 2.24a	Akasofu, S.-I. and Chapman, S., Planet. Space Sci., **24**, 785, 1962.
Figure 2.24b	Akasofu, S.-I. and Chapman, S., Planet. Space Sci., **24**, 785, 1962.
Figure 2.25	Wagner, J.S., Sydora, R.D., Tajima, T., Hallinan, T., and Akasofu, S.-I., J. Geophys. Res., **88**, 8013, 1983.
Figure 2.26	Stenbaek-Nielsen, H.C., Hallinan, T.J., Osborne, D.L., Kimball, J., Chaston, C., McFadden, J., Delory, G., Temerin, M., and Carlson, C.W., Geophys. Res. Lett., **25**, 2073, 1998.
Figure 2.27	Voots, G.R., Gurnett, D.A., and Akasofu, S.-I., J. Geophys. Res., **82**, 2259, 1977.

Figure 2.28 Akasofu, S.-I., *Upper Atmosphere Research in Antarctica,* Antarctic Research Series, vol. 29, ed. by L.J. Lanzerotti and C.G. Park, AGU, Washington, D.C., 1978.

Figure 2.29 Bhardwaj, A. and Gladstone, G.R., Rev. Geophys. **38**, 295, 2000, Hubble Space Telescope Project.

Figure 2.30 Courtesy of Frank, L.

Figure 2.31 Courtesy of Geophysical Institute, University of Alaska.

CHAPTER 3

Figure 3.1 Chapman, S., Terr. Magn., **40**, 349, 1935.

Figure 3.2 Chapman, S., Terr. Magn., **40**, 349, 1935.

Figure 3.3a Axford, W.I. and Hines, C.O., Can. J. Phys., **39**, 1433, 1961.

Figure 3.3b Courtesy of Greenwald, R.A.

Figure 3.4 Kawasaki, K. and Akasofu, S.-I., Planet. Space Sci., **19**, 543, 1971.

Figure 3.5 Akasofu, S.-I. and Meng, C.-I., J. Geophys. Res., **74**, 293, 1969.

Figure 3.6a-d Akasofu, S.-I. and Kamide, Y., J. Geophys. Res., 103, 14,939, 1998.

Figure 3.7 Ahn, B.-H., Kamide, Y., and Akasofu, S.-I., J. Geophys. Res., **89**, 1613, 1984.

Figure 3.8 Akasofu, S.-I., Sun, W., and Ahn, B.-H. (in preparation 2002).

Figure 3.9 Boström, R., J. Geophys. Res., **69**, 4983, 1964.

Figure 3.10 Akasofu, S.-I., Geophys. J. Inst., **109**, 191, 1992.

Figure 3.11a,b Akasofu, S.-I. and Sun W., EOS, **81**, 361 2000.

Figure 3.11c DeForrest and McIlwain, C.E., J. Geophys. Res., **76**, 3587, 1971.

Figure 3.12 Waters, C.L., Anderson, B.J., and Liou, K., Geophys. Res. Lett., **28**, 2165, 2001.

Figure 3.13a	Akasofu, S.-I. and Meng, C.-I., J. Atm. Terr. Phys., **29**, 965, 1967.
Figure 3.13b	Akasofu, S.-I. and Meng, C.-I., J. Atm. Terr. Phys., **29**, 1015, 1967.
Figure 3.14	Akasofu, S.-I. and Meng, C.-I., J. Atm. Terr. Phys., **30**, 227, 1968.
Figure 3.16	Akasofu, S.-I., Nature, **284**, 248, 1980.
Figure 3.17	Akasofu, S.-I., Meng, C.-I., and Makita, K., Planet. Space Sci., **40**, 1513, 1992.
Figure 3.18	Frank, L.A., Sigwarth, J.B., and Paterson, W.R., *Substorms-4,* ed. by Kokubun S. and Kamide, Y., Terra Scientific Pub. Co., Tokyo, 1998.
Figure 3.19	Akasofu, S.-I., Ann. Geophys., **26**, 443, 1970.
Figure 3.20	Burch, J., Green, J.L., and Fuselier, S.A., EOS, **82**, 241, 2001.
Figure 3.21	Akasofu, S.-I., EOS, **73**, 209, 1992.
Figure 3.22	Akasofu, S.-I., Scientific American, p. 90, May, 1989.

CHAPTER 4

Figure 4.1	Akasofu, S.-I.
Figure 4.2	Akasofu, S.-I.,
Figure 4.3	Saito, T., Oki, T., Akasofu, S.-I., and Olmsted, C., J. Geophys. Res., **94**, 5453, 1989.
Figure 4.4	Saito, T., Oki, T., Olmsted, C., and Akasofu, S.-I., J. Geophys. Res., **94**, 14,993, 1989.
Figure 4.5	Saito, T., Kozuka, Y., Oki, T., and Akasofu, S.-I., J. Geophys. Res., **96**, 3807, 1991.
Figure 4.6	Akasofu, S.-I., Lee, L.-H., and Saito, T., Planet. Space Sci., **39**, 1259, 1991.
Figure 4.7	Akasofu, S.-I., Lee, L.-H. and Saito, T., Planet. Space Sci., **39**, 1259, 1991.

Figure 4.8 Akasofu, S.-I. and Saito, T., Planet. Space Sci., **38**, 1203, 1990.

Figure 4.9a,b,c Saito, T., Akasofu, S.-I., Kozuka, Y., Takahashi, T., and Numazawa, S., J. Geophys., Res., **98**, 5639, 1993.

Figure 4.10a,b Akasofu, S.-I. and Fry, C.D., J. Geophys. Res., **91**, 13,679, 1986.

CHAPTER 5

Most of the figures in Chapter 5 are in:
Akasofu, S.-I., Planet. Space Sci., **32**, 1469, 1984.

Figure 5.2a Courtesy of Big Bear Solar Observatory.

Figure 5.2b Courtesy of Kitt Peak Solar Observatory.

Figure 5.4 Akasofu, S.-I., Planet. Space Sci., **33**, 275, 1985.

Figure 5.8 Solar flare photo: Big Bear Solar Observatory.
Aurora photograph: Kamazawa Astronomical Society.

Figure 5.9 Courtesy of Choe C.S., and Lee, L.C., 1995.

Figure 5.10 Wang, H., Ewell, M.W. Jr., and Zirin, H., Astroph. J., **424**, 436, 1994.

Figure 5.11 Courtesy of The Transition Region and Coronal Explorer, TRACE, which is a mission of the Stanford-Lockheed Institute for Space Research (a joint program of the Lockheed Martin Advanced Technology Center's Solar and Astrophysics Laboratory and Stanford's Solar Observatories Group) and part of the NASA Small Explorer program.

CHAPTER 6

Most of the figures in this Chapter are in:
Akasofu, S.-I. and Fry, C.F., Planet. Space Sci., **34**, 77, 1986.

Akasofu, S.-I., *Space Weather,* Geophysical Monograph 125, AGU Workshop, 2001.

Hakamada, K. and Akasofu, S.-I., Space Sci. Rev., **31**, 3, 1982.

Akasofu, S.-I. and Lee, E.-H., Planet. Space Sci., **37**, 73, 1989 and 38, 575, 1990.

Akasofu, S.-I., Hakamada, K., and Fry, C., Planet. Space Sci., **31**, 1435, 1983.

Figure 6.7 Akasofu, S.-I., EOS, **77**, 225, 1996.

Figure 6.8c Hewish, A., Tappin, S.J., and Faggen, G.K., Nature, **314**, 137, 1985.

Figure 6.12 Courtesy of Maurits, S. and Watkins, B. See also Maurits, S.A., Ph. D. Thesis, University of Alaska, 1996.

Figure 6.13 Akasofu, S.-I. and Aspnes, J.D., Nature, **295**, 136, 1982.

Figure 6.14 Wescott, E., Sackinger, W., and Akasofu, S.-I., (unpublished report).

Figure 6.15 Tang, F., Akasofu, S.-I., Smith, E., and Tsurutani, B., J. Geophys. Res., **90**, 2703, 1985.

Figure 6.16 Fry, C.D. and Akasofu, S.-I., Planet. Space Sci., **33**, 925, 1985.

Figure 6.17 Courtesy of NASA Sky Lab Project, the Solar Physics Group, America Science and Engineering Inc.

Figure 6.18 Akasofu, S.-I.

Figure 6.19a Courtesy of Big Bear Solar Observatory.

Figure 6.19b Courtesy of SOHO [Instrument] consortium. SOHO is a project of international cooperation between ESA and NASA.

Figure 6.19c The solar X-ray images are from the Yohkoh mission of ISAS, Japan. The X-ray telescope was prepared by the Lockheed-Martin Solar and Astrophysics Laboratory, the National Astronomical Observatory of Japan, and the University of Tokyo with the support of NASA and ISAS.

Figure 6.20 Courtesy of Dave Rust.

Figure 6.21 and
Figure 6.22 Sun, W., M. Dryer, C.D. Fry, C.S. Deehr, F. Smith, S.-I. Akasofu, M.D. Kartalev, and K.G. Grigorov, Geophys. Res. Lett, (in press) 2002.

Figure 6.23 Data Center for Geomagnetism, Kyoto University.

Figure 6.25a Akasofu, S.-I and Chapman S., Planet. Space Sci. **12**, 607, 1964.

Figure 6.25b Akasofu, S.-I., and Chapman, S., *Solar-Terrestrial Physics,* Oxford University Press, 1972.

CHAPTER 7

Figure 7.1 Akasofu, S.-I. and Covey, D.N., Planet Space Sci., **29**, 313, 1981.

Figure 7.2a,b Akasofu, S.-I. and Covey, D.N., Planet. Space Sci., **29**, 313, 1981.

Figure 7.3 Akasofu, S.-I., Fillius, W., Sun, W., Fry, C. and Dryer, M., J. Geophys. Res., **90**, 8193, 1985.

Figure 7.4 Yoshida, S., Ogita, N., and Akasofu, S.-I., J. Geophys. Res. **76**, 7801, 1971.

Figure 7.5 Yoshita, S., Ogita, N., Akasofu, S.-I., and Gleeson, L.J., J. Geophys. Res., **78**, 6409, 1973.

PREFACE

My purpose in writing this book is to describe my own experiences, from my graduate student days in the 1950s to the present (2001), when I came upon phenomena or facts that did not support the prevailing ideas and theories, or even contradicted them. In some instances, the encounters began with nothing more than the naïve questions I posed as a graduate student to my professors regarding a well-established fact; others were the result of questions my graduate students asked me. Essentially, this is an account of my personal encounters with some of the ideas and theories that once prevailed but were later eliminated in the history of auroral science.

I believe that young researchers becoming successful as scientists depends on how they deal with new phenomena or facts that do not fit established theories. One cannot be a researcher unless he/she can encounter such a problem. This is because such an encounter is the very first step for new progress. When encountering such problems, some may put the discordant facts on the shelf or sweep them under the rug, so to speak, at least for a while. Others may try hard to find a way to make new facts fit into prevailing ideas by modifying or improving them. Yet others may try to establish a new idea, scheme, or theory by adapting their findings and those of others, but abandoning the prevailing interpretation of the phenomena or facts. It has been my experience that it is the people in this last group who produce epoch-making progress in science.

The choice of what to do when facing this situation is not easy and depends on many factors. First of all, researchers have to know where they stand at that point in the history of their scientific discipline. It is therefore crucial to have a deep historical knowledge of the background of a prevailing idea or the established interpretation of a phenomenon. To choose a course of action without knowing the background would be like starting to run in the dark without a sense of direction or of the surroundings. Unfortunately, I see too many young scientists doing just that, particularly those who believe that technological advance is everything. Often, a mentor provides the history, not necessarily in a classroom setting, but through daily interactions. I was fortunate to have a very good mentor, Sydney Chapman, who guided me during my early days.

It is also my hope for this book that young researchers will learn that even a simple, one-line statement in a standard textbook, such as The aurora lies along an oval-shaped belt, endured a decade of struggle before acceptance by the scientific community. My point here is that it is important to learn how to proceed during the period of controversy and struggle, which requires skills not taught in a textbook. However, it is not the intent of this book to provide a general methodology, even if one existed, on how to overcome such problems. I show several examples, right or wrong. The creative approach taken by individual researchers is crucial at this point. In science, we may eventually reach the same or a similar conclusion, but the creative approach taken depends greatly on the individual, as the history of science proves. Science is a human endeavor and is not a dry subject at all.

It is obvious, first of all, that new ideas or theories in science should explain more observational facts than the old ones did. However, that an idea is great (or better) does not guarantee its immediate acceptance by the scientific community. Scientific accuracy is a necessary condition for acceptance, but is not in itself a sufficient condition for it. The readers of this book will see examples, not a methodology, of how such situations were dealt with in the history of auroral science by researchers who made significant advances in understanding auroral phenomena. The most serious problem in a scientific discipline occurs when a given idea or theory dominates utterly. The longer a particular prevailing idea dominates, the more damage it does, retarding progress as researchers, young and old, begin to feel that there is nothing major left to be done.

Looking back at the history of auroral science, one can find that our pioneers had dreams. Our generation also had dreams. Some of the recent advances have made their and our dreams a reality. In order to make this book a little more than just my own ramblings, I have added several highlights concerning those advances in some of the chapters.

However, in spite of the considerable progress in the disciplines of solar-terrestrial physics, a number of long-standing fundamental problems have remained unsolved for many decades. It is my belief that some of these problems remain unsolved because no doubt has been cast on the guiding concept behind the prevailing ideas, not necessarily because we are presently unable to solve them technically. In order to stimulate new or different ways of thinking, I have decided to provide here some unconventional ideas, although they will certainly be criticized or ignored by those who believe that they are on the right track and that their difficulties are only technical in nature. However, it must be noted that all the materials used here were at least accepted and published in standard scientific journals; many of my unsuccessful geophysical research projects will be described elsewhere.

Space physics must evolve. The future of space physics depends on the creativity of the young generation with a wide range of interests in other fields of science. With a solid background in space physics and at least one other field, the young generation should be able to create a new field of science. I have suggested the exploration for life on planets of distant stars by searching for oxygen emissions in their aurora. That is just an example, and there may be many new unexpected fields of science.

Obviously, this book is not a textbook, or an autobiography, or a treatise of facts and theories for a particular prevailing idea or two. It is a sort of reflection on my research endeavor during the last forty years or so. Since I have an instinctive tendency to avoid prevailing ideas and theories, I am perhaps not a normal scientist, but I hope nevertheless that this book will be useful, particularly for graduate students and young scientists, especially in helping them think beyond the box of accepted wisdom.

I thank my senior and junior colleagues in many countries, and my former graduate students, who participated in my research activities and helped guide me. Without their close interaction over my research career, I would not be writing this book. Those who are not mentioned in the main text are acknowledged in the figure captions. I have also worked with many other close colleagues who are not mentioned in this book, but could not mention them in order to focus on the subject areas specifically dealt with in this book.

Note: At the end of the book, further readings are listed for those who are interested in the history of auroral science, but they are not a reference list. The names of authors with the year in parentheses in the following chapters may look like citations but instead simply indicate the year their papers were published. I have used the full first name of those authors with whom I had at least some acquaintance. For all the rest, I have given only their initials.

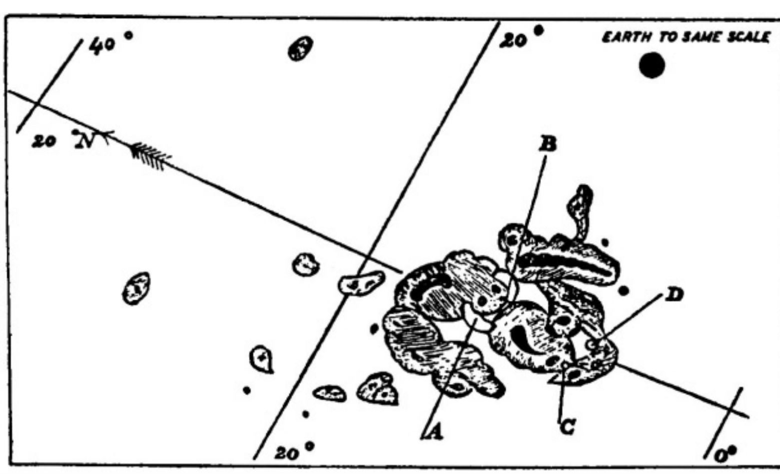
PROLOGUE

The story in this book had a fascinating beginning that can best be described by R. C. Carrington (1860) with his own words:

> *While engaged in the forenoon of Thursday, September 1, in taking my customary observations of the forms and positions of the solar spots, an appearance was witnessed which I believe to be exceedingly rare – two patches of intensely bright and white light broke out...*

Simultaneous with this first sighting of what is now called a white-light solar flare (a most intense type of solar activity), the terrestrial magnetic field record made at the Kew Magnetic Observatory in Greenwich, England, showed a distinct magnetic variation.[1] About 16 hours after this remarkable event, a great geomagnetic storm began and a brilliant auroral display appeared over northern Europe and many other places. Carrington suspected that the geomagnetic storm was related to what he had observed on the Sun, but hesitated to assert the connection. The footnote in Carrington's report to a meeting of the Royal Astronomical Society reads:

> *While the contemporary occurrence may deserve noting he would not have it supposed that he even leans towards hastily connecting them. "One swallow does not make a summer."*



R. C. Carrington's sketch of a sunspot group. He was the first to witness a solar flare (A, B, C, D): R.C. Carrington (1860).

---

[1]This magnetic change is a result of *augmentation* of the ionospheric current by an enhanced conductivity of the Earth's ionosphere caused by the flare's radiations.

Lord Kelvin (1824-1907)

It is in this way that solar-terrestrial physics was born. Lord Kelvin (1892) took up Carrington's extremely modest suggestion of the solar-terrestrial connection during the Anniversary Meeting of the Royal Society of London, England, in 1892.

Kelvin, then the president of the society, attempted to explain the observed geomagnetic variations in terms of the solar magnetic changes observed at a distance of 200 solar radii and found that the expected changes of the dipole moment of the Sun were too large to be reasonable. Thus, he concluded:

> *...Guided by Maxwell's "electro-magnetic theory of light", and the adulatory theory of propagation of magnetic force which it includes, we might hope to perfectly overcome a fifty years' outstanding difficulty in the way of believing the Sun to be the direct cause of magnetic storms in the Earth, though hitherto every effort in this direction has been disappointing. It seems as if we may also be forced to conclude that the supposed connection between magnetic storms and Sunspots is unreal, and that the seeming agreement between the periods has been mere coincidence*[2] .

His difficulty is understandable. Without the concept of a medium (which now is known as solar plasma flow) that carries the effects of solar disturbances out into interplanetary space, it is not possible for the Sun to cause the magnetic changes recorded on the Earth.

---

[2]It may be of interest to note that Kelvin estimated the age of the Earth to be less than 40 million years, based on a heat conduction theory; at that time, radioactivity was not known as the heat source in the Earth.

E.W. Maunder (1851-1928)

E.W. Maunder (1905) made a new approach to this problem by noting that geomagnetic disturbances generally reoccur every 27 days, the so-called *27-day recurrence* tendency. After an extensive study of magnetic and solar records, he concluded:

> *First: The origin of our magnetic disturbances lies in the Sun: not any body or bodies affecting both. This is clear from the manner in which those disturbances mark out the solar rotation period...*
>
> *Second: The areas of the Sun giving rise to our magnetic disturbances are definite and restricted areas...not due to a general action or influence diffuse over the whole solar surface.*
>
> *Third: The areas of the Sun, wherein the magnetically active areas are situated, rotate with the speed of the chief spot-bearing zones, viz., latitudes 0° to 30°.*
>
> *Ninth: ...though Sunspots and magnetic disturbances are intimately connected, large Sunspots will often be observed when no disturbances are experienced, whilst sometimes disturbances will be experienced when no spots with which they can be associated are visible...*

The first statement was the most definitive in history in suggesting the solar-terrestrial connection. The other remarks are also quite accurate in spite of the very limited amount of data available to Maunder at that time. The spot-free region he referred to is what we now call a *coronal hole.* In his concluding remark, Maunder noted:

> *That, therefore, which Lord Kelvin spoke of twelve years ago as "the fifty years outstanding difficulty" is now rendered clear...*

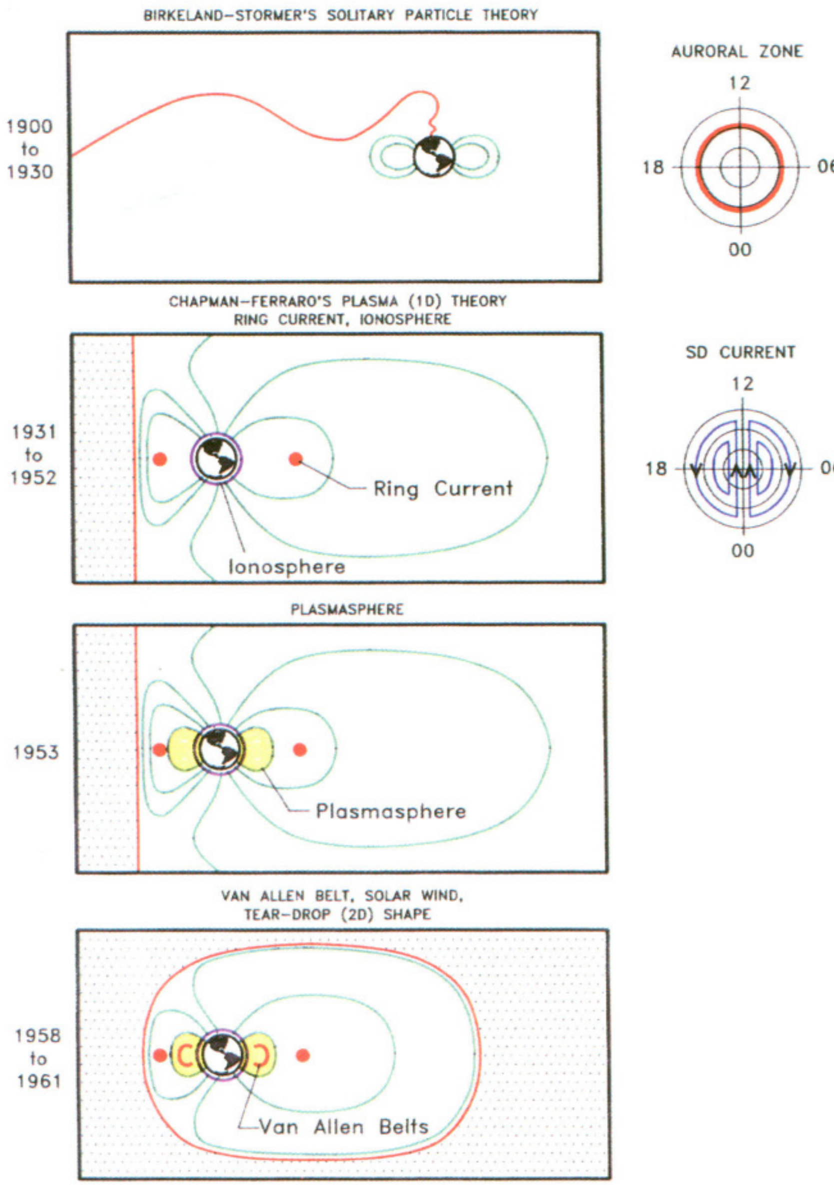
A. Schuster (1905) immediately criticized Maunder's conclusion by an argument similar to that presented by Kelvin:

> *...I cannot, therefore, agree with his somewhat boastful claim that he has rendered clear what Lord Kelvin has called a "fifty years' outstanding difficulty." He has, no doubt, added a new fact and made an important contribution to the subject. He has given a renewed interest to it and brought out the urgent importance of further investigation, but the mystery is left more mysterious than ever. The facts have become harder to understand and more difficult to explain.*

In the history of solar-terrestrial physics, as in any other field of science, such controversies among experimenters, observers, and theorists have been a common occurrence. However, through such controversies, their efforts have been interwoven, resulting eventually in a better understanding of natural phenomena.

After such exciting beginnings, the concept of the Earth's electromagnetic environments has evolved dramatically (Figure I). K. Birkeland viewed the interaction between the solar gas and the Earth's magnetic field in terms of motions of solitary charged particles in a dipole field. He set up an elaborate discharge chamber to study the trajectories of electrons around what he called a *terrella.* Stimulated by Birkeland, C. Störmer began his lifelong study of trajectories of charged particles in a dipole field. His life work was summarized in his book, *The Polar Aurora* (1955). In their studies, both Birkeland and Störmer assumed that the Earth's magnetic field was unaffected by the advancing solar gas.

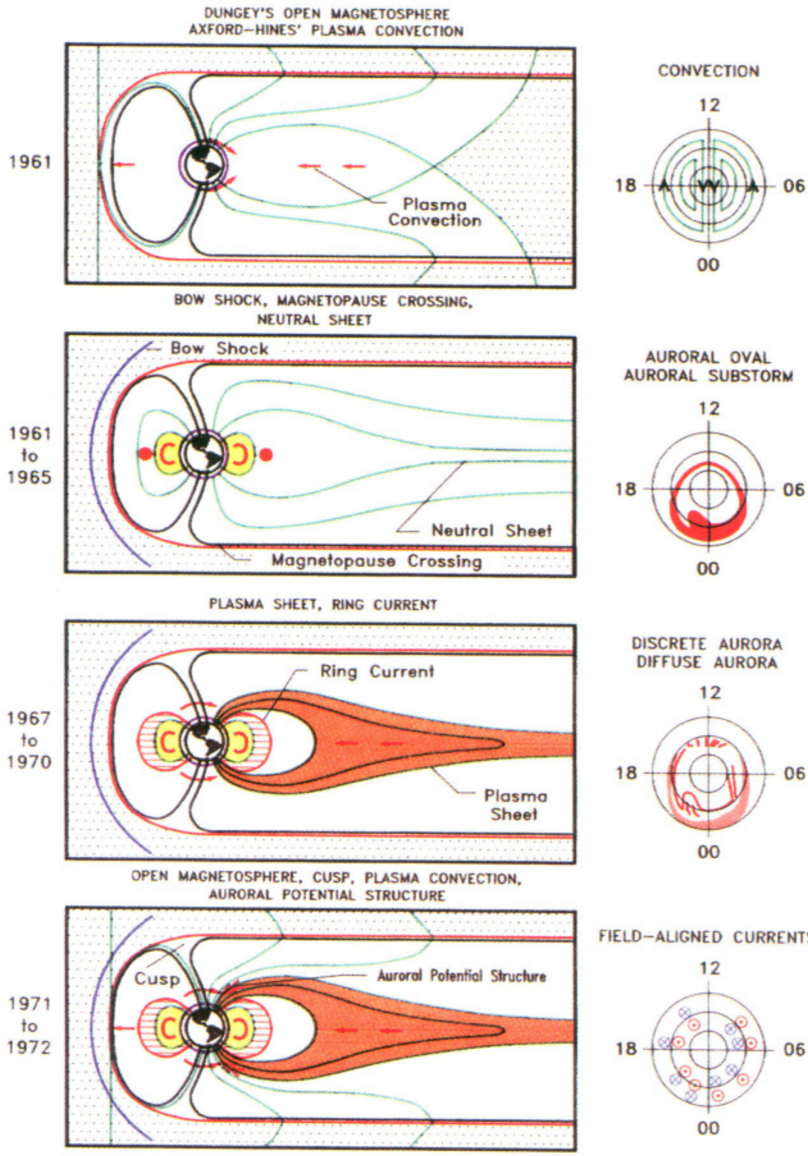
In order for this particular field of science to make substantial progress, however, we had to wait for Sydney Chapman and Vincenzo Ferraro (1931) to introduce the concept of confinement of the Earth's magnetic field in a cavity carved in the solar gas flow. Chapman and Ferraro considered the solar gas to be plasma (in present terminology) and attempted to understand the behavior of the plasma flow as it approached a dipole field. They inferred that the solar plasma flow forms a cometlike structure around the Earth, extending in the anti-solar direction and confining the Earth and its magnetic field in it. Chapman and Julius Bartels summarized the development of the field in their classic treatise *Geomagnetism* in 1940.



BIRKELAND–STORMER'S SOLITARY PARTICLE THEORY
1900
to
1930
AURORAL ZONE
12
18
06
00
CHAPMAN–FERRARO'S PLASMA (1D) THEORY
RING CURRENT, IONOSPHERE
1931
to
1952
Ring Current
Ionosphere
SD CURRENT
12
18
06
00
PLASMASPHERE
1953
Plasmasphere
VAN ALLEN BELT, SOLAR WIND,
TEAR–DROP (2D) SHAPE
1958
to
1961
Van Allen Belts

Figure I

Figure I continued



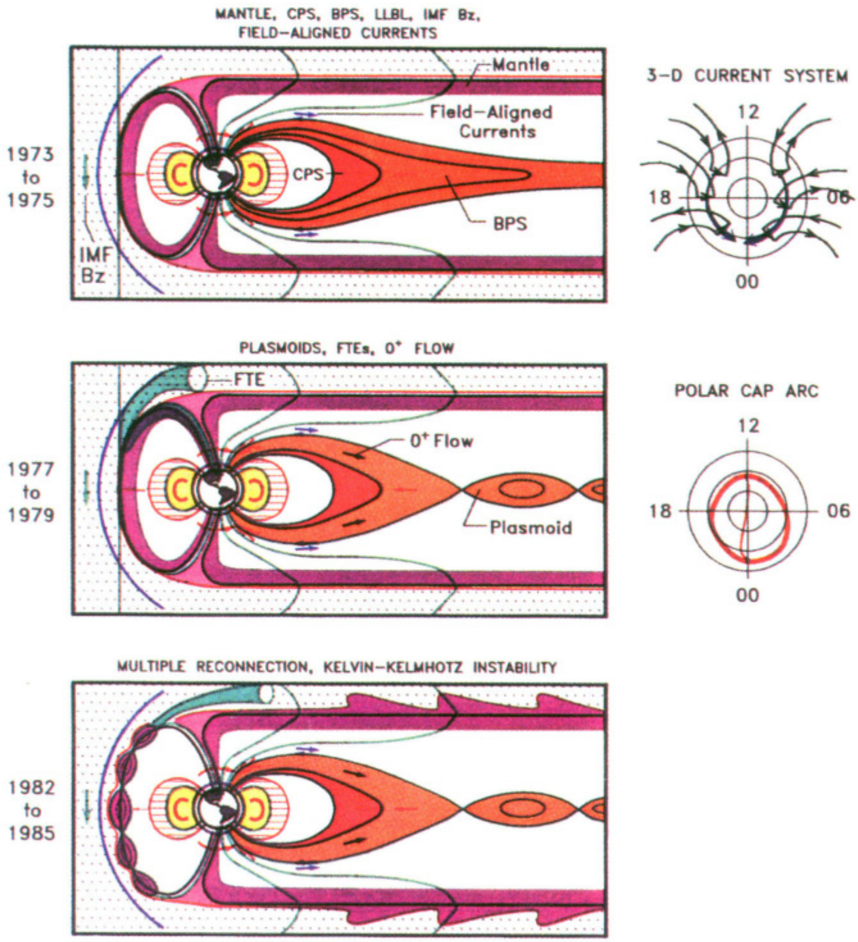


Figure I: Schematic presentation of the development of the concept of the magnetosphere: S.-I. Akasofu (1991).

The discipline of geomagnetism evolved into magnetospheric physics after the International Geophysical Year (IGY), the historic geoscience enterprise in 1957-1958, namely during the beginning of the space age. Tomy Gold (1959) coined the term *magnetosphere* by defining it as "the region above the ionosphere in which the magnetic field of the Earth has a dominant control over the motions of gas and fast charged particles."

The Earth's electromagnetic environment is continuously monitored by recording changes of the Earth's magnetic field. The record shows from time to time characteristic changes $\Delta \boldsymbol{B}$ of the Earth's magnetic field. At a low-latitude observatory, the magnetic variations begin with a steplike increase for a few hours, which is then followed by a decrease of a larger magnitude for a day or so. The upper diagram of Figure II shows magnetic records of the north-south component from several low-latitude stations widely separated in longitude; northward changes are recorded as positive changes, while southward changes are recorded as negative changes. The first increase and the subsequent larger decrease are observed at all stations, indicating that those changes occurred on a global scale. This phenomenon is called the *geomagnetic storm.* The development of the study of geomagnetic storms is one of the important subjects of this book. It may be noted that the term *magnetic storm* was coined by A. von Humboldt in his treatise *Cosmos* (1871).

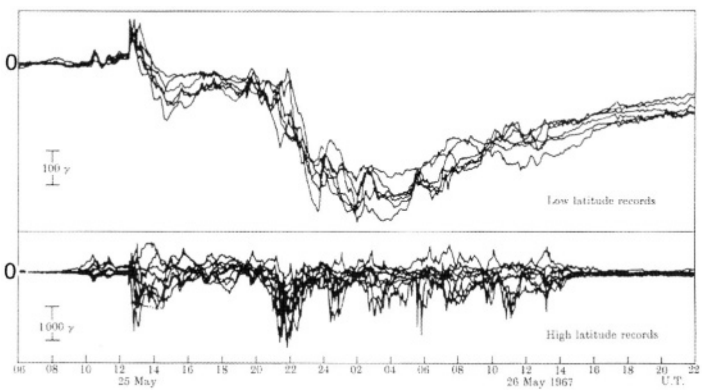


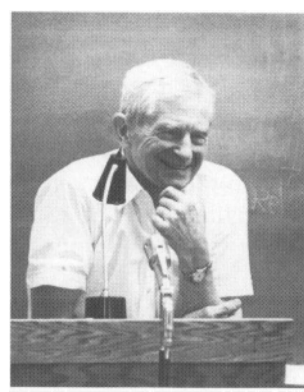
Figure II:
Upper – Superimposed magnetic records (the north-south component) on May 25-26, 1967, from six low-latitude observatories separated widely in longitude.
Lower – Superimposed magnetic records on the same dates from nine high-latitude observatories separated widely in longitude. Note the difference of the scale between the upper and lower diagrams: S.-I. Akasofu and S. Chapman (1972).

The geomagnetic storm field $\Delta\boldsymbol{B}$ is produced by various electric current systems that develop around the Earth when solar disturbances reach the Earth. The field $\Delta\boldsymbol{B}$ is thus superposed on the Earth's main field $\boldsymbol{B}_0$ which does not change in days or months.

During a geomagnetism storm, at high-latitude observatories, fluctuations of a much greater magnitude than those seen in low latitudes, consisting of a number of simultaneous impulsive changes, can be observed. In the lower diagram of Figure II, magnetic records from a number of high-latitude stations are shown; note the difference of the scale for the low- and high-latitude records. Those impulsive changes are magnetic manifestations of what we now know as magnetospheric substorms. During a geomagnetic storm, a number of such intense impulsive disturbances occur.

Birkeland classified fluctuations of the Earth's magnetic field in terms of equatorial positive/negative and polar positive/negative changes. As far as I am aware, Chapman was the first who established the present concept of the *geomagnetic storm.* It consists of the *storm sudden commencement* (SSC), a step-function-like increase in the horizontal (north-south) component and the *main phase,* a larger decrease that follows the SSC. There is often a relatively steady period of a few hours after the SSC, which is followed by the main phase; this period is called the *initial phase.* The SSC is caused by the impact of the shock wave on the magnetosphere; the shock wave is generated by some solar activities. The main phase is caused by the formation of a belt of energetic particles that surround the Earth. This belt is called the *ring current belt.* After reaching the maximum decrease during the main phase, the storm tends to recover slowly; this phase is called the *recovery phase.* In the book *Geomagnetism,* by Chapman and Bartels (1940), an early account of the development of a study of geomagnetic storms is outlined. Chapman told me that there was great difficulty in publishing it, as world tension was mounting before World War II.

K. Birkeland (1867-1917)



Sydney Chapman (1888-1970)

After World War II, in the 1950s and 1960s, there were several important developments in a study of the electromagnetic environment between the Sun and the Earth and beyond. First of all, until that time, interplanetary space was thought to be practically a vacuum, except for the streams suggested by Maunder and clouds ejected by solar flare activity. Thus, the Chapman-Ferraro cavity was thought to form only *occasionally,* as the solar plasma engulfed the Earth. Meanwhile, the suggestion of a continuous flow of the solar wind by Ludwig Biermann (1951, 1953) and Gene Parker (1958), and the subsequent detection of the solar wind by the Mariner 2 spacecraft in

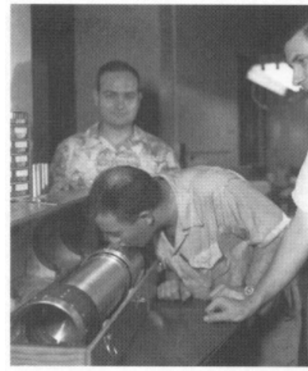
1962, brought about a significant change in the concept of the magnetosphere. The magnetosphere is now considered a permanent feature of the Earth, so long as the solar wind blows, rather than forming only occasionally when the Earth is engulfed by intermittent solar plasma flows.

Second, an extensive tail of the magnetosphere was first revealed by the IMP-1 satellite, reported by Norman Ness and his colleagues (1965), as had been suggested by Jack Piddington (1960). It was found later by space probes on their way to outer planets that the magnetotail extends to a distance of about 1000 Earth radii and perhaps farther.

Third, it was found that the Earth is surrounded by an extensive atmosphere of ionized gases. Based on the study of atmospherics (radio emissions generated by thunderstorm lightning), L.R.O. Storey (1953) found that atmospherics can propagate approximately along the geomagnetic field lines from one hemisphere to the other. The propagation requires an extensive ionized atmosphere to a distance of several Earth radii. This ionized atmosphere has been named the *plasmasphere.* The ionosphere feeds the ionized gases to the plasmasphere.

The Space Age and space research by rockets and satellites were initiated by James Van Allen. In his effort to explore the origin of auroral electrons and cosmic rays, his first attempt was to study auroral electrons near Greenland by *rockoons,* a combination of a rocket and a balloon. It is worth noting that the space age was initiated by the curiosity of scientists like him, who were pursuing the causes of auroral and geomagnetic phenomena. It was his pursuit of auroral electrons by satellites which led him to the discovery of the Van Allen radiation belts. Subsequently, the ground-based discipline of geomagnetism, together with satellite-based studies, developed into magnetospheric physics. In theoretical space research, Hannes Alfvén stimulated my generation most by introducing many creative concepts, including the concept of the guiding center, magnetohydrodynamics (MHD), Alfvén waves, dusty plasmas, and many others.

Hannes Alfvén (1908-1995)



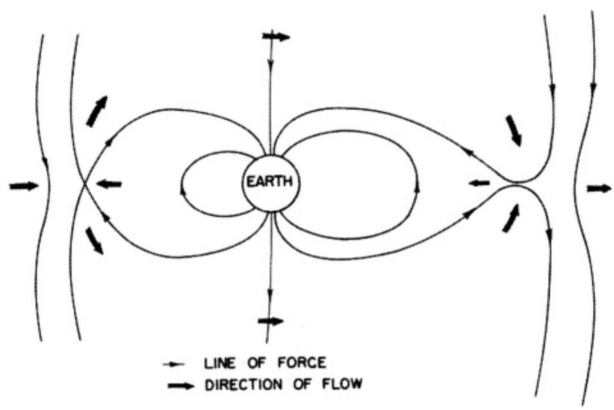
James Van Allen with Carl McIlwain (left) and George Ludwig (right), giving a farewell kiss to their detector to be carried by one of the first U.S. satellites.

In 1968, Sam Bame and his colleagues at Los Alamos National Laboratories discovered the most extensive region of plasma, called the *plasma sheet,* in the tail region of the magnetosphere. Thus, the magnetosphere has been found to be not an empty cavity, but to consist of several plasma domains. In the 1970s, solar wind-like plasmas were found well inside the boundary of the magnetosphere, and the region occupied by such plasmas is called the *plasma mantle.* The plasma in the plasma mantle flows in the antisolar direction with a speed of about 100 km/sec, appreciably less than that of the solar wind. Certainly, the plasma in the plasma mantle is of solar wind origin. This finding indicates that the magnetospheric boundary does not exclude completely the solar wind from the magnetosphere, as Chapman and Ferraro originally envisioned in their theory. This was also a major change of the concept of the magnetosphere. Another unexpected finding by D.C. Hamilton and his colleagues in 1988 was that oxygen ions (O^+) of ionospheric origin, instead of solar wind protons, become the dominant ions in the ring current belt during geomagnetic storms.

Jim Dungey (1961) made the most drastic addition to, or more appropriately the most fundamental revision of, Chapman-Ferraro's original theory. He suggested that the magnetic field lines carried by the solar wind are connected with some of the geomagnetic field lines across the boundary of the magnetosphere. Such a magnetosphere is said to be *open,* while the Chapman-Ferraro model is called a *closed* magnetosphere. The difference between the two theories is that Dungey considered magnetized solar wind plasma, while Chapman and Ferraro considered diamagnetic plasma.

Dungey envisaged that the connection process, called *reconnection,* takes place on the dayside magnetopause and that the connected field lines are then transported in the antisolar direction by the solar wind, resulting in the magnetotail. Subsequently, the field lines are reconnected there and then transported back to the dayside magnetosphere. Dungey's view was that this transport process may occur intermittently and that magnetospheric disturbances, such as magnetospheric substorms, are a manifestation of such a transient process.

In this book, we consider that this interaction between the magnetized solar wind and the magnetosphere constitutes a dynamo that converts a small fraction of the kinetic energy of the solar wind into electrical energy. Chapter 1 describes efforts toward this understanding based on my own experience.



Dungey’s open magnetosphere: J.W. Dungey (1961).

The aurora can then be understood as the only visible manifestation of electrical discharge processes that are powered by the dynamo. Its output power is usually one million megawatts or more. The discharge takes place in an oval-shaped belt, called the *auroral oval,* in the polar upper atmosphere. On the basis of this finding, it could be expected that a magnetized planet with an atmosphere, such as Jupiter, Saturn, Uranus, and Neptune, would have a similar auroral oval, while a nonmagnetized planet, such as Venus and Mars, would have no auroral oval. Indeed, the Hubble Space Telescope Project succeeded in imaging the auroral ovals of Jupiter and Saturn (see Figure 2.29), while the Venus and Mars orbiters could not image any indication of the auroral oval.

As mentioned earlier, geomagnetic disturbance fields $\Delta \boldsymbol{B}$ are the magnetic fields produced by the discharge current generated by an enhanced solar wind-magnetosphere dynamo power. *Thus, auroral activity and geomagnetic disturbances are only different manifestations of an enhanced dynamo power.*

Obviously, the two subjects cannot be discussed separately in understanding magnetospheric disturbances. One purpose of studying auroral phenomena and geomagnetic disturbances is, among other things, to infer the configuration of the discharge current system in the magnetosphere and the dynamo process that feeds the current. Chapters 2 and 3 describe our efforts in this endeavor during the early days of the space age.

A typical geomagnetic disturbance field $\Delta\boldsymbol{B}$ undergoes a specific sequence of changes, as shown in Figure II. We now understand that a geomagnetic storm is the magnetic manifestation of what we call a *magnetospheric storm* that results from a large increase of the dynamo's power. Similarly, an auroral storm is its visible manifestation.

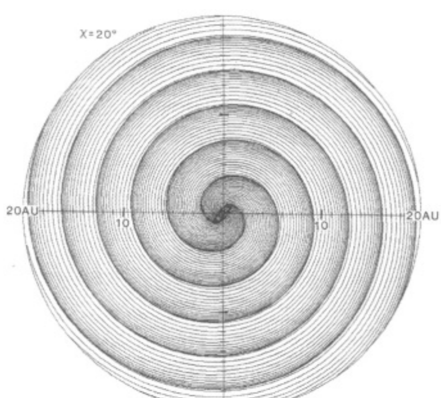
It also has been found that the magnetosphere has a specific response to an increased power for a few hours. The results of these responses are called *magnetospheric substorms,* their manifestations being the *polar magnetic* and *auroral substorms.* As Chapter 3 explains, a magnetospheric storm results from a frequent occurrence of intense magnetospheric substorms. That is to say, *the substorms are the elements of a storm.* This is because each substorm feeds oxygen ions (O^+) from the ionosphere into the ring current belt.

Planetary magnetism is an important subject for all geophysicists, solar physicists, and astrophysicists. It has been a great surprise that the dipole fields of both Uranus and Neptune appear to be inclined considerably with respect to their rotation axis and are greatly off-centered. So long as the generation of planetary magnetism relies on the planetary rotation, it is difficult to explain why the magnetic axis is inclined greatly from the rotation axis. Chapter 4 provides a nontraditional interpretation of planetary magnetic fields. In this attempt, it is assumed that the photosphere of the Sun corresponds to the core surface of the magnetized planets and a spherical surface of 2.5 solar radii, called the *source surface,* corresponds to the planetary surface, where the field is dipolar. Thus, a study of the relationship between the magnetic fields of the photosphere and the source surface might provide a new way of interpreting the observed planetary magnetic field.

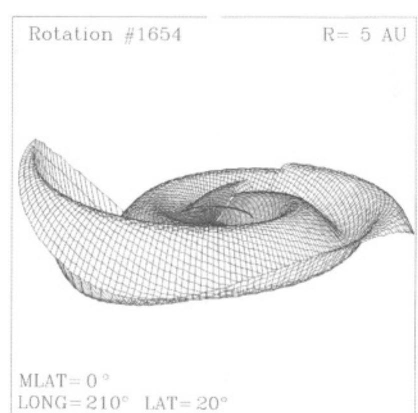
The solar wind stretches the dipolar field lines on the source surface all the way to the outer boundary of the heliosphere, where the solar wind interacts with interstellar gas. As the Sun rotates with a period of about 25 days, the stretched field develops a spiral structure. The heliospheric current sheet is formed as the extension of the magnetic equator of the Sun. The current sheet divides the stretched dipolar field lines (the interplanetary magnetic field lines) into two regimes, away and toward the Sun, in terms of the orientation.

As the solar wind and its magnetic field are continuously changing, the power of the solar wind-magnetosphere dynamo varies as well. In particular, after a few days of intense solar activities (called solar flares, coronal mass ejections (CME) and others), an intensified solar wind, together with its shock wave, reaches the magnetosphere. The shock wave compresses the magnetosphere. As a result, the Alfvén waves are generated at the front of the magnetosphere and propagate into the magnetosphere, causing the storm sudden commencement (SSC). Subsequently, coronal mass and its magnetic structures (either in the form of magnetic clouds or magnetic loops) are thought to arrive at the front of the magnetosphere and increase the dynamo power, causing a frequent occurrence of magnetospheric substorms and thus subsequently generating the ring current belt and the magnetospheric storm.

The ultimate cause of magnetospheric storms is thus a variety of transient solar activities. In spite of more than a half-century of intense research, however, the causes of sunspots, solar flares, and coronal mass ejections still remain as long-standing unsolved problems. Most solar physicists think that solar activities are directly related to hypothetical thin magnetic flux tubes beneath the photosphere, their uplift by buoyancy and magnetic reconnection among them after their uplift. Later, in Chapter 5, it will be pointed out that magnetic flux tubes are nothing but a hypothesis, perhaps an unworkable one. It will also be pointed out that a dynamo process in the solar atmosphere must generate the source of energy for solar activities, since solar activities are basically electromagnetic phenomena. An attempt will also be made to show that a dynamo process in the photosphere is responsible for the power supply.

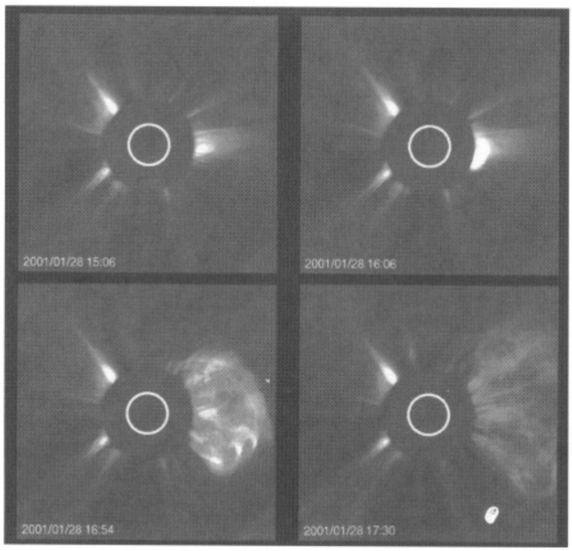


The spiral structure of the interplanetary magnetic field lines; the Sun is located at the center: K. Hakamada and S.-I. Akasofu (1982).

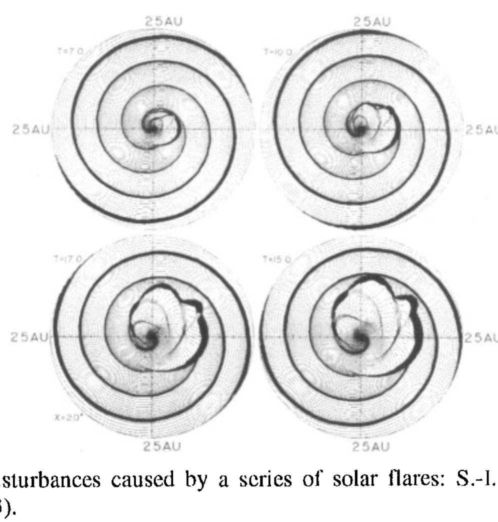


Heliospheric current sheet that extends from the Sun for Carrington Rotation 1654. The Earth's orbit is also shown: S.-I. Akasofu and C.D. Fry (1986).

In this short review on the progress of solar-terrestrial physics, one can see that investigators of the four disciplines–solar physics, interplanetary physics, magnetospheric physics, and aeronomy–have made considerable progress in the twentieth century after Carrington's finding. However, for these disciplines to progress further, in particular in terms of space weather research, it is important for solar physicists, interplanetary physicists, magnetospheric physicists, and upper atmospheric physicists to work together. There are many missing links among the four disciplines that will only be noticed if one attempts to succeed in space weather research. Chapter 6 is devoted to the integration of the four fields.



Intense solar activity observed by SOHO on January 28, 2001: NASA/ESA Project.



Interplanetary disturbances caused by a series of solar flares: S.-I. Akasofu and K. Hakamada (1983).

The disturbed solar wind caused by various solar activities advances into this interplanetary structure, well beyond the distance of the Earth. Shock waves also form a barrier for cosmic ray particles that enter from the outer boundary of the heliosphere, causing a reduction of the cosmic ray intensity in the heliosphere. This phenomenon was discovered by Scott Forbush and is called the *Forbush effect.* Chapter 7 describes the magnetic field structure of the heliosphere and how it is disturbed by solar activities.

It is hoped that the readers of this book will find a number of longstanding unsolved problems in the four disciplines. I believe that many of the difficulties the present generation is facing are not due to the lack of basic knowledge and technical problems (for example, the capability of a supercomputer), but to our inability to recognize fundamental flaws in the presently prevailing concepts, namely paradigms. The Epilogue is devoted to discussing this issue. The new generation of scientists are encouraged to challenge the present paradigms and advance our understanding of electromagnetic phenomena around the Earth, in interplanetary space, and the heliosphere.

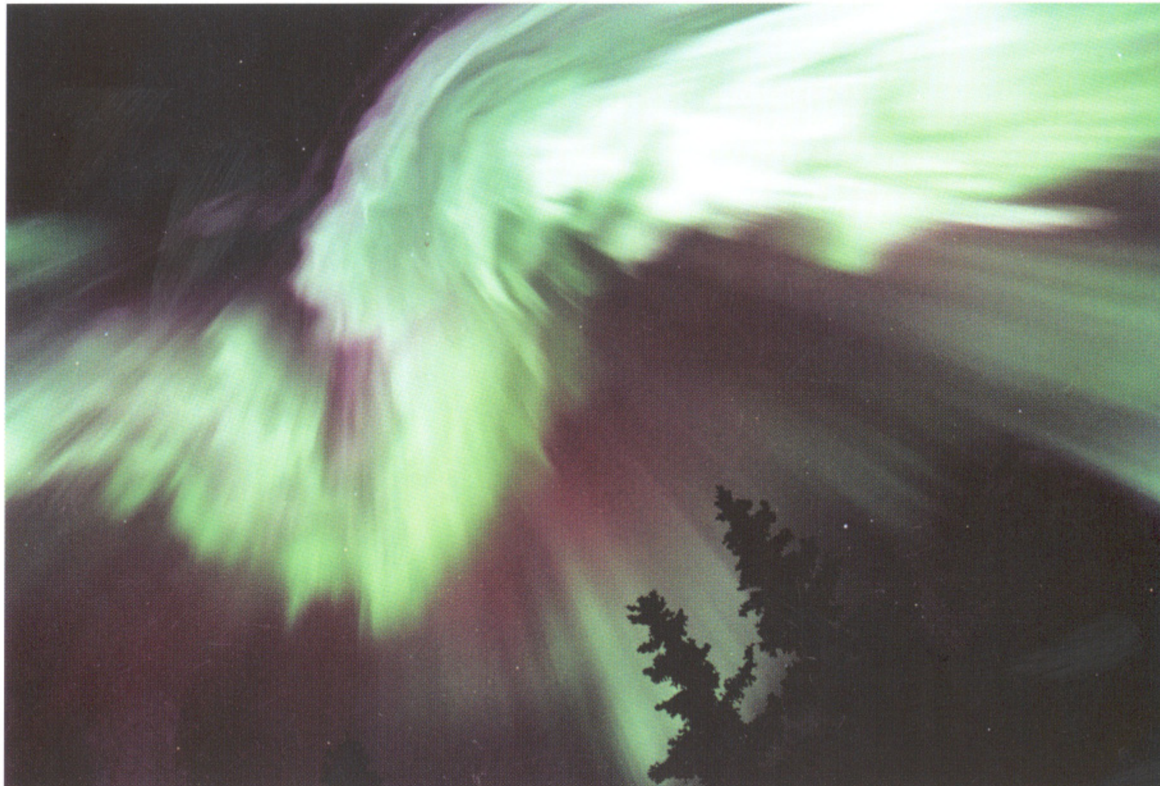


Plate 1: An active auroral curtain near the zenith. This form is sometimes referred to as the corona-type display. Photographed by Jan Curtis.

Plate 2: The curtain-like form of the aurora. The upper part of the curtain is rich in the dark red emissions from atomic oxygen. Photographed by Jan Curtis.

Plate 3: An active auroral display called the westward traveling surge. The upper part of this particular aurora shows the dark red emission (660 nm) from atomic oxygen. Photographed by Jan Curtis.

Plate 4: A typical red aurora. Photographed by Jan Curtis.

Chapter 1

SEARCH FOR THE UNKNOWN QUANTITY IN THE SOLAR WIND

1.1 Solar Corpuscular Streams

One of the most remarkable things about E.W. Maunder, as mentioned in the prologue, is that most of his conclusions have remained valid. In his sixth conclusion, Maunder (1905) stated:

> *...such a relation can only be explained by supposing that the Earth has encountered, time after time, a definitive stream. A stream which, continually supplied from one and the same area of the Sun's surface, appears to us, at our distance, to be rotating with the same speed as the area from which it rises.*

During the first ten years of the twentieth century, Maunder's idea was gradually accepted. As indicated by the statements by Lord Kelvin and A. Schuster quoted in the prologue, the acceptance did not result from an elaboration and extension of the theoretical framework, but from intuitive associations between solar flares and geomagnetic disturbances, between geomagnetic disturbances and the aurora, and also between the aurora and the glow in a cathode ray tube. Recognizing the close association of geomagnetic storms and auroras is one of the important contributions toward the acceptance of the nature of the solar-terrestrial relationship Kelvin doubted. In his book *Cosmos*, A. Humboldt (1871) described his own finding:

> *The mysterious course of the magnetic needle is equally affected by time and space, by the Sun's course, and by changes of place on the Earth's surface. ...It is affected instantly, but only transiently, by the distant northern light as it shoots from the pole, flashing in beams of coloured light across the heavens.*

This statement was based on his own incredibly strenuous observations. In a biography of Humboldt, L. Kellner (1963) noted:

> *Humboldt had rented a small cottage in the garden of a rich brandy distiller on the outskirts of Berlin where he set up his instruments....Here, Humboldt carried out more than six thousand observations, from May 1806 until June 1807. Glued to his post, he spent, at one time, seven days and nights, in succession, at his*

instruments, taking half-hour readings....In December, he was lucky enough to observe a display of the aurora and simultaneously a violent perturbation of the magnetic needle.

Meanwhile, many physicists in the second half of the nineteenth century were convinced that the aurora was an electrical discharge phenomenon based on some similarity between the auroral phenomena and phenomena observed in high-voltage vacuum discharge tubes. Electrical discharge in a vacuum tube was one of the hottest topics among physicists in those days. Their studies eventually led to the discovery of electrons by J.J. Thomson in 1897. In the classic book on gaseous discharges *Conduction of Electricity through Gases,* J.J. and G.P. Thompson (1903) noted:

We may, thus, regard the Sun, and probably any luminous star, as a source of negatively electrified particles that stream through the solar and stellar systems. Now, when electrons moving at a high speed pass through a gas they make it luminous; thus, when the electrons from the Sun meet the upper regions of the Earth's atmosphere they will produce luminous effects. Arrhenius has shown that we can explain, in a satisfactory manner, many of the periodic variations in the Aurora Borealis; if we assume that it is caused by electrons from the Sun passing through the upper regions of the Earth's atmosphere.

Recognizing the possible association between the aurora and an electron beam, K. Birkeland became one of the first proponents of what was once called the *corpuscular school* the students of which proposed that the aurora and geomagnetic storms were caused by an electron beam ejected from the Sun. Birkeland (1908) stated:

It has gradually come to be recognized that aurora and magnetic perturbations should be regarded as rather moderate manifestations – at present, the only ones there are for us to observe – of an unknown cosmic agent of solar origin, and quite different from light, heat, or gravitation. It has long been supposed that this unknown aspect was in some way or other of an electrical nature.

Birkeland's observational, analytical, laboratory, and theoretical activities have been well documented in his three-inch-thick book, *On the Cause of Geomagnetic Storms and the Origin of Terrestrial Magnetism.* I am fortunate to have a copy of this book; it was given to me by the Committee of the Birkeland Symposium on Aurora and Magnetic Storms, which was held at Sandefjord, Norway, on September 18-22, 1967.

Birkeland's *terrella experiment* is displayed at the Auroral Observatory, University of Tromsø (Figures 1.1a and 1.1b). Stimulated by Birkeland, C.

Störmer began his lifelong research on the aurora (Figures 1.2a and 1.2b). He computed trajectories of single electrons in a dipole field by devising a special integration method of the equation of motion. He also made extensive observations of the aurora and he summarized his research in his book *The Polar Aurora* (1955). He dedicated the book to his wife: *To my wife Ada who never ceased to encourage me to work hard till this book was safely finished.*

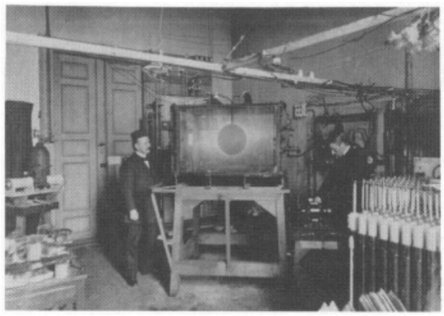


Figure 1.1a K. Birkeland and Olav Devik experimenting on the terrella in a large vacuum box: University of Oslo.

Sydney Chapman (1918) also considered theoretically how a beam of either positive or negative charged particles could produce a global motion of the upper atmosphere as a cause of geomagnetic storms. The title of his paper was *An Outline of a Theory of Magnetic Storms.* He noted later (1967):

> *I certainly misnamed this paper in calling it "An outline of a theory of magnetic storm." The observational part was useful, the theory was quite phony...*

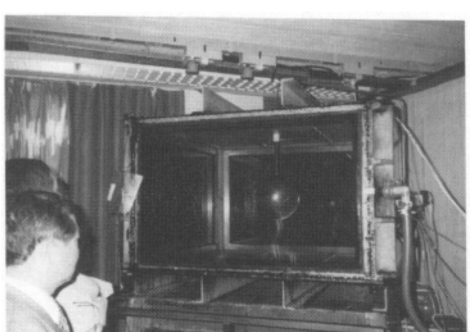


Figure 1.1b Terrella experiment at the University of Tromsø: N. Fukushima.

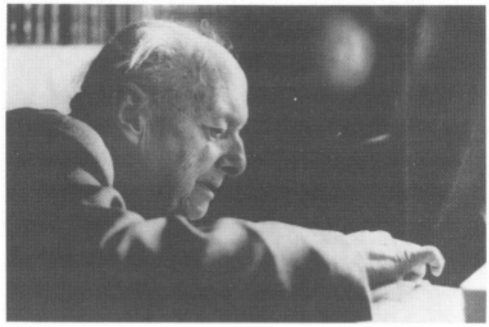
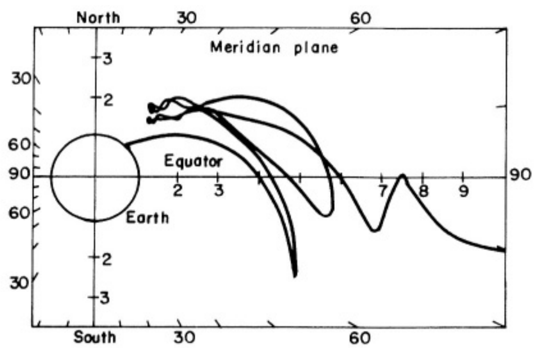


Figure 1.2a C. Störmer (1874-1957): University of Oslo.



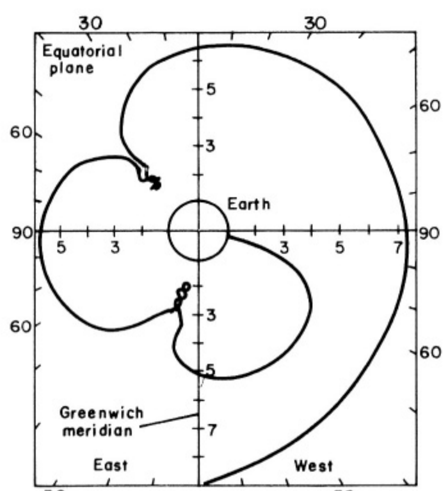


Figure 1.2b An example of the trajectory of an electron in a dipole field computed by C. Störmer: K.G. McCracken, U.R. Rao, and M.A. Shea (1962).

Eventually, A. Schuster (1911), Chapman's professor, became aware of the possibility of the existence of beams of charged particles and stated:

This verdict (Lord Kelvin's argument) was generally accepted until recently, when the theory of a direct solar action has been revived in a form, which is assumed to be free from the objection raised. The magnetic actions being supposed to be due to a swarm of electrified corpuscles ejected by the Sun.

Similarly, F.A. Lindemann (1919) stated:

There seems to be no doubt that terrestrial magnetic storms are connected in some way with solar disturbances.

On the other hand, both Schuster and Lindemann noted that something is not quite right in this idea and criticized Chapman's theory based on a beam of positive or negative charged particles. Schuster (1911) stated:

...We must conclude that a swarm of electrons packed with sufficient density to cause a magnetic effect would soon get dissipated laterally into space until its magnetic action becomes negligible.

Lindemann (1919) examined the ionization rate of a hydrogen gas cloud by developing an equation similar to the Saha equation, which provides the ionization rate of a gas for a given temperature. Based on it, he concluded that the solar gas is highly ionized. Thus, he pointed out the importance of the electrostatic force between positive ions and electrons, which can prevent the repulsion and the subsequent diversion of charged particles of one sign. This suggestion became the basis for Chapman to launch a fresh approach to the problem of the interaction between solar corpuscles and the Earth's dipole field.

Chapman and Ferraro (1931) developed the first theory of magnetosphere formation by considering the solar corpuscles as a gas consisting of positive ions and electrons in equal number. Such a gas is now called *plasma.* The term plasma had been introduced by I. Langmuir (A.H. Rosenfeld, 1966) for a gas consisting of an equal number of positive and negative ions, but they were not aware of Langmuir's work.

During the first half of the twentieth century, interplanetary space was considered practically a vacuum, and it was believed solar corpuscular streams or clouds were emitted from time to time from two kinds of source regions. The first, discovered by Maunder, is the spot-free region which emits a stream of plasma for many months. Julius Bartels later named this region the M Region. It was said that "M" signifies *magnetically active* or *mysterious,* because it causes magnetic disturbances in spite of the fact that it is a region with no spots. This region is now called the *coronal hole.* The second one is an active region around a large sunspot group. In modern terms, an intense transient activity in the solar atmosphere ejects an isolated magnetic cloud (detached from the Sun), or produces an expanding magnetic rope or loop (rooted on the Sun) during an intense solar flare or a coronal mass ejection (CME).

It may be noted that the solar flare-geomagnetic storm relationship was doubted by some even in the 1950s. This was because flares near the limb of the solar disk do not necessarily cause major magnetic disturbances; the center of such flare disturbances are directed 90° away from the Sun-Earth line. H.W. Newton clarified this point. I mentioned his work on this issue in one of my early papers. Newton's wife found my paper and read it to him (he was blind by then). He was very pleased, and sent me a copy of his book *The Face of the Sun* (Penguin Books; 1958). It is a delightful book that I recommend to today's investigators. Every sunspot cycle produces a new generation of solar-terrestrial scientists who are most welcome to the discipline. However, many do not learn what earlier sunspot generation researchers learned. As a result, many phenomena are rediscovered every sunspot cycle.

1.2 The Chapman-Ferraro Theory

Chapman took up a theoretical study of the interaction between Maunder's stream and the Earth's dipole field. Chapman and his graduate student Vincent Ferraro recognized that Maunder's stream should be treated as what we today call plasma, not a swarm of protons (or electrons). Here is Chapman's account of the birth of their theory (Chapman, 1968):

> *In my first year, there were only two honors students in mathematics....The other student was an Anglo-Italian, Vincenzo Antonino Ferraro whose father had emigrated from Italy and was in the hotel business; he was manager or head of one of the big high-class restaurants in London. ...His son was born in England, but he was very much influenced by his family; I think the mother was a dominant figure, too. He went to a good English school in London, and then came to the Imperial College, and did very well in the degree examination and went on to do research.*

By that time, the Ph.D. had become established in England. So I was expected to provide him with a problem and guide him in it. Now, Larmor would never have given to me the suggestion of tackling the central problem in kinetic theory, which was the one that I myself discovered and sought to attack and solve. I felt I was doing a dangerous thing in giving Ferraro the problem I chose and asking him to work on the causation of magnetic storms. We tried to work out deductively what would be the consequence of the impact upon the Earth of a stream of what is now called plasma, neutral ionized gas. This had been suggested by Lindemann when he criticized my first phony theory on magnetic storms. He had not only destroyed my theory, he had proposed the constructive alternative suggestion, that the influence from the Sun must not be as I had supposed (like Birkeland and Störmer and others); namely, a stream of gas formed of charges of one sign only. He said it must consist of charges of opposite signs in practically equal numbers, so that it could hold together.

Lindemann never tried to develop what would be the consequences on the Earth of the impact of such a stream of gas. I made an attempt at that while I was Professor at Manchester in 1919-1924, but, unfortunately, I started at the wrong end. I tried to find out what would be the steady state, as if the stream had been going on forever. It didn't work out; so I was still wanting to find out what would happen. I proposed this subject to Ferraro. We played about with this problem, often being quite at a loss to know what would happen and how to approach the problem. But, finally, there did come the breakthrough of realizing that the stream would be a good conductor. Looking into Maxwell's great work on electricity and magnetism, and using at first a very crude model, namely, the approach of a conducting metal sheet towards Earth, we considered what currents would be induced in this sheet by the approach to Earth's magnetic field, and would add their own field in the space around the Earth. There would be a repulsion between the sheet currents and the field, which would tend to slow down the sheet. The sheet in this model was rigid. If you think of it as a gas instead of a rigid sheet, the current having at first a plane front, the current would be induced in its surface, but owing to the unequal repulsion in different parts, as we realized, the sheet would be subject to weaker forces. So, we inferred that a cavity would be formed around the Earth, enclosing the magnetic field – which would be confined in this cavity. This was the first published note in "Nature" in 1930; Ferraro came to the College in 1924, so this was in his sixth year. The work was later published in full, over a period of time, in 1931-32, in the journal that is now called the "Journal of Geophysical Research."

For a long time people didn't believe in it or they were very dubious about it, didn't read it, or took no notice of it. Not long before it was actually demonstrated by satellites that this is what does happen, it came to be at last considered, and on the whole, accepted by a number of people. But, one of my American friends, Hulburt, who was an excellent director of the Naval Research Laboratory, developed an alternative theory of Ferraro's and mine, in one of his publications, as a matter now only of historical interest. I criticized Hulbert's theory, and I think it is quite dead now, as dead as my first theory of magnetic storms. However, Hulburt and I are very good friends.

In his article in *The Earth and Science* edited by A.M. Cook and T.F. Gastell for the occasion of Chapman's eightieth birthday, Vincenzo Ferraro wrote:

In 1927, I became one of Chapman's first research students at Imperial College; this was the beginning of a long and fruitful collaboration which afforded me much pleasure and in which, Chapman once told me, he much enjoyed, as indeed he must have enjoyed his collaboration with other people. Chapman and I undertook afresh the problem of the approach of the neutral ionized stream in the Earth's magnetic field and during the years 1927-33 we developed a new theory of magnetic storms. Only the theory of the first phase was then fully developed. We found that during its advance in the Earth's magnetic field electric currents are induced in the surface of the solar corpuscular stream. The surface currents shield the interior of the stream from the Earth's magnetic field so that particles in the stream can describe a rectilinear path up to the point where they enter the surface (current) layer of the streams. The action of the Earth's magnetic field on the surface currents repels the surface of the stream, the retardation being greatest over the part of the surface nearest the Earth. A cavity is thereby formed in the surface of stream, which deepens until equilibrium is reached between the kinetic and magnetic pressures. The geomagnetic field is thereby compressed by the solar cavity, the resulting increase in the horizontal force at the Earth's surface being identified as the increase in the horizontal force during the first phase of a magnetic storm.

In the first of a series of their papers on this subject, Chapman and Ferraro (1931) obtained a formula, which is basically similar to the Debye length (p. 94 in their paper), and showed that protons and electrons in the stream are strongly coupled in motion as they flow around the Earth's dipole field. Thus, they showed that the solar gas must be treated as *plasma,* not as a cloud of solitary particles.

The basis of the Chapman-Ferraro theory is to regard the solar corpuscular stream as diamagnetic superconducting plasma. Thus, the gas cannot penetrate into the Earth's magnetic field, as strong shielding currents flow on the front surface of the advancing stream. This current is called the *Chapman-Ferraro current* (Figures 1.3a, 1.3b, 1.3c, 1.3d). As a result, the Earth and its magnetic field are completely confined or compressed in a cavity. In this way, Chapman and Ferraro explained successfully the storm sudden commencement (SSC) as a result of the impact of the solar plasma on the Earth's dipole field. In modern terms, the SSC is caused by the impact of a shock wave on the magnetosphere. The shock wave is generated when a high-speed solar plasma cloud or loop advances into the background slow-speed solar wind. It may be noted that Chapman and Ferraro correctly envisioned the geometry of what we now call a high-speed stream from the coronal hole.

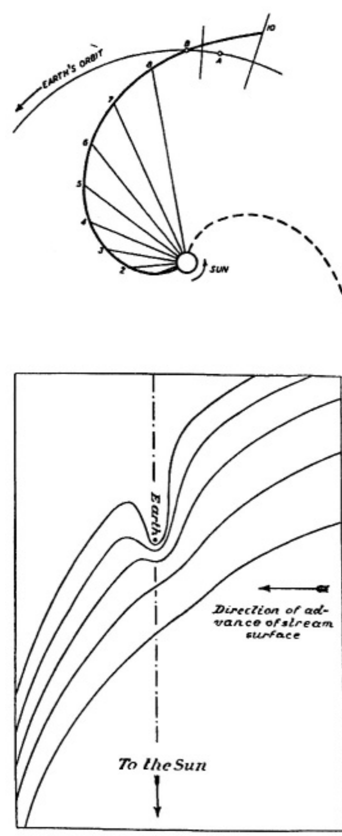


Figure 1.3a, b The geometry of a stream from the Sun: S. Chapman and J. Bartels (1940).

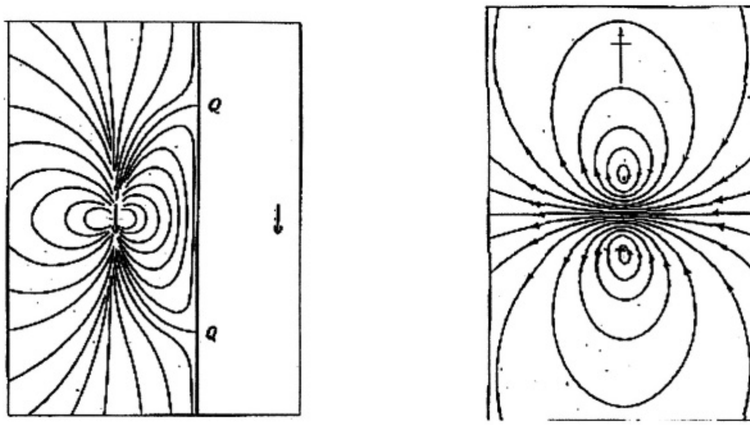


Figure 1.3c, d The compressed Earth's magnetic field and the shielding current (the Chapman-Ferraro current) on the advancing front of the stream: S. Chapman and V.C.A. Ferraro (1931).

However, it was unfortunate that Chapman was convinced by this success that a theory of the main phase of geomagnetic storms (namely, the formation of the ring current belt) should be built upon the Chapman-Ferraro theory, namely, the interaction between *diamagnetic* plasma and the Earth's magnetic field. As is now known, an efficient energy transfer of as much as $10^{19} \sim 10^{20}$ erg/sec from the solar wind to the magnetosphere cannot be achieved by a diamagnetic plasma, because the diamagnetic plasma stream tends to flow around the magnetosphere without introducing much energy into it. In fact, Chapman agonized for almost 30 years after the Chapman-Ferraro theory was established because he was unable to find the energy transfer mechanism from the solar wind into the magnetosphere for almost thirty years after the Chapman-Ferraro theory was successfully established.

However, Chapman was considerably encouraged when he learned that the Explorer 12 satellite crossed the magnetopause in 1962, and demonstrated that the magnetic field just inside the magnetopause is close to twice that of the dipole field value, as expected from the Chapman-Ferraro theory. This satellite observation event is described in detail by Larry Cahill (1997). I recall that Chapman mentioned that it is rather rare that a theory becomes confirmed more than thirty years after its inception. He was all the more convinced that a theory explaining the main phase must be an extension of the Chapman-Ferraro theory. This was an exciting period during the early days of space exploration. I personally witnessed the occasion when Explorer 10 crossed the magnetopause. Jim Heppner, Norman Ness, and Joe Cain`, at the Goddard Space Flight Center, were trying to understand this newly observed phenomenon (Heppner et al., 1963).

1.3 The Solar Wind

The view that interplanetary space is a vacuum into which the Sun intermittently emitted corpuscular streams was changed radically by Ludwig Biermann (1951, 1953) who proposed, on the basis of his study of comet tails, that the Sun continuously blows its atmosphere out in all directions at supersonic speed. At that time, it had generally been accepted that the solar radiation pressure was responsible for the tendency of the comet tail to trail in the antisolar direction. However, referring to the fact that the magnitude of acceleration of the ionized component of cometary tails is of the order of ~10^2–10^3 times solar gravity, he concluded:

> *It is, thus, found that the acceleration of the CO^+ and N_2^+ formations observed in the tails are easily explained in terms of the friction between the solar and the cometary ions.*

In 1957-58, Chapman was interested in zodiacal light and tried to explain it in terms of an extended *static* solar corona, but told me that he could not publish his paper in a regular scientific journal. Later, Gene Parker pointed out that Chapman's solution has a problem in that the pressure of his corona is finite at infinity.

In order to explain Biermann's conclusion, Parker (1958) examined thermal conditions of the solar corona that could lead to a supersonic flow at the Earth's distance and found that a supersonic flow can occur if the temperature in the solar corona decreases less rapidly than $1/r$, where r denotes the solar radius. This conclusion requires that the corona must be heated over a very extensive height range. In suggesting such a possibility, he coined the term *solar wind.*

The first observations of the solar wind by the Mariner 2 space probe were considered to be the confirmation of Parker's theory (Marcia Neugebauer and C.W. Snyder, 1962). During the period between 1960 and 1980, his idea of the generation of the solar wind was considerably rearticulated and elaborated by a number of solar physicists. However, there is so far no conclusive theory on both the high temperature of the corona and the acceleration process of the solar wind. It is puzzling that a strong solar wind tends to flow out from a dark region in the solar corona, what we now call a coronal hole, although it is known to be a magnetically open region; the magnetic field lines originating in a coronal hole extend into interplanetary space rather than forming closed loops in the corona. The source process of the solar wind is still one of the long-standing unsolved problems of space physics.

1.4 Interplanetary Shock Waves

Tomy Gold (1955) intuitively associated the sudden rise of the horizontal component of the geomagnetic field at storm onset (SSC) with the impact of a shock wave that propagates in interplanetary gas. During the symposium titled *Gas Dynamics of Cosmic Clouds,* held at Cambridge, England, in 1953, he stated:

> *I should like to discuss, in connection with the subject of shock waves, some of the magnetic disturbances on the Earth that are caused by solar outbursts. The initial magnetic disturbance at "Sudden Commencement" of a magnetic storm can be accounted for, very roughly, by an increase of pressure of the tenuous gas around the Earth. This increase of pressure may perhaps be described as the effect of a wave sent out by the Sun through the tenuous medium. This description would then correspond to that of a stream of particles, while in the presence of a medium the correct description may lie anywhere between an acoustic wave, a supersonic shock wave, or an unimpeded corpuscular stream. The observations of magnetic storms may, hence, give us a fairly direct proof of the existence of shock waves in the interplanetary medium.*

However, H.W. Liepmann objected to Gold's suggestion (1955). Liepmann argued:

> *I would like to ask whether the picture of a shock wave really is applicable. The mean free path in the residual gas between the Sun and the Earth appears to be 4 or 5 times the solar radius...In order to get agreement with Gold's values, the mean free path would have to be considerably shorter, i.e., by a ratio of about 100, or else the mechanism of interaction of the wave with the field of the Earth has to explain the very sudden rise observed...*

Gold refuted this objection by stating:

In considering the interaction between the stream and the residual gas, one must not restrict oneself to the collision cross section of neutral particles, but one has to consider the much stronger electromagnetic interactions that may occur between the two ionized gases.

Gold's view has since been elucidated in great detail by researchers. Some solar wind ions are accelerated by colliding with both the advancing shock wave and the shock wave that forms at the front of the magnetosphere.

1.5 The Modern Interpretation of the Chapman-Ferraro Theory

Chapman and Ferraro theorized, first, the storm sudden commencement (SSC) and the initial phase in terms of encounter by the Earth and its magnetic field with a discrete solar corpuscular stream. However, since the Sun is known to expel the solar wind continuously, the Earth and its magnetic field are always confined in a comet-like cavity (Figure 1.4). As mentioned earlier, this cavity is called the magnetosphere. Since Chapman and Ferraro predicted the cavity, their theory of the SSC and initial phase was later regarded as the first theory of the formation of the magnetosphere. The SSC and initial phase of a geomagnetic storm can now be explained as a simple extension of their theory; it is caused by an enhanced solar wind pressure associated with the shock wave's compressing the magnetosphere (Figure 1.5), not by an impact of isolated solar gas clouds or streams. Although these findings may sound trivial from the present understanding of the magnetosphere, each finding was an epoch-making advance in magnetospheric physics in its early days of development.

1.6 The Main Phase of Geomagnetic Storms and the Ring Current

Chapman and Ferraro thought that the morning side of the boundary of the magnetosphere, the magnetopause, is slightly positively charged, while the evening side of the magnetopause is negatively charged.

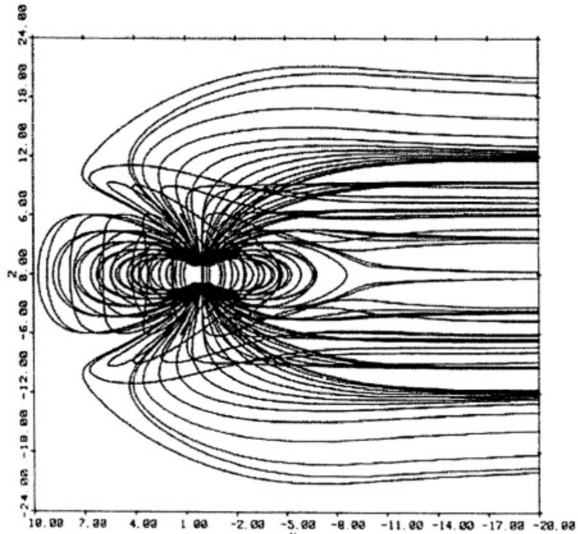


Figure 1.4 The Chapman-Ferraro theory predicted that the magnetic field configuration in the magnetosphere is completely confined in a comet-shaped cavity: M. Roederer.

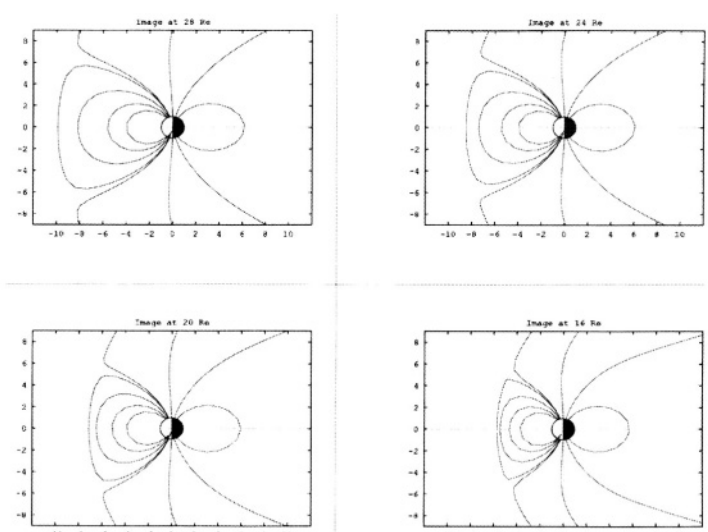


Figure 1.5 The arrival of an interplanetary shock wave compresses the magnetosphere (the noon-midnight cross-section): S.-I. Akasofu.

They suggested that a dawn-dusk electric field, thus established, brings some of the particles into the cavity. Several theorists have independently considered this idea a few times during the last thirty years. Chapman and Ferraro thought that the particles thus brought into the magnetosphere form a toroidal westward-directed ringlike current around the Earth, explaining the large decrease in the horizontal component of the Earth's field during the main phase.

Unfortunately, as mentioned earlier, after his initial success in 1931, Chapman did not make much progress in explaining the development of the main phase of geomagnetic storms and was still struggling with the problem when I joined him in Alaska in 1958.

I never took the opportunity to ask Chapman why he decided to come to Alaska after retiring from Oxford University. One possibility was that Störmer criticized his theory during the 1950 London, Ontario, conference – the first major conference on the aurora after WWII – saying that the Chapman-Ferraro theory cannot explain any specific aspect of the aurora. Störmer claimed that his theory, on the other hand, could explain details of many aspects of the aurora, including the auroral zone and the curtain-like structure of the aurora. Although the title of Chapman's paper was *Theories of the Aurora Polaris,* all he could do was to describe the Chapman-Ferraro theory; he could not find the mechanism by which the solar wind transfers its energy to the magnetosphere in causing the aurora. In Alaska, Chapman could continue to search for that mechanism.

When I was a graduate student at Tohoku University, Japan, an organization called *The Ionospheric Research Committee* was dedicated to research on solar-terrestrial physics. Top-level researchers in Japan attended its meetings and their discussions were stimulating for the young scientists. It was said in their meetings that a good understanding of the Chapman-Ferraro papers was a prerequisite to studying geomagnetic storms. Thus, I began that study and found the papers difficult to grasp, leaving me with a number of questions. I learned that Chapman worked at the Geophysical Institute, University of Alaska, and wrote to him in Spring 1958. I included in my letter the questions I had and did not expect a quick response. Chapman was the greatest authority on geomagnetism, contemporary of the great British astronomer A. Eddington, and I was simply a student who had just started learning geomagnetism. Thus, it was a great surprise to receive his letter in a matter of a few weeks. In his response, he said in essence that he could not answer some of my questions and asked if I would be interested in studying those questions under his guidance.

His response was totally unexpected. I wrote to him immediately, saying how delighted I was, but that I was too poor to study abroad. To my great surprise again, I received a check from Chris Elvey, director of the Geophysical Institute, soon afterward. By then, however, I had been asked to be a member of the Japanese solar eclipse expedition party to go to the South Pacific. Thus, it was not until December 13, 1958, that I arrived in Fairbanks, Alaska.

Soon after my arrival in Alaska, the Van Allen belts were discovered. It became the dawn of scientific space exploration. James Van Allen correctly pointed out that energetic particles execute the motions studied by Störmer in his pioneering work on motions of charged particles in a dipole field (Figure 1.6). At the time of the discovery of the Van Allen belts, several researchers, including Fred Singer, suggested those motions constituted a westerly electric current around the Earth, causing the main phase of the geomagnetic storms.

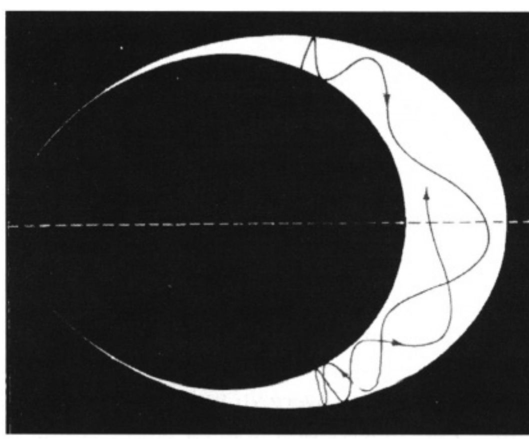


Figure 1.6 C. Störmer found a group of trajectories of electrons which are confined in a region isolated from infinity. An example of the trajectories is shown. James Van Allen thought that high-energy particles in the radiation have similar trajectories: J.A. Van Allen (1962).

In my Ph.D. thesis work (1961) I examined quantitatively the magnetic field produced by trapped particles in the Earth's dipole field, although I had to postulate the growth of what I called the *storm time radiation belts,* which consist of protons of a few kilovolts. The computed magnetic field, produced by the motions of protons in the belts, was found to point almost uniformly southward around the Earth, explaining the large depression of the field during the main phase (Akasofu and Chapman 1961); see Figure 1.7. It should be noted that, unlike what most researchers believe today, the ring current field arises mainly from diamagnetism of the trapped particles in the dipole field, not the westward drift motion of positive ions. In fact, for an isotropic pitch angle distribution, the westward drift does not contribute to the current. In the inner half of the belt, the current flows eastward, while in the outer half, the current flows westward. The outer current is more intense than the inner current. It is for this reason that the ring current belt flows westward. An IBM 7090 computer at the Goddard Space Flight Center was used for this computation; it was the fastest computer of its time. The computed results agreed with the ground-based observations and some of the earliest satellite observations of the magnetic field produced by the ring current reported by Paul Coleman and Larry Cahill. Further, the belt I postulated was surprisingly not too far from what Lou Frank, University of Iowa, detected later with a satellite.

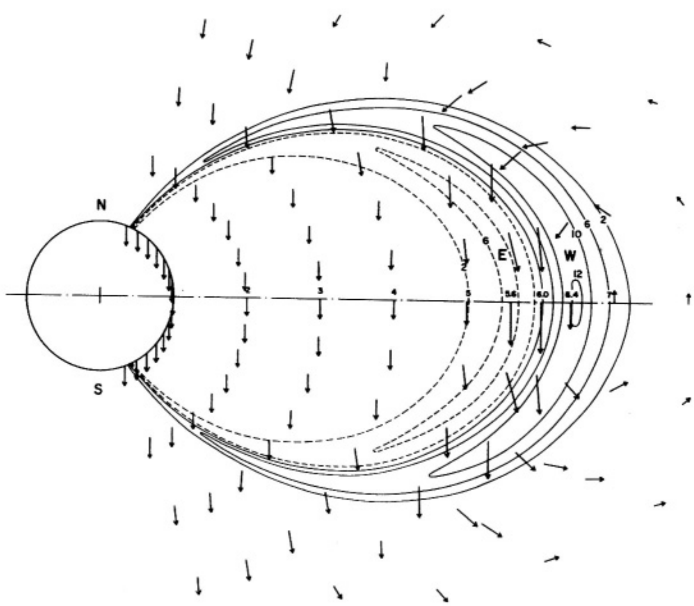


Figure 1.7 The distribution of electric currents produced by a belt of trapped particles in the Earth's dipole field (E, eastward and W, westward) and the magnetic field produced by the currents: S.-I. Akasofu, J.C. Cain, and S. Chapman (1961).

1.7 Variety of the Development of Geomagnetic Storms

Chapman and I were quite happy about the computed results. Confirming that protons of a few kilo electron-volts in the trapping region can produce the desired westward current for the main phase decrease, we tried to find ways to bring solar wind protons deep into the trapping region across the magnetopause on the basis of the Chapman-Ferraro theory, but finally concluded that the Chapman-Ferraro theory actually tells us that a diamagnetic plasma tends to flow around the Earth, confining it into a cavity without transferring much energy into the magnetosphere. There is no way to bring solar wind protons to a distance of several Earth radii across the dayside magnetopause. However, the prevailing idea at that time was that only a tiny fraction of the solar wind energy was needed to cause the main phase, and that the problem should not be difficult to solve (Alex Dessler, and his colleagues, 1961). So our conclusion did not get any attention in the community.

After much thinking, I proposed to Chapman that I should examine the development of a number of *individual* geomagnetic storms, instead of a typical or an idealized storm, in an effort to study how the main phase actually develops. It immediately became evident that geomagnetic storms develop in a great variety of ways. In order to demonstrate the point, we published a paper that included Figure 1.8a (Akasofu and Chapman, 1963). If I had to choose three of the most important figures published in my research career, this would be one of them, although I believe it would be impossible to publish such an unsophisticated figure in the *Journal of Geophysical Research* today. Figure 1.8b shows also the great variety of ways in which geomagnetic storms can develop. In the first storm the main phase did not develop, in spite of the fact that a strong solar wind blew around the magnetosphere for many hours after the SSC, as can be seen by the fact that the increased field level was maintained for many hours.

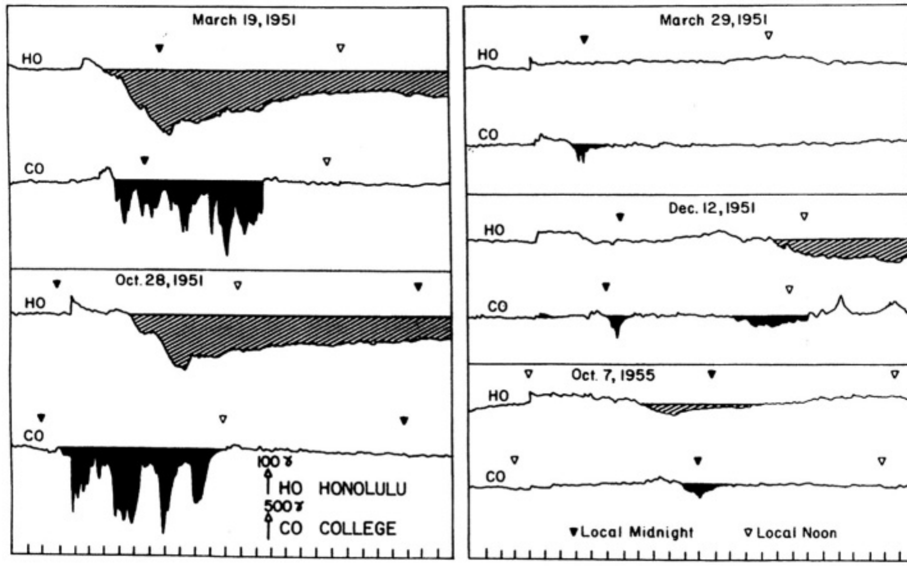


Figure 1.8a Two types of geomagnetic storms which begin with SSC of similar magnitudes. Left: Storms with an intense main phase (HO: the Honolulu record) and polar magnetic substorms (CO: the College record); Right: Storms with weak or no appreciable main phase and substorms: S.-I. Akasofu, and S. Chapman (1963).

The third one is what has been thought to be a *typical* storm as Chapman conceptualized it. However, in many cases, the main phase can start to develop even before the SSC. In many other cases, a large main phase can develop without the SSC (see the last example in Figure 1.8b).

This type is known as *gradually commencing storms,* but was ignored because they are not *typical.* In fact, a major geomagnetic storm can develop even after what is known to be a *negative sudden impulse,* namely a sudden decrease of the solar wind pressure. In modern terms, such a sudden change is associated with the passage of an interplanetary discontinuity in the solar wind. An obvious conclusion from this study was that an increased solar wind pressure is not a necessary and sufficient condition for a geomagnetic storm to occur. However, such a conclusion was not acceptable to the scientific community at that time.

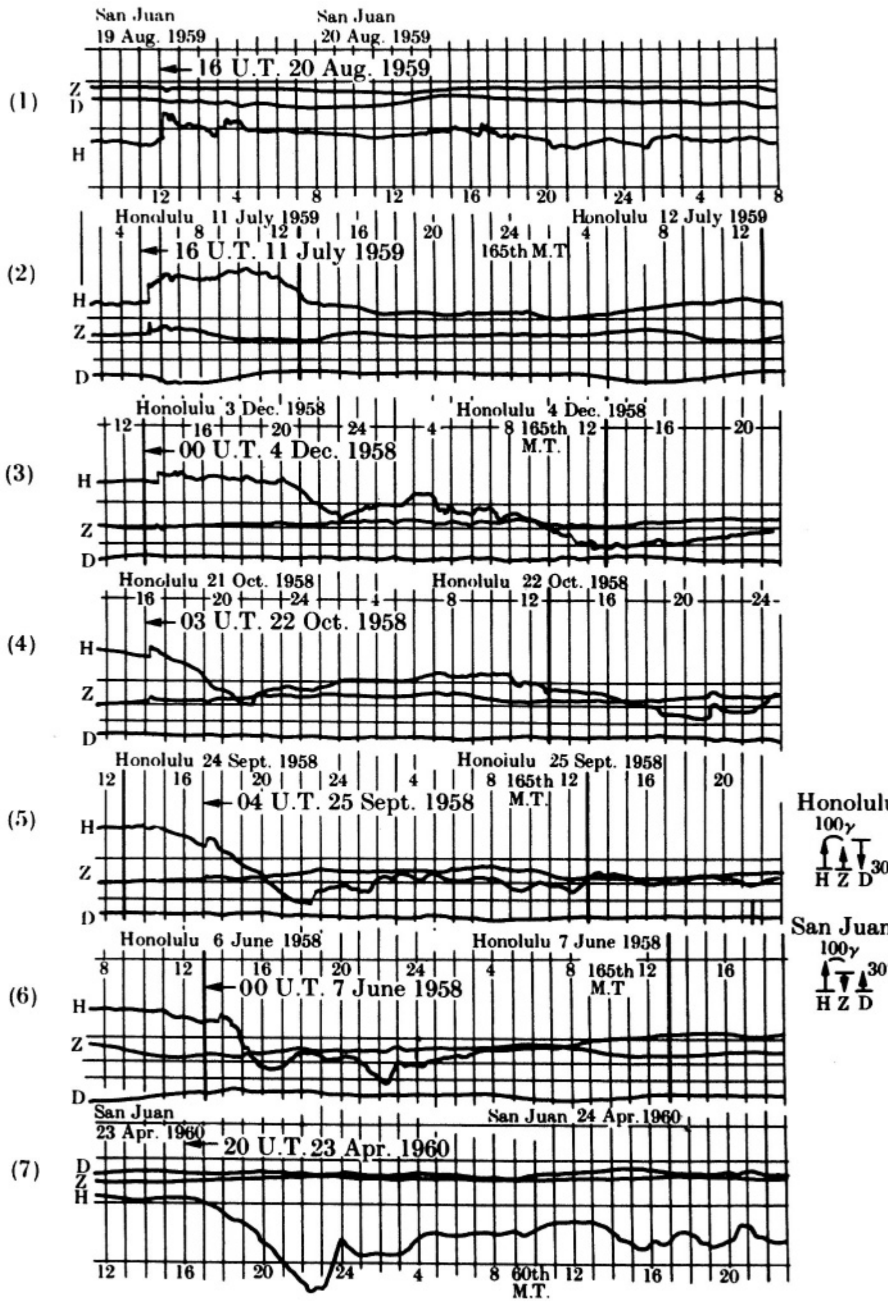


Figure 1.8b A collection of magnetic records showing a great variety of ways in which geomagnetic storms develop: S.-I. Akasofu and S. Chapman (1972).

1.8 Unknown Quantity

After much discussion, Chapman and I reached an important realization: the Chapman-Ferraro theory implies that a diamagnetic plasma flow around the Earth can confine it in a cavity (the magnetosphere), but does not transfer the energy into it. We concluded (Akasofu and Chapman, 1963, p. 129):

> *The variety of development of the storms seems to suggest some intrinsic differences between the solar streams far beyond what we would expect from a mere difference between their pressures. The nature of their intrinsic differences is at present unknown.*

This conclusion annoyed and even outraged some prominent theorists, since it was so firmly believed that the Chapman-Ferraro theory was all that was needed and that a geomagnetic storm is a result of a stronger solar wind. Some even told Chapman that he was trying to destroy his own life work. Most researchers thought that there could not be any unknown quantity in physics, except for some elementary particles.

Nevertheless, we thought that the theory of the main phase must be built upon the Chapman-Ferraro theory, because it is successful in explaining the SSC. Thus, I once thought that the unknown quantity was neutral hydrogen atoms, which can penetrate across the dayside magnetopause without difficulty and can become energetic protons of the ring current belt after a charge exchange process (Akasofu, 1964). In Figure 1.9, the variety of the development of magnetic storms shown in Figure 1.8 was interpreted in terms of the degree of the ionization of the solar wind. The first type is caused by fully ionized plasma, which corresponds to the case of the Chapman-Ferraro theory. On the other hand, the last type is produced by essentially un-ionized plasma atoms (which do not cause the compression (SSC) of the magnetosphere, but cause the main phase after penetrating into the magnetosphere and exchanging the charge).

As will be explained in the next section, this unknown quantity is now identified as the southward component (-Bz) of the interplanetary magnetic field (IMF), or more accurately, a specific combination of the solar wind speed V, the IMF magnitude B, and its polar angle θ. The magnetosphere responds to the southward component of the IMF lasting for a few hours in a very specific way. This mode of magnetospheric disturbance is called the *magnetospheric substorm.* The polar magnetic substorm is the magnetic manifestation of it, while its auroral manifestation is called the auroral substorm.

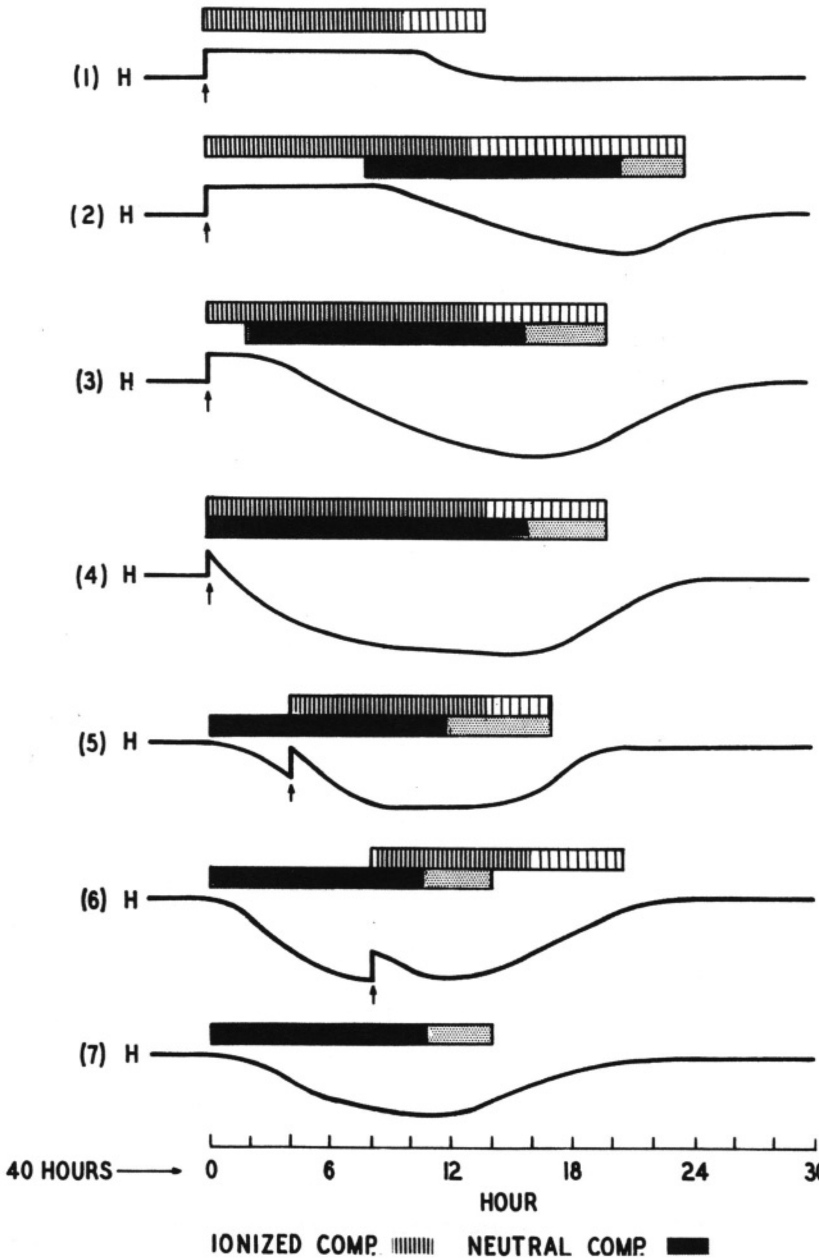


Figure 1.9 An attempt to explain the great variety of development of geomagnetic storms in terms of differences of the ionization rate of the solar plasma flow: S.-I. Akasofu (1981).

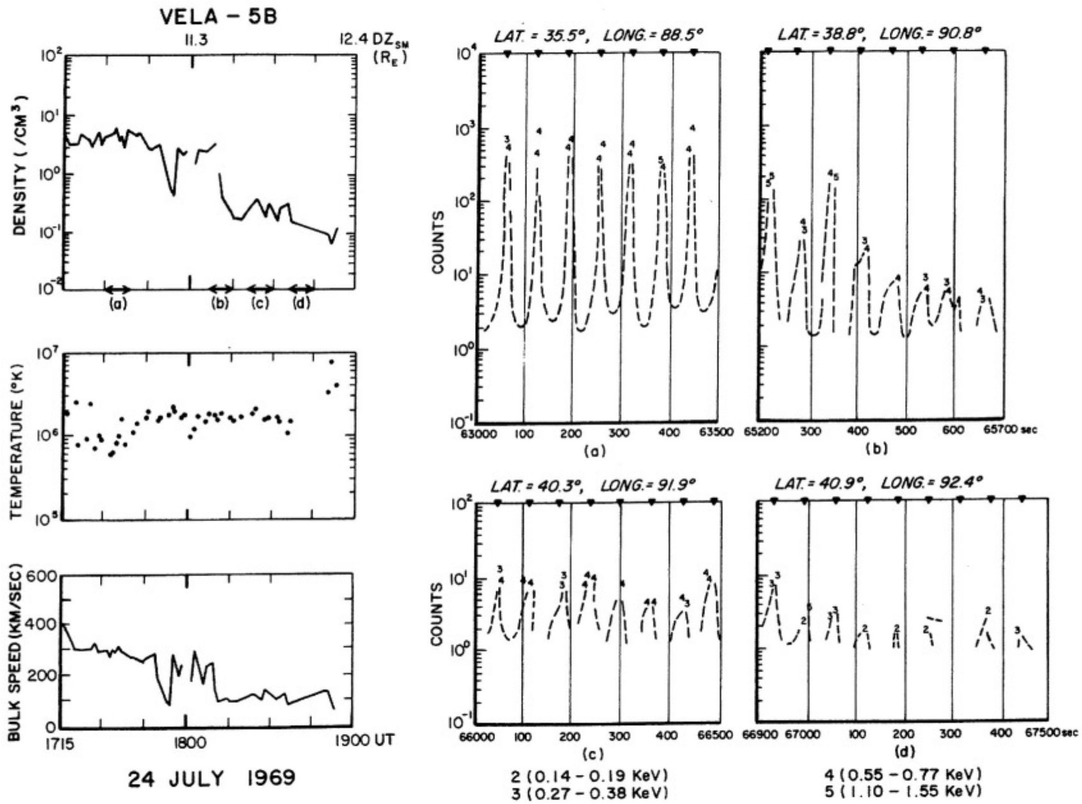


Figure 1.10 Example of observation of the plasma flow as a satellite (Vela-5B) enters from the solar wind regime into the magnetosphere. As the satellite spins, the unidirectional solar wind produces a large-intensity modulation in the detector (a). However, even after the satellite entered into the magnetosphere (b), a weak flow from the direction of the Sun is still observed (c and d): S.-I. Akasofu, E.W. Hones, Jr., S.J. Bame, J.R. Asbridge, and A.Y. Lui (1973).

Incidentally, there was an interesting aftermath of this neutral hydrogen story. When I was studying magnetotail phenomena in Los Alamos with Ed Hones, I found an antisunward flow of particles in the lobe of the magnetotail (Figure 1.10). At that time, as the Chapman-Ferraro theory indicated, the fully ionized solar wind was thought to flow around the magnetosphere, so that the entrained particles could not flow inside the magnetosphere and the magnetotail. One obvious interpretation of the observation was that this flow was composed of neutral hydrogen atoms across the magnetopause; they could become ionized by colliding with the detector. After a few sleepless nights, however, I found that a simple calculation showed that such a possibility was unlikely. On the other hand, this finding led to the discovery of the mantle flow (1973). Speaking of neutral hydrogen atoms, I should point out a common misconception, which is that a solar prominence is fully ionized plasma. Actually, it is only partially ionized plasma, because the observed prominence emissions are from neutral hydrogen atoms (the Balmer alpha line). Further, it is clear from the observations of exploding prominences that neutral hydrogen atoms can escape from the Sun before they become ionized. Unfortunately, so far there has been no attempt to observe them in interplanetary space.

I became more convinced of the validity of our conclusion on the existence of the unknown quantity when I examined the intense magnetic storm illustrated in the Honolulu and College magnetic records shown in Figure 1.11. After the sudden commencement at about 13:40 165° Local Mean Time (LMT) on December 3, 1958, a strong solar wind blew for as long as 6 hours, but the main phase began to develop only after 20:10 165° LMT without an additional large enhancement of the solar wind pressure, which would be recorded in the horizontal component of magnetic records if it happened; intense auroral activities also began at the same time. Some unknown quantity in the solar wind must have arrived around the Earth at that moment to cause the main phase and the auroral activity. Thus, our research for the unknown quantity began.

1.9 The ε Parameter

A new understanding of the energy transfer process from the solar wind to the magnetosphere began when Don Fairfield (1967) found a close association between the so-called *southward turning* of the interplanetary magnetic field (IMF) vector and geomagnetic disturbances, or the southward component of the interplanetary magnetic field. Fairfield concluded that the IMF southward component can be identified with what Chapman and I thought to be the unknown quantity. This finding was based on Jim Dungey's suggestion of magnetic reconnection; Dungey credited this to F. Hoyle. He elaborated on his suggestion later and published a paper in *Physical Review Letter* in 1971. However, the significance of his open model in substorm processes was not well recognized by most magnetospheric physicists for almost ten years. In fact, in that period, many sketches of the magnetosphere model did not include the interplanetary magnetic field lines (see Figure I in the Prologue).

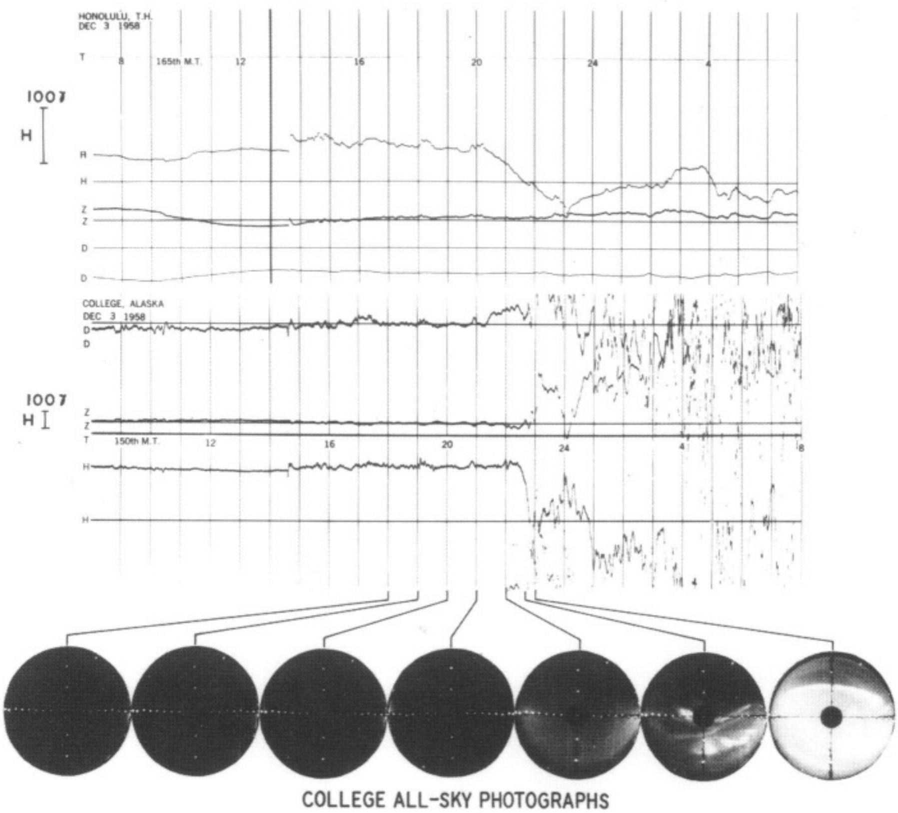


Figure 1.11 The development of the geomagnetic storm of December 4, 1958. From the top, the Honolulu magnetic record, the College magnetic record, and the College all-sky photograph on that day. Note that after the storm sudden commencement at about 1340 165° West Mean Time on December 3, the magnetosphere was exposed to a higher solar wind pressure for many hours. However, the main phase, high-latitude disturbances, and auroral activities developed 6 hours after the SSC. The nature of the solar wind must have changed at about 20 165° LMT. Note that College (Fairbanks) was dark enough to observe the aurora during the intense flow: S.-I. Akasofu (1964).

Figure 1.12 shows the development of the geomagnetic storm of February 15-16, 1967. One can see clearly the arrival of the interplanetary shock wave at about 23:50 UT on February 15, manifested by a step function-like increase of the field magnitude *B,* which nearly coincided with the SSC on the ground. However, the intense polar magnetic substorm activity (indicated by the auroral electrojet index AE) and the associated development of the main phase (indicated by the Dst-ASY index) did not begin until about 09 UT on that day. One can see clearly that this time coincided with the arrival of the southward component (-*Bz*) of the interplanetary magnetic field, namely of the *unknown quantity* suggested by Fairfield. Fairfield's study demonstrated that geomagnetic disturbances must be closely associated with changing the IMF.

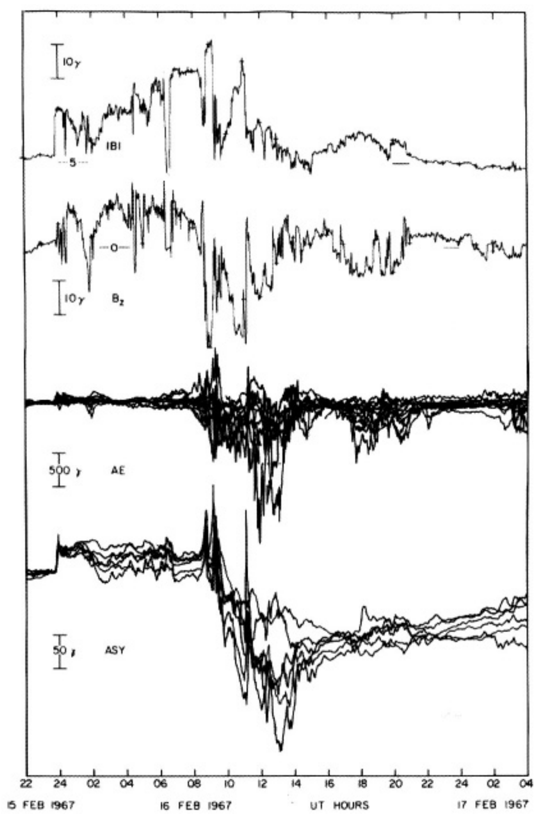


Figure 1.12 Superimposed magnetic records from both high and low latitudes, together with the interplanetary magnetic field magnitude *B* and the *Bz* component for the geomagnetic storm of February 16-17, 1967. The differences among the low- latitude observatories indicate the asymmetry of the main phase (ASY): S.-I. Akasofu (1977).

However, in the 1960s and 1970s, most theorists in magnetospheric physics were preoccupied with the hypothesis of magnetic reconnection in the magnetotail as the energy supply process for the ring current. This was under the premise that magnetic energy was *gradually* and *continuously* accumulated in the magnetotail and that *spontaneous* magnetic reconnection suddenly converted the magnetic energy thus accumulated into substorm energy. It was said that the magnetotail had more than enough energy for thirty substorms and that all we had to find was the process leading to magnetic reconnection. What Fairfield found was that each substorm requires a significant amount of input energy. However, Fairfield's paper and those that followed did not get the attention they deserved for many years.

As I mentioned earlier, there is little doubt that the energy for magnetospheric substorms is delivered from the Sun to the magnetosphere by the solar wind. Thus, in the last few decades, one of the most profound issues in magnetospheric physics, both theoretical and observational, has been to uncover the processes associated with the energy transfer from the solar wind to the magnetosphere and the subsequent transmission and conversion processes that lead to various magnetospheric substorm processes. Further, various polar upper atmospheric phenomena (such as the auroral substorm, the ionospheric substorm, the polar magnetic substorm, etc.) and also various disturbance phenomena in the inner magnetosphere and the magnetotail are mostly different manifestations of the magnetospheric substorm. Further, the magnetospheric substorm is perhaps the most basic type of magnetospheric disturbance, as a response to an increased energy input from the solar wind.

In understanding these energy transfer and conversion processes, the hypothesis of magnetic reconnection has become such a powerful paradigm that reconnection has been considered to be the cause of most magnetospheric processes. Most theorists thought they had to base their theories upon it and many experimenters felt they had to prove it. In one of the standard references on this subject, Vytenis Vasyliunas (1975) stated:

> *The process variously known as magnetic merging, magnetic field annihilation, or magnetic field line reconnection (or re-connexion), plays a crucial role in determining the most plausible, if not the only, way of tapping the energy stored in the magnetic field in order to produce large dissipative events, such as solar flares and magnetospheric substorms.*

Indeed, from 1960-1980, understanding explosive magnetic reconnection was considered to be one of the most important theoretical problems to be solved in magnetospheric physics, as documented in reports by the National Academy of Sciences, the National Aeronautics and Space Administration (NASA), and various committees. For example, Colgate et al. (1978), in the Colgate Report, state:

...This magnetic reconnection may occur gradually or explosively. When it occurs explosively, it can lead to auroral substorms and solar flares....

In the same report, magnetic reconnection is identified as the most important problem among six problems, which are:

...vital to further understanding of space plasmas...

In fact, much of the past theoretical effort has been focused on finding mechanisms that make magnetic reconnection explosive in order to explain explosive phenomena, such as solar flares and substorms. At the same time, the resulting neutral line or the X-line has become a *magic* line. Many phenomena are blindly ascribed to unknown and unproved physical processes associated with the X-line. For example, it was proposed without any definitive proof that auroral electrons were accelerated along the X-line, causing auroral arcs.

I avoided this particular paradigm and decided to go my own way. I must confess that this decision was not based on any rational thinking. It may be that I have an instinctive tendency to avoid a popular view.

In spite of this trend, until the 1970s there was no serious attempt to examine observationally how the energy input rate $I(t)$ and the output rate $O(t)$ of the magnetosphere are related on a global scale, although such a study is crucial to examining whether explosive magnetic reconnection would be responsible for the magnetospheric substorms. For the purpose of this particular study, one may consider here three systems with very different relationships between I(t) and $O(t)$. In the first system, time variations of the energy output rate $O(t)$ are almost identical to those of the energy input rate $I(t)$ (Figure 1.13a). In the second system, the energy is initially accumulated to a critical value, at which value it is suddenly unloaded. Therefore, in such a system (Figure 1.13b), $I(t)$ and $O(t)$ are expected to have different time variations. These two systems may be schematically represented by the so-called *pitcher model* and the *tippy bucket model,* respectively. In the pitcher model, $O(t)$ is more or less directly controlled by $I(t)$, and such a system may be called a directly driven system. On the other hand, in the tippy bucket model the amount of water in the bucket and the spring constant (equivalent to some magnetospheric threshold parameters) play important roles in controlling *O(t)*, and such a system may be called an unloading system.

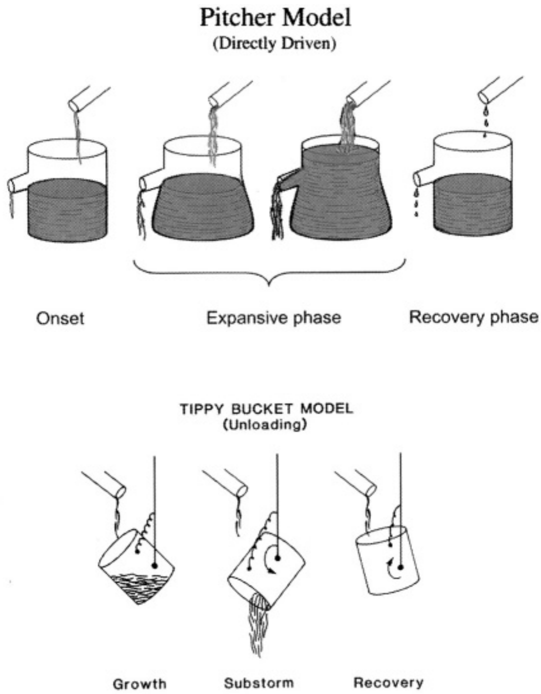


Figure 1.13a, b The pitcher (directly driven) model and the tippy bucket (unloading model): S.-I. Akasofu (1985).

If explosive magnetic reconnection is considered to be the primary process in generating substorm energy, we would expect that *O(t)* will be significantly different from *I*(*t*). This is because substorm energy would have to be accumulated in the magnetotail prior to its explosive conversion. There should be a delay, a time during which the energy is being accumulated - namely, the period between identifiable increases of *I*(*t*) and *O*(*t*); after substorm onset, *O(t)* should increase sharply, regardless of how *I*(*t*) varies, and *O*(*t*), at the peak time, should be much greater than *I*(*t*) at any time. Conversely, if *O(t)* is found to be very similar to *I*(*t*), there is little basis for hypothesizing explosive magnetic reconnection. It is for this reason that the relationship between *I*(*t*) and *O(t)* provides important information on the basic magnetospheric substorm process. In early energy transfer studies investigators attempted to determine the correlation coefficient between a geomagnetic index (chosen from Kp, ΣKp, AE, Dst, etc.) and solar wind quantities (such as the solar wind speed V, the mass density mn, the southward component of the solar wind magnetic field $-Bz$, etc.). Among such correlation studies, the auroral electrojet index AE (which is a substorm index) is found to be highly correlated (the correlation coefficient being 0.7-0.8) with $-Bz$ or $V^2|-Bz|$ The high correlation coefficients have suggested that $I(t)$ and $O(t)$ are closely related. Unfortunately, however, neither $-Bz$ nor $V^2|-Bz|$ is $I(t)$; likewise, AE is not $O(t)$. It is not possible to compare apples and oranges.

When I was attempting to identify the magnetosphere as a pitcher-type or a tippy-bucket-type system, I thought about the possibility of a system that is an intermediate between the two. One lesson I learned in this study was that a natural system is always complex and likely is neither of the first two extreme cases. Thus, it was best to propose an intermediate case, instead of one of the two. If I had chosen one of them and had been wrong, I would have been criticized or ignored. Thus, I tried to consider an intermediate type, as illustrated in Figure 1.13c. The three cases were presented during a substorm conference in Los Alamos in 1978 and are illustrated a little more quantitatively in Figure 1.13d.

In order to examine the relationship between $I(t)$ and $O(t)$, Paul Perreault and I estimated the total output $U_T(t) = O(t)$, in units of power (ergs/sec), on the basis of the two geomagnetic indices AE(t) and Dst(t) for a large number of geomagnetic storms, and then tried to find a combination of solar wind parameters that has the dimension of power and that resembles the output function in terms of time variations (Perreault and Akasofu 1978). The first input function we examined was the kinetic energy flux $((1/2)mnV^2)$. However, we found that there is no obvious relationship between this quantity and the output function. Actually, it was obvious even from the early study by Chapman and myself that an enhanced solar wind flow is not a necessary condition for the development of geomagnetic storms (Figure 1.14).

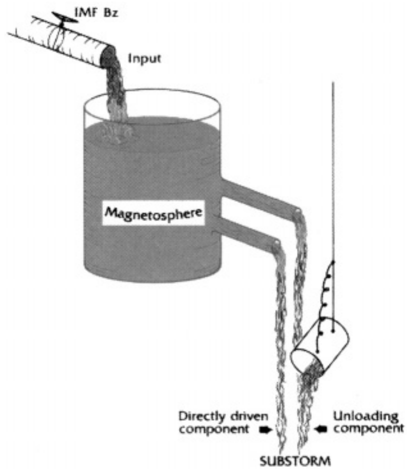


Figure 1.13c Model combining both the pitcher and tippy bucket models: S.-I. Akasofu (1985).

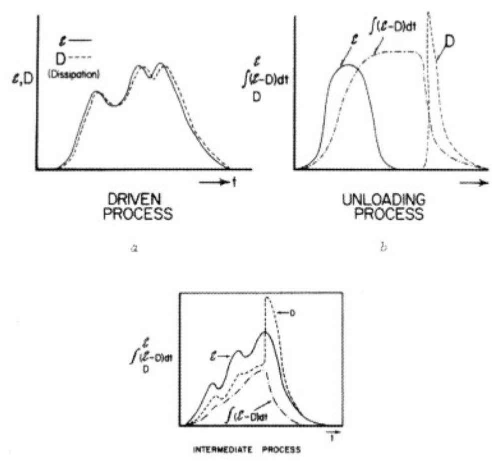


Figure 1.13d Semiquantitative representations of Figures b and c: S.-I. Akasofu (1980).

By then, it had been gradually confirmed by Roger Arnoldy (1971), Ching Meng et al. (1973), and many others that each substorm is associated with a specific change of the IMF *B*z component, namely from a positive value to a negative value, as Fairfield observed first. Thus, obviously, the next simple combination of solar wind parameters that has the dimension of power (erg/sec) and that considers the *B*z effect has the form of:

$$\varepsilon = VB^2 \sin^4(\theta/2)\ell_0^{\,2}$$

where V, B, and θ denote the solar wind speed, the interplanetary magnetic field magnitude, and its polar angle; ℓ_0 is a constant $\simeq$ 7 Earth radii.

In this regard, an important development was that Mikhail Pudovkin and his colleague (1986) identified ε as the Poynting flux across the magnetopause. This is a theoretical confirmation that the ε function can be identified as the power generated by the solar wind-magnetosphere dynamo. More specifically, the magnetopause is where the solar wind-magnetosphere dynamo is located.

By considering the range of variability of V, B, and θ in ε, θ is most crucial, then B, while effects of V are very small. We were surprised at how well the ε function reproduces the output function U_T. In Figure 1.14, we estimated the total energy dissipation rate U_T (the total output rate) from the AE and Dst indices (namely, pure magnetospheric quantities) and compared it with the kinetic energy flux (K) and ε (namely, pure solar wind quantities). One can easily recognize a close relationship between U_T (a magnetospheric quantify) and ε (a solar wind quantity), but not between U_T and K.

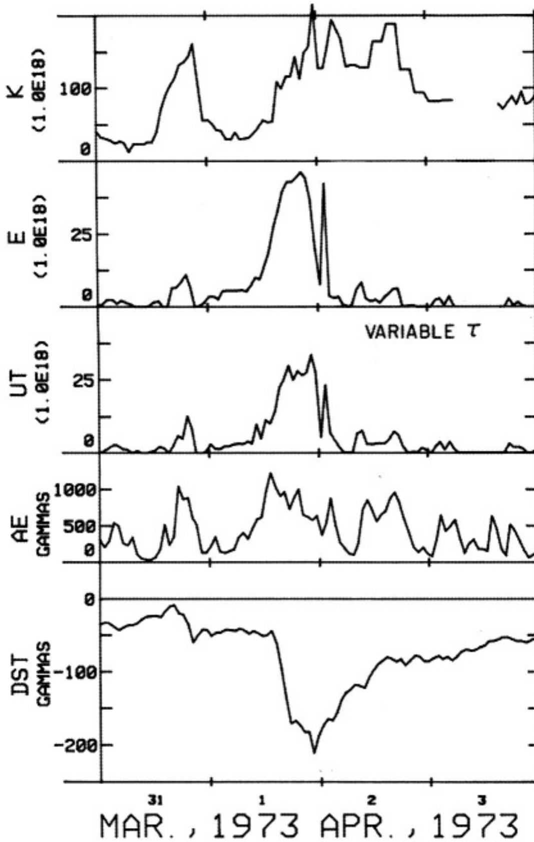


Figure 1.14 The kinetic energy flux (K), the ε function, the total energy dissipation rate (U_T) and the geomagnetic storm indices AE and Dst for the geomagnetic storm of March 31 – April 3, 1973: S.-I. Akasofu (1981).

One of the most important conclusions derived from this study is that the magnetospheric substorm is the *element* of global magnetospheric disturbance. It is the response of the magnetosphere to a significant increase of the solar wind-magnetospheric coupling for a few hours or more. Most of what we call *auroral phenomena* are visible manifestations of electromagnetic energy dissipation processes of this particular global disturbance. Therefore, the solar wind-magnetosphere coupling must constitute a dynamo that can supply the power for the dissipation process. Indeed, the ε parameter represents the power of this dynamo process for magnetospheric substorms. A typical magnetospheric substorm occurs when ε exceeds about 10^{18} erg/sec for a few hours. Soon after the publication of our results, Pat Reiff, et al. (1981) found the polar cusp potential is proportional to $\varepsilon^{1/2}$. This potential is approximately the voltage developed by the magnetospheric dynamo process; it is about 30-200 kilovolts (KV).

It so happened that the first libration point satellite, S3, was launched at that time; at the point of about 200 Earth radii distance, the gravitational pull of the Earth and that of the Sun are supposed to balance. It takes a little less than 60 minutes for the solar wind to reach from that point to the Earth. I was told that I could get the solar wind data from S3 on a real-time basis, free of charge, so long as my request was limited to two digits. Since ε in units of erg/sec is about 10^{18}-10^{20} erg/sec., I asked the S3 operations people to give me two digits, 2 and 8, if $\varepsilon = 2\mathrm{x}10^{18}$ erg/sec and 5 and 9 if $\varepsilon = 5 \mathrm{~x~} 10^{19}$ erg/sec, and so on. This scheme worked well. Since we could receive the data every five minutes on a real-time basis, my graduate students and I could wait for the aurora on the roof of the Geophysical Institute building, when ε went up above 10^{18} erg/sec, assuring the occurrence of a substorm a little more than 30 minutes or so later.

In summarizing this section, I think it is important to note that the magnetosphere should be considered a system that converts the kinetic component of the solar wind energy into electromagnetic energy, since geomagnetic and auroral phenomena are various manifestations of electromagnetic energy dissipation processes. The magnetosphere must thus be a *generator* for this conversion. It transforms the kinetic (input) energy of the solar wind into substorm energy and eventually into heat (output) energy in the ionosphere. The southward component of the IMF facilitates this energy transfer process.

During the course of studying the development of geomagnetic storms, I realized that a geomagnetic storm occurs when intense substorms occur frequently. This is clearly seen in Figure 1.8a. Chapman and I concluded that this relationship suggested that substorms are *essential elements* of a geomagnetic storm. In the early days, substorms were considered to be unrelated to a geomagnetic storm. In fact, in *Geomagnetism* by Chapman and Bartels, substorms were treated as *magnetic bays* in Chapter 10; substorms are observed as bay-like figures in midlatitude magnetic records. It should be noted that the concept of substorms is different from that of Birkeland's polar elementary storms; see also Section 2.5 on the same subject.

On the basis of my observation of the storm-substorm relationship, I concluded that substorms are the cause of the ring current belt, injecting high-energy protons from the magnetotail into that belt. Carl McIlwain and his colleagues (1974) showed that both protons and electrons are injected into the ring current belt and drift around the Earth. Meanwhile, there was great surprise that oxygen ions (O^+) become the dominant ions in the ring current belt during an intense geomagnetic storm. Since the oxygen ions in the solar wind are highly ionized (O^{+7}), O^+ ions must be of ionospheric origin. Indeed, a recent observation shows that O^+ ions are ejected out from the ionosphere into the magnetotail at substorm onset. After reaching the magnetotail, these ions are injected into the ring current belt by a convective motion of plasma in the magnetotail (see also Section 3.8).

1.10 The Directly Driven and Unloading Components

The component of the output function that closely follows the ε function in time is now called the directly driven component, which is illustrated in Figure 1.13a. The rest is the unloading component, as illustrated in Figure 1.13b. The existence of the directly driven component had not been considered for many years, since the spontaneous reconnection paradigm was so powerful at that time. The directly driven component was *officially* recognized for the first time as late as 1987, in a joint paper by Gordon Rostoker and his colleagues (1987). That the magnetosphere must be driven first for substorms to occur and that substorms are not caused by a spontaneous process finally became clear.

The directly driven component is the one in which the energy derived from the solar wind is directly deposited in the magnetotail, the ionosphere, and elsewhere with a slight time delay. Thus, time variations of this component have approximately the same time variations as that of ε(t). For this component, the equivalent current pattern in the polar ionosphere features two vortices (Section 3.1.2).

The unloading component must be caused by a magnetosphere-ionosphere (MI) coupling process that is presently unknown. The equivalent current pattern in the polar region associated with the unloading component has a single vortex involving a longitudinally confined westward electrojet centered around the midnight sector. Its time variations do not resemble the rate of energy derived from the solar wind. Magnetic reconnection and various instability processes beyond ten Earth radii in the magnetotail have been proposed as the cause for the unloading component, but recent satellite observations have not revealed the expected (***E***x***B***) earthward plasma flow. Simply put, the magnetotail (tail) cannot wag the ionosphere (dog). Many fascinating phenomena occur in the magnetotail, but we should not forget the very significant ionosphere.

In theorizing about the causes of the unloading component, it is crucial to know its characteristics, at least its time variations. It is rather surprising that proponents of magnetic reconnection have been theorizing substorm processes while ignoring characteristics of the time variations of the unloading component. I have learned that theorists tend to formulate their own problem in their own way and that they try to learn about only what they are interested in. Observations are forgotten. In this particular case, they formulate a spontaneous and explosive reconnection problem, but are not concerned with the associated time variations. They examine how it can possibly be stopped after it begins. If magnetic reconnection is so fundamental, each substorm should last until the entire tail is burned up.

However, we know that the magnetosphere has both components, as illustrated in Figures 1.13c and 1.13d. Now, the question is whether there is any method for separating the two components, so that we can learn about time variations of the unloading component. For this purpose, Wei Sun and his colleagues (1998) applied the Method of Natural Orthogonal Components (MNOC) to this difficult problem.

From this analysis, they found that the first natural component has a two-cell pattern, which is well known to be associated with the convection in the magnetosphere. It is enhanced during the growth and expansion phases of substorms and decays during the recovery phase of substorms. Further, it has a fair correlation with the ε function with a time lag of 20-25 min. Thus, this may be identified as the directly driven component (Figure 1.15).

The second natural component reveals itself as an impulsive enhancement of the westward electrojet, around midnight, between 65° and 70° latitude, during the expansion phase only. It is much less correlated with the ε parameter than is the first one. Thus, as a first approximation, we may identify it as the unloading component. Sun et al.'s analysis showed that the directly driven component tends to dominate over the unloading component, except for a brief period soon after substorm onset. This is the first clear determination of the time profile of the unloading component. Thus, knowing its characteristics and its time profile, it has now become possible to examine, for the first time, the physics of the unloading component and its cause. This problem will be further discussed in subsequent chapters.

1.11 The Open Magnetosphere

It appears that magnetospheric physicists did not consider seriously the open model until the beginning of the 1970s.

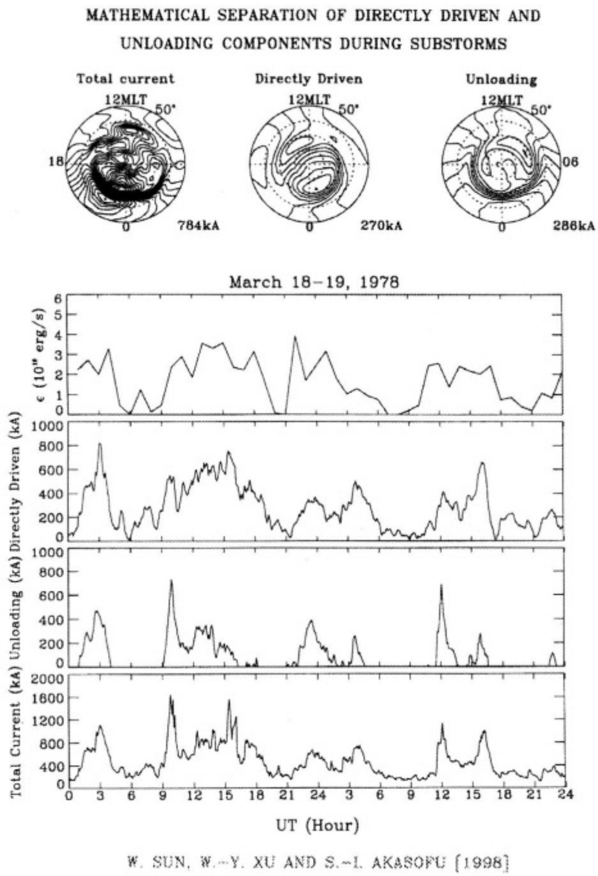


Figure 1.15 Separation of the directly driven component and the unloading component for the period of March 18-19, 1978: W. Sun, W.-Y. Xu, and S.-I. Akasofu (1998).

They began to pay attention to the concept of an open magnetosphere when A. Vampola (1971) detected solar electrons of about 400 KeV uniformly over the entire polar region (Figure 1.16). According to Störmer's cut-off latitude calculation, these electrons could reach only very near the geomagnetic pole. There is no way to explain this phenomenon without considering the magnetosphere to be open. The only possible interpretation of this phenomenon is that these electrons reach the polar cap along the magnetospheric magnetic field lines, which are connected with the interplanetary magnetic field lines; these field lines are in turn connected to the Sun. Here, the polar cap is defined as the area where the open field lines originate. Further, the dayside and nightside boundaries of the area where the electrons had been detected coincided with those of the auroral oval (Chapter 2), indicating that the auroral oval delineates approximately the boundary of the polar cap.

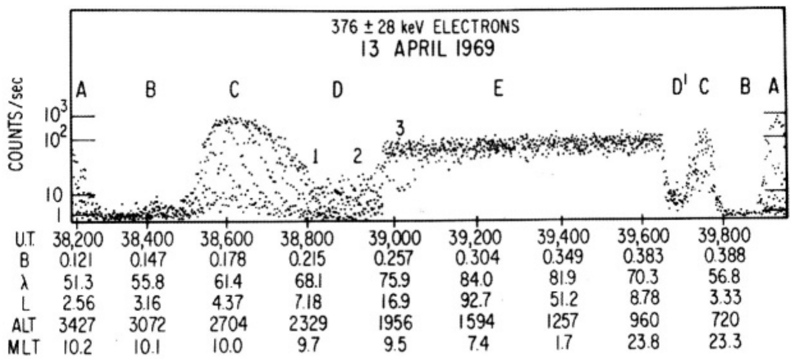


Figure 1.16 Observation of solar electrons as a satellite crossed the polar region from the late morning sector (9.5 MLT) to the late evening sector (23.8 MLT). Across the polar cap, the electron flux was uniform: A. Vampola (1971).

I recall that during my visit to the University of Iowa in the early 1960s, my colleagues and I found a very strange phenomenon. Protons of energies well below Störmer's cut-off energies were sometimes observed deep in the so-called *forbidden region* (Figure 1.17). However, we had no idea how to explain this phenomenon, since the Chapman-Ferraro theory predicts that the equatorial boundary of the forbidden region is even higher when the Earth's dipole field is confined by the solar wind. Now it may well be that this anomalous phenomenon is related to the fact that the magnetosphere is open.

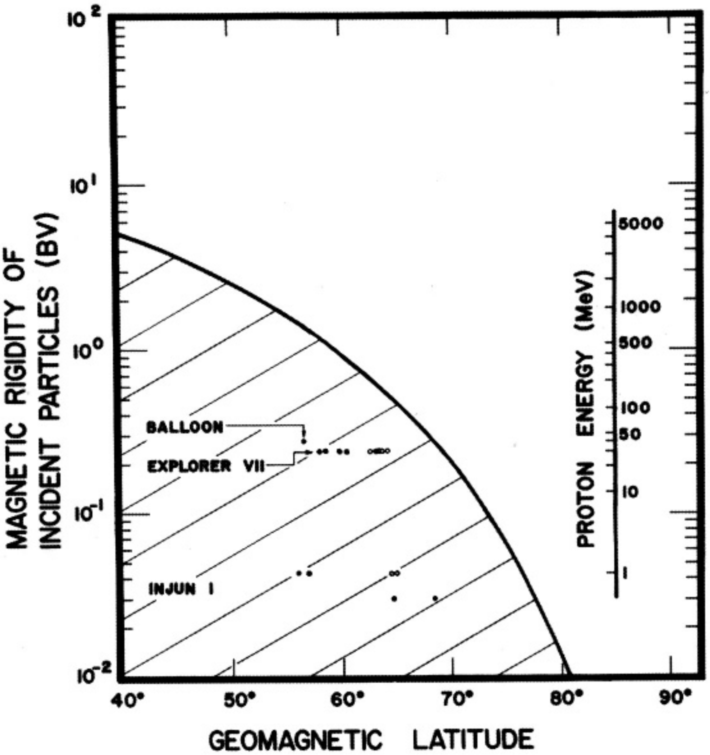
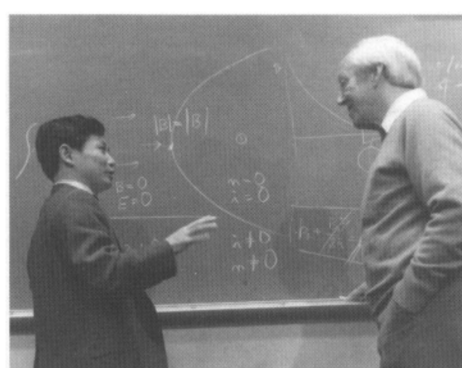
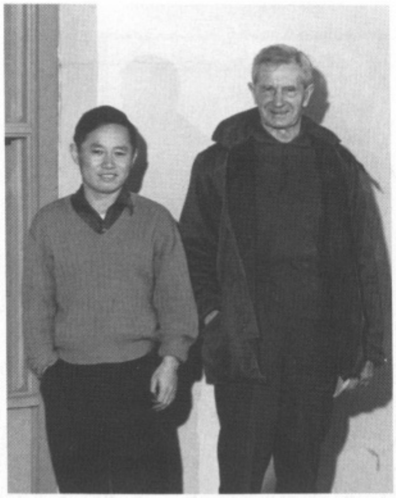


Figure 1.17 Störmer's *forbidden region* in magnetic latitude-magnetic rigidity diagram. Solar protons were observed deep in the forbidden region: S.-I. Akasofu, W.C. Lin, and J.A. Van Allen (1963).

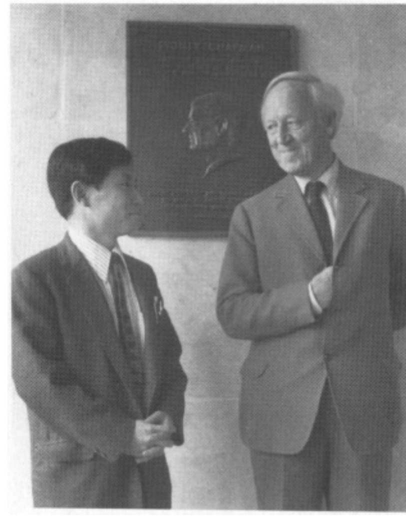
Chapman and Akasofu: Geophysical Institute, University of Alaska (1961).



Alfvén and Akasofu: Geophysical Institute, University of Alaska (1974).



Chapman and Akasofu: Geophysical Institute, University of Alaska (1964).



Alfvén and Akasofu in front of Chapman's bust at the Chapman Building of the University of Alaska Fairbanks: Geophysical Institute, University of Alaska (1974).

Chapter 2

CONFRONTING PARADIGMS: AURORA RESEARCH DURING THE EARLY SPACE AGE

2.1 My Earliest Association with the Aurora

My mother had a favorite song. It was a sentimental popular song, which she used to sing ever since she was a young girl. Its title is something like "A Drifter's Song" and starts with "I have to decide to go ahead or return home under the aurora. Russia is a big country, sunset in the western part, sunrise in the eastern part, a noon bell in the middle..." My mother would sing this song to me, when I was five years old. The only word I did not understand in the song was *aurora*, and I asked my mother about it. If I remember correctly, she told me that it was something she hadn't seen, but is a beautiful phenomenon in a far northern country. This was my first encounter with the word *aurora.*

I was born in a small town in the mountainous region of central Japan, only ten miles from Mt. Asama, one of the most active volcanoes in Japan. One of my earliest childhood memories is a gigantic nighttime eruption, which I observed from my mother's back, crying in fear. The elementary school I attended was small, but was very well equipped with scientific instruments. I recall I was, and still am, fascinated by lights from vacuum discharge tubes, which are closely related to the aurora, although I was obviously not capable then of associating the lights with the aurora.

The Department of Geophysics of Tohoku University, which I attended, was staffed with famous professors. Among them were Yoshio Kato (geomagnetism), Gi-Ichi Yamamoto (atmospheric sciences), Kokichi Honda (seismology), Hiroshi Kamiyama (ionospheric physics), and others. The department operated a magnetic observatory where I worked from time to time to earn wages. There, several magnetometers recorded magnetic changes. In the magnetometers, a light beam was deflected from a mirror attached to a magnet and produced a spot on a photographic paper wrapped on a rotating cylinder in a dark room. I was greatly attracted by movements of the light spot and learned that the movements were caused by the aurora, an electrical discharge phenomenon, in Siberia and Alaska. It was fascinating to imagine how a distant phenomenon like the aurora could cause delicate movements of the spot. It was during this time, my student days, when I associated the memory of my mother's song with what I was learning.

However, as mentioned in Chapter 1, it was Chapman-Ferraro's paper that brought me to Alaska.

2.2 The Auroral Zone to the Auroral Oval

E. Loomis (1860) assembled the first extensive collection of auroral appearances over the Earth and found that the aurora tends to appear along a fairly narrow belt centered around a point at the northwestern tip of Greenland, not at the geographic pole (Figures 2.1 and 2.2). H. Fritz (1873), using much more data covering the period from 503 B.C. to A.D. 1872, confirmed Loomis' findings and constructed his well-known map of isochasms, the lines of equal average annual frequency of auroral visibility expressed by ‘M’ nights per year. The maximum frequency of auroral visibility thus defined was found to lie approximately along Loomis' belt. This auroral belt has been called the auroral *zone.* The centerline of the aurora zone coincides well with a geomagnetic latitude (gm lat.) of 67°. The width of the auroral zone is about 5° - 6° of latitude. Thus, on a geomagnetic longitude-latitude map centered around the geomagnetic pole (located near the northwestern tip of Greenland), the auroral zone is a circumpolar belt (Figure 2.3). Harry Vestine (1944) refined Fritz's isochasm map with the aid of additional data covering more than a century, including the two International Polar Years.

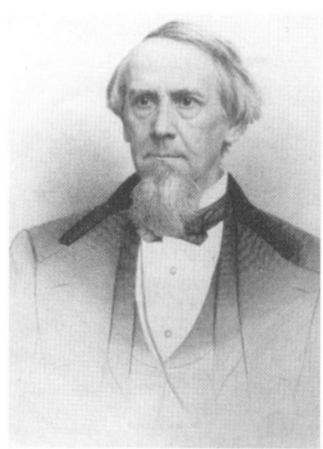


Figure 2.1 E. Loomis (1811-1889)

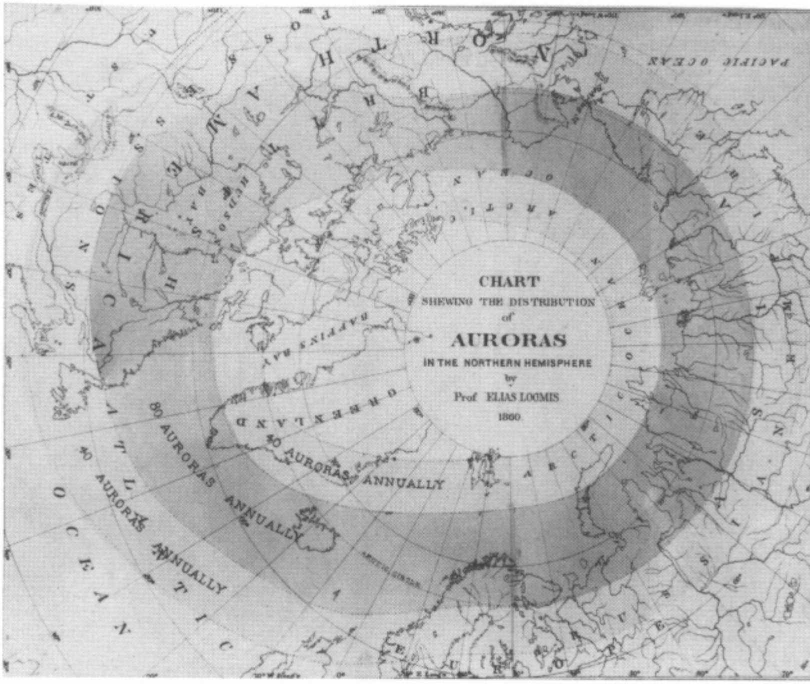


Figure 2.2 The auroral zone determined by Loomis: E. Loomis (1860).

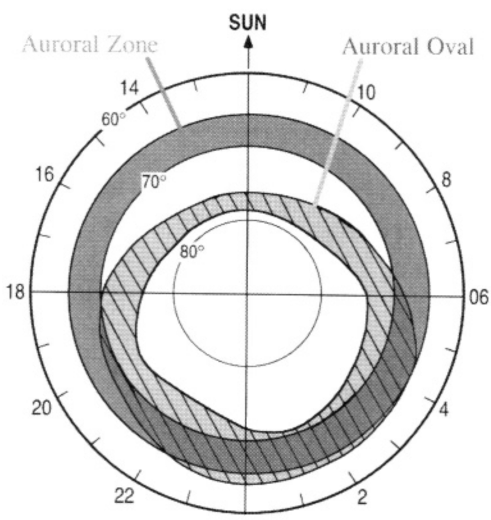


Figure 2.3 The auroral zone (shaded), the auroral oval (line shaded), and the outer boundary of the outer radiation belt (projected to the ionosphere), all on the geomagnetic latitude-magnetic local time coordinate system: S.-I. Akasofu (1967).

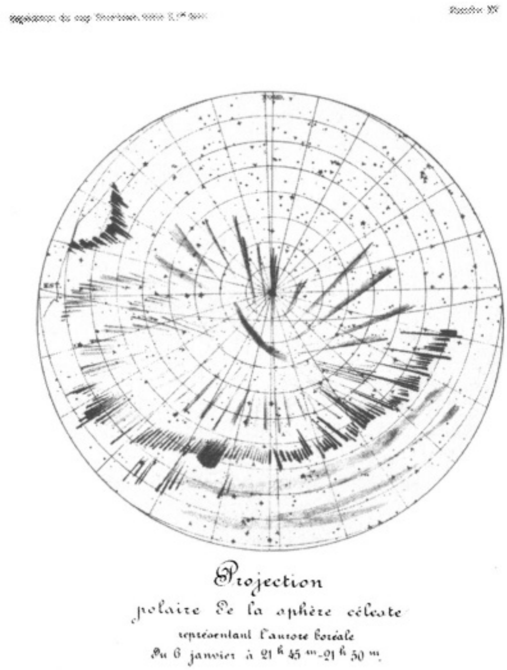


Figure 2.4 Sketch of the aurora by N. Carlheim-Gyllensköld at Svalbard (1907).

Figure 2.5 Photography was introduced in auroral science at the beginning of the twentieth century. C. Störmer (sitting) used it extensively in determining the height of the aurora: University of Tromsø.

Figure 2.4 shows an auroral sketch made by N. Carlheim-Gyllensköld at Cape Thordsen in Svalbard during the First Polar Year (1907). This was one of the first scientific recordings of the aurora. A photographic method was introduced in auroral physics at the beginning of the twentieth century (Figure 2.5). A number of auroral expedition parties were dispatched to Greenland, Siberia, Canada, and many other countries during the Second Polar Year (1932). The isochasm map was further refined by Yasha Feldstein and his colleagues (1961) and Bengt Hultqvist (1961), based on International Geophysical Year (IGY) data.

Based on such studies, it had been tacitly believed for more than 100 years that the auroral zone was the actual belt along which the aurora lies. It was Sydney Chapman, president of the IGY, and Chris Elvey, director of the Geophysical Institute, University of Alaska, who thought that the actual belt of the aurora should be determined photographically, not by statistics as done by Loomis, Fritz, and Vestine. For this purpose, they took the leadership in constructing all-sky cameras (Figures 2.6a and 2.6b).

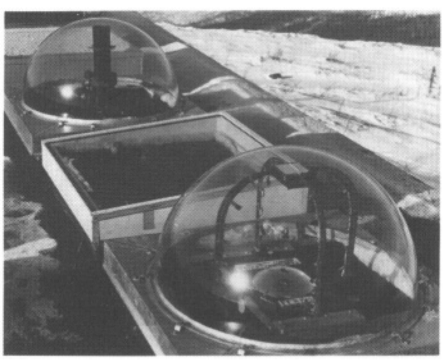


Figure 2.6a All-sky camera and a photometer: Geophysical Institute, University of Alaska.

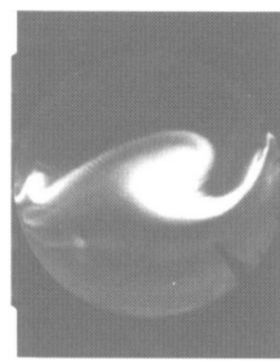


Figure 2.6b An example of a photograph taken by an all-sky camera: Geophysical Institute, University of Alaska.

Auroral researchers in several countries responded to Chapman and Elvey by designing and constructing their own all-sky cameras. During the IGY, such cameras were operated at more than 100 locations and took photographs of the sky at one-minute intervals, regardless of sky conditions. The films were then sent to the World Data Center in Moscow and the Geophysical Institute, University of Alaska.

When I became a graduate student of the Geophysical Institute in December 1958, I had an opportunity to observe the aurora with my colleagues, including Gene Wescott and Charles Deehr. I observed that the aurora tends to appear in the northern sky in the evening, advances toward the zenith (or even the southern sky) of Fairbanks (gm lat. 64.6°), and recedes toward the northern sky in the morning. This north-south shift of auroral arcs was a well-known fact by then (V.R. Fuller and E.H. Bramhall, 1937; Jim Heppner, 1954). I recall that I asked Elvey why this shift occurs, if auroral arcs were supposed to lie along the auroral zone. His response was that it was perhaps that auroral arcs tend to form at the centerline of the auroral zone (gm lat. 67°) and then the auroral arcs, after their formation, move equatorwards.

My question was simply that if the concept of the auroral zone was correct, we should be able to see auroral arcs near the zenith of the sky above Fairbanks at 6 p.m. when the sky becomes dark enough (actually, in Fairbanks, the sky is dark enough to observe the aurora even before 5 p.m. around the winter solstice). Instead, auroral arcs almost always appear near the northern horizon first and advance equatorward. My question to Elvey was the naïve one of a graduate student.

After this conversation with Elvey, I examined newly arrived IGY all-sky films taken at Fort Yukon, Alaska (gm lat. 66.6°), which is located at about the center line of the auroral zone. To my great surprise, auroral arcs behaved in a similar way at Fort Yukon as in Fairbanks. That is, auroral arcs appeared first near the northern horizon. Therefore, I also examined all-sky films from Barrow, Alaska (gm lat. 68.5°), well north of the centerline of the auroral zone. It was even more surprising to me that auroral arcs behaved in a similar way at Barrow. The only difference is that the local time of the first appearance is in the northern sky and that the arc arrives at the zenith earliest at Barrow, than at Ft. Yukon, than at Fairbanks. Figure 2.7a shows simultaneous all-sky camera photographs from Sachs Harbor (gm lat. 76.0°), Inuvik (gm lat. 71.0°), Fort Yukon, and College. The photographs show the equatorward shift of the aurora in the evening; see also Figure 2.8a.

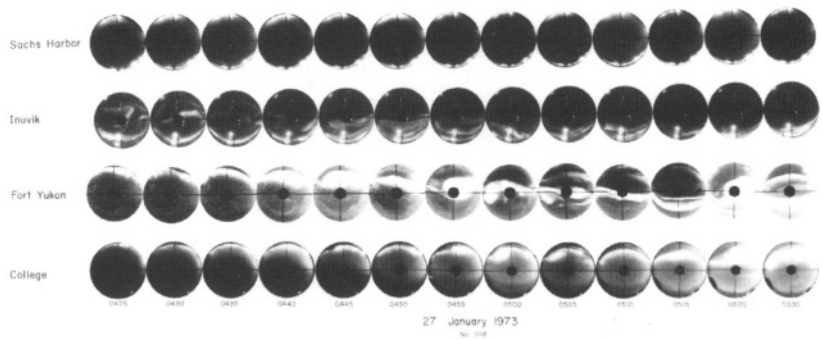


Figure 2.7a All-sky photographs from the Alaska meridian chain (Figures 2.8a and 2.8b), showing the equatorward shift of the aurora. A substorm began toward the end of this series: S.-I. Akasofu (1974).

It was quite obvious to me at that time that auroral arcs do not lie along the auroral zone. I realized, also, that Loomis, Fritz, and others did not and could not take into account the local time dependence of the auroral distribution (namely, only how many nights per year) in their statistical study, meaning that the instantaneous belt of auroral arcs can be quite different from the auroral zone.

All-sky films from many IGY arctic stations started to arrive at the Geophysical Institute in 1959 and 1960. It was my finding that the actual distribution of auroral arcs agrees with the auroral zone only during the midnight hours and deviates greatly from the auroral zone at the other local times. However, I could not determine the auroral distribution on the day-side of the Earth because of the lack of data at that time.

Yasha Feldstein (1963) determined the complete distribution of the aurora at all local times, using the films from Heiss Island and other sites that can observe midday auroras (Figure 2.7b). His distribution showed that the belt of the auroral zone is located at about 78° during midday hours, instead of 67° (Figure 2.3). Further, the center of the belt is shifted by about 3° from the geomagnetic pole toward the midnight sector. This belt is called the auroral *oval.* Since the results obtained by Feldstein were basically the same as mine for the dark hours, I supported his results immediately. On the other hand, Knud Lassen, in Copenhagen, proposed once that there were two belts of aurora instead of the auroral oval.

Figure 2.7b From the left, S.N. Vernov, O.V. Khorosheva, and Yasha Feldstein at the University of Moscow campus at the occasion of my first visit to Russia: S.-I. Akasofu (1968).

That time was a sort of golden age of auroral spectroscopy. All-sky cameras were not considered even to be a scientific instrument for auroral spectroscopists, compared with their then-sophisticated spectroscopic instruments. In fact, some of my senior colleagues advised me that the aurora should be the same in Alaska, Siberia, Canada, and Norway, that physics of the aurora should be the same everywhere, that the distribution of the aurora is thus not a major issue, and thus that it is a waste of time to work on it. I objected to this argument. Auroral arcs appear in a very specific belt, the auroral oval, and not along the auroral zone, and not all over the polar region. This fact tells us something about their cause and origin, therefore it is important to determine their actual distribution accurately.

In such an atmosphere, Feldstein's results got little attention from the scientific community. Worse, since the auroral zone had been believed to be the belt of auroral arcs for more than 100 years, it was difficult for both of us to convince our colleagues of the validity and significance of the auroral oval.

In order to convince the scientific community that Feldstein’s and my views about the auroral oval were valid, I planned several projects. The first was to establish the Alaska meridian chain of all-sky cameras (Figures 2.8a). Taking advantage of the Earth's rotation, a meridian chain of all-sky cameras can *scan* the entire polar sky (like an azimuth-scanning radar at an airport) once a day, and delineate the auroral oval that is fixed with respect to the Sun (Figure 2.8b). As far as I am aware, this is the largest scanning device on Earth. This project was funded by my first grant from the National Science Foundation.

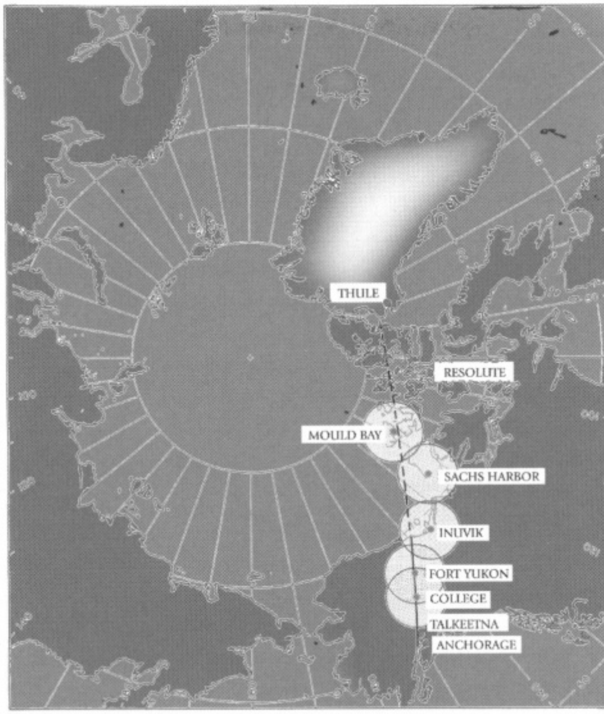


Figure 2.8a Alaska meridian all-sky cameras. A circle indicates the field of view of each camera: S.-I. Akasofu.

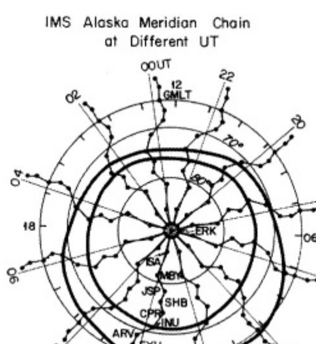


Figure 2.8b Position of the Alaska meridian cameras at different universal times (UT) on geomagnetic latitude-magnetic local time (MLT) coordinates: A.L. Snyder (1972).

Figure 2.8c shows an example of the results from this investigation. If auroral arcs were distributed along the auroral zone, they should appear in a horizontal belt approximately along the latitude of Fort Yukon (gm lat. 66.6°).

Instead, auroral arcs appear at about gm 76-77° at 0 UT (14 MLT, Magnetic Local Time) and shift toward the latitude of Fairbanks. The line-dot curve shows Feldstein's oval for the magnetic index Q = 3. Therefore, the meridian chain of all-sky cameras could delineate the auroral oval. The width of the oval changes intermittently, a phenomenon that will be discussed later.

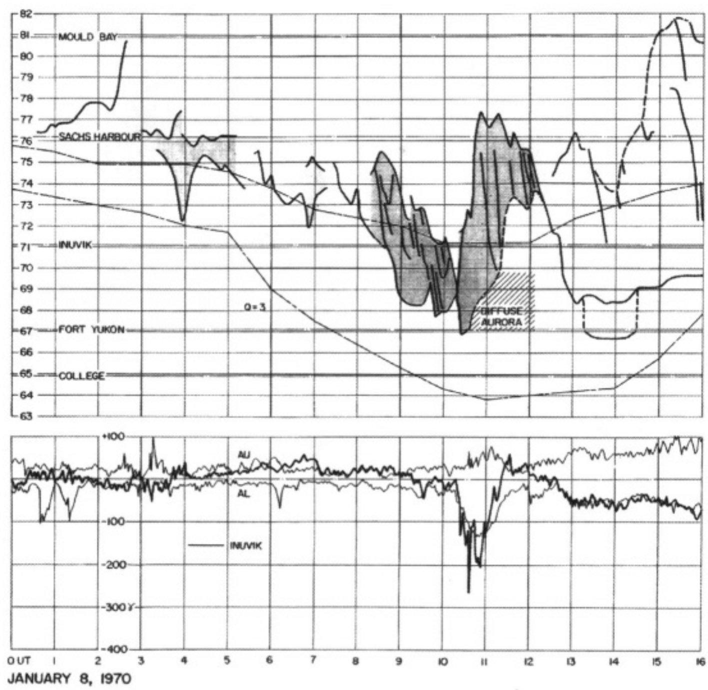


Figure 2.8c Location of the aurora observed by the Alaska meridian all-sky cameras: A.L. Snyder and S.-I. Akasofu (1972).

The second project was to fly along auroral arcs, since the flight path should be able to delineate the auroral oval. Both a US Air Force jet from Hanscom Air Force Base and a NASA jet from Ames Research Center participated in the operation. The results were as predicted: the flight paths delineated clearly the auroral oval (Figures 2.9a and 2.9b). George Gassmann, Jurgen Buchau, Charlie Pike, Rosemarie Wagner, and Jim Whalen of the Air Force Geophysics Laboratory, and Walter Heikkila and Dave Winningham of the University of Texas were instrumental in accomplishing this task. However, I felt that the scientific community in general was not much interested in such observational results at that time.

Figure 2.9a A U.S. Air Force aircraft that participated in auroral research: G. Gassmann.

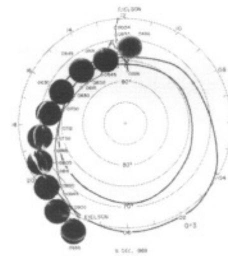


Figure 2.9b All-sky photographs taken from a U.S. Air Force jet that flew along the evening half of the auroral oval: J. Buchau, J.A. Whalen, and S.-I. Akasofu (1970).

2.3 Significance of the Auroral Oval

One lesson I learned in elucidating the auroral oval is that one specific finding alone would not get much attention from the scientific community. When one finds an interesting phenomenon, it is necessary to relate it to other significant phenomena and demonstrate that a new finding is worth paying attention to. Thus, the third attempt was to find other geophysical phenomena that have a distribution similar to the auroral oval. Fortunately, I had an opportunity to work with the space physics group of the University of Iowa. I found one day that Lou Frank, James Van Allen, and John Craven were plotting the outer boundary of the outer radiation belt onto the Earth's surface.

I was greatly surprised that the boundary they delineated coincides fairly well with the auroral oval (the solid curve in Figure 2.3). This result suggested to me that auroral electrons penetrate into the polar upper atmosphere by moving along the outer boundary of the outer radiation belt. I remember that I reported the results immediately to Van Allen. It was a time when the initial hope of associating auroral phenomena with the radiation belts had faded, so initially convincing my colleagues of this finding's significance was difficult. After the discovery of the plasma sheet, this result had long been forgotten, and only during the last few years are some researchers coming back to the boundary of the outer radiation belt in their search for the origin of auroral arcs.

Thus, it was fortunate that Al Zmuda and his colleagues (1966) found on the basis of TRIAD satellite data that field-aligned currents flow in or out from a belt that is basically identical to the auroral oval. He told me that he plotted the location of the field-aligned currents on my figure (Figure 2.3) after finding it in one of the University of Iowa reports. This fact suggested that auroral electrons carry field-aligned currents. It so happened that the tape recorder aboard the satellite failed, so Zmuda asked me to help install his satellite receiving station at the top of the Geophysical Institute building; it was installed when the temperature was 50 degrees below zero Fahrenheit. Using his satellite data and the simultaneous all-sky data, we found that auroral arcs appear where there is upward field-aligned current (Figure 2.10).

Takeshi Iijima and Tom Potemra (1976) completed Zmuda's work by showing the distribution of field-aligned currents at the ionospheric level (Figure 2.11). Further, solar protons of energies on the order of 1.5 Mev were found to penetrate uniformly over the polar region bounded by the aurora oval. Energetic solar electrons were also found in the area bounded by the auroral oval (Figure 1.16). These results indicate that the auroral oval delineates approximately the boundary of the polar cap. The field lines that originate at the polar cap are connected with the interplanetary magnetic field lines, so that they are "open" field lines. It was in this way that the significance of the auroral oval was firmly established. I learned thus that it is very important to find as many relevant results as possible in proving the importance of a newly observed result.

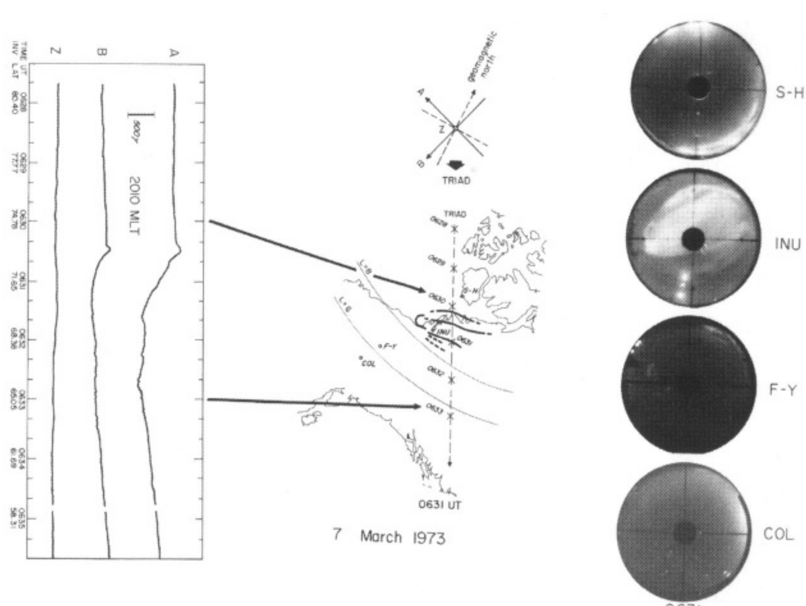


Figure 2.10 Simultaneous observations of the aurora (all-sky photograph) and the field-aligned currents (TRIAD satellite): Y. Kamide and S.-I. Akasofu (1976).

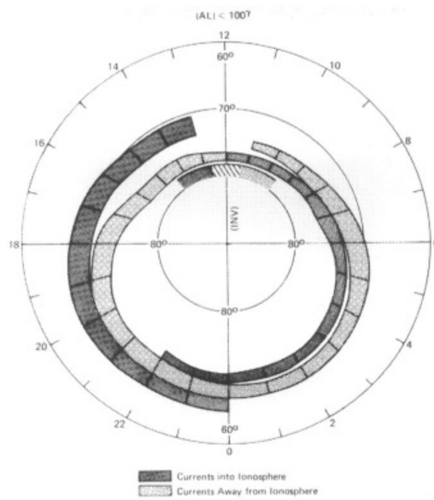


Figure 2.11 Distribution of the field-aligned currents: T. Iijima and T.A. Potemra (1976).

The validity and significance of the auroral oval began to be recognized toward the beginning of the 1970s. However, we had to wait for full recognition of the auroral oval until 1971, when a scanning instrument devised by Cliff Anger, and installed on the ISIS-2 satellite imaged the entire oval (Figures 2.12a and 2.12b). Tony Lui came to Alaska as a postdoctoral fellow, starting joint projects on ISIS-2 data with the University of Calgary group; their work extended Anger's observation. After this, the concept of the auroral oval was accepted as if there had been no controversy about it in the past. In any modern monograph on the aurora, one can find a simple statement that auroral arcs lie along the auroral oval. Thus, it is interesting to recognize that such a simple fact – perhaps one sentence in modern textbooks – had a long history; it took about a decade of struggle for its acceptance by the scientific community.

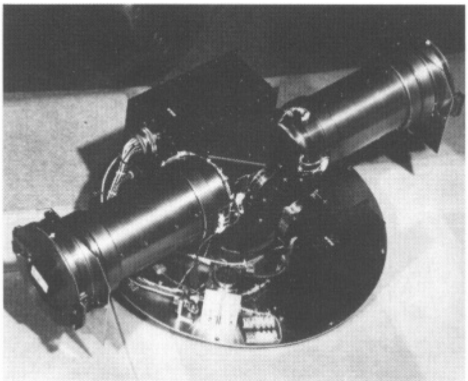


Figure 2.12a Cliff Anger's scanning instrument aboard the ISIS-2 Satellite: C. D. Anger.

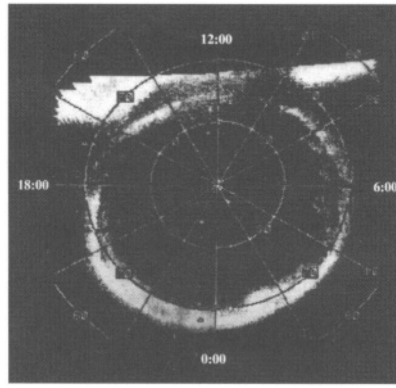
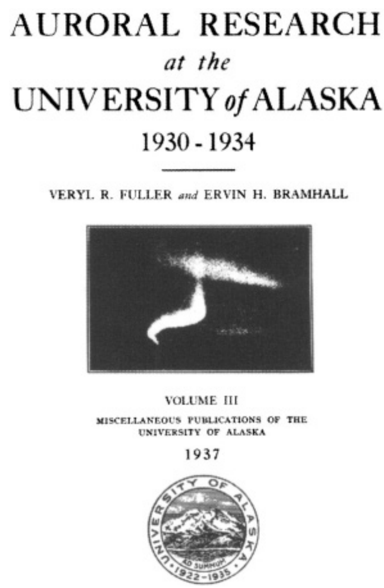


Figure 2.12b First image of the auroral oval depicted by Cliff Anger's instrument: C. D. Anger.

For me, it started out with a naïve question about daily auroral behavior that was well known at the time. Recalling those days, I appreciate the foresight and courage of both Chapman and Elvey for taking the leadership of the all-sky camera project in spite of the fact that auroral spectroscopists and auroral physicists in general paid little attention to it. I should note that, as far as I know, Neil Davis, Carl Gartlein, Alexander Lebedinsky, and W. Stoffregen were the first who used an all-sky camera for aurora studies.

2.4 Auroral Substorms: Fixed Pattern to Substorm Pattern

It had long been believed, on the basis of the pioneering study of the aurora by V.R. Fuller and E.H. Bramhall (1937) and also by Jim Heppner (1954), that in the evening sky, auroral arcs *always* had a quiet and homogeneous form, that auroral arcs were *always* very active in midnight hours and that arcs were broken up into patchy forms in the morning sky (Figure 2.13). In this view, the auroral activity pattern is fixed with respect to the Sun (and thus to magnetic local time), and the Earth, with observers at points on it, rotates under such a fixed pattern of activity once a day. That is, a single observer at a point observes a quiet form, an active form, and a patchy form in the evening, midnight, and morning skies, respectively, during the course of the night. This is certainly true statistically.


AURORAL RESEARCH
at the
UNIVERSITY of ALASKA
1930 - 1934

VERYL R. FULLER and ERVIN H. BRAMHALL

VOLUME III
MISCELLANEOUS PUBLICATIONS OF THE
UNIVERSITY OF ALASKA
1937

Figure 2.13 Cover page of the report by V.R. Fuller and E.H. Bramhall.

At the beginning of the IGY (1957-1958), however, little was known about how auroras behave *simultaneously* in Siberia (in evening hours) and Canada (in morning hours), when auroras became suddenly active over the Alaska sky (in midnight hours). There had not been any simultaneous observations of auroras over a long local time span up to then. At the Rasmuson Library of the University of Alaska Fairbanks, I found a letter from C. Störmer to Fuller and Bramhall, urging them to conduct, jointly, simultaneous observations of the aurora in Norway and Alaska. It must have been Störmer's dream to make such a joint observation.

As I began to examine IGY films, I found that the view commonly held on the auroral activity pattern was incorrect. Indeed, all-sky films even at a single station on the same night showed that auroral arcs can transform themselves from quiet to active and back to a quiet form two or three times (Figure 2.14). This fact suggested to me either that the fixed pattern concept was not correct or that the Earth rotated two or three times in a single night! As a graduate student I was obviously puzzled, but was overwhelmed at that time by the firm believers in the fixed pattern. Thus, I decided to examine *simultaneous* all-sky photographs from Siberia, Alaska and Canada, when Alaska was in the midnight sector. It was my finding that when an auroral arc is quiet in Alaska (in the midnight sky), it is also quiet over Siberia (in the evening sky), and Canada (in the morning sky), in addition to the fact that the aurora over Alaska can be quiet even in the midnight hours. When an auroral arc suddenly brightens and moves rapidly poleward over the Alaska sky (Figure 2.15 and 2.16a), this activity generates a large wavy or folding structure (the westward traveling surge), which propagates along the arc toward Siberia (toward the evening sky, Figure 2.16b). This surge-like activity was recorded first at the Siberia station closest to Alaska several minutes after its formation over Alaska and, subsequently, at other earlier evening stations in Siberia. This activity could propagate all the way to the day side of the oval with a speed of a few kilometers per second. At the same time, auroras over Canada became active, often forming an inverted Ω-shaped form (called the omega band). To the south of the omega band, auroral arcs became folded in a very complicated way. Folded portions of an arc appear as shafts of light, or patchy forms, scattering all over the sky.

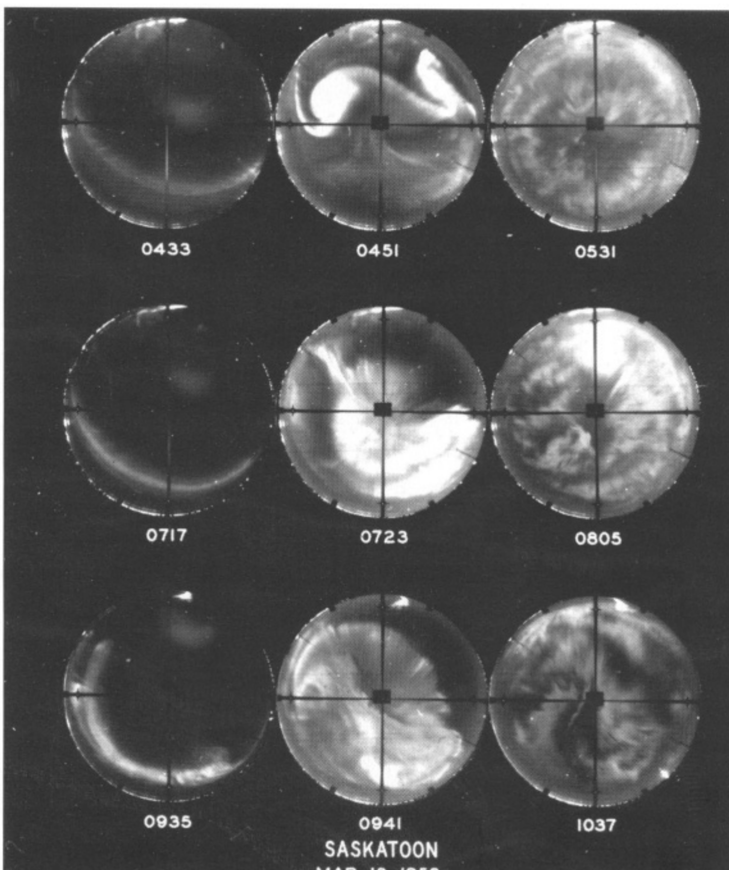


Figure 2.14 All-sky photographs taken during the night of March 13, 1958, at Saskatoon, Saskatchewan, Canada. Note that the aurora underwent three cycles of its activity during a single night: S.-I. Akasofu.

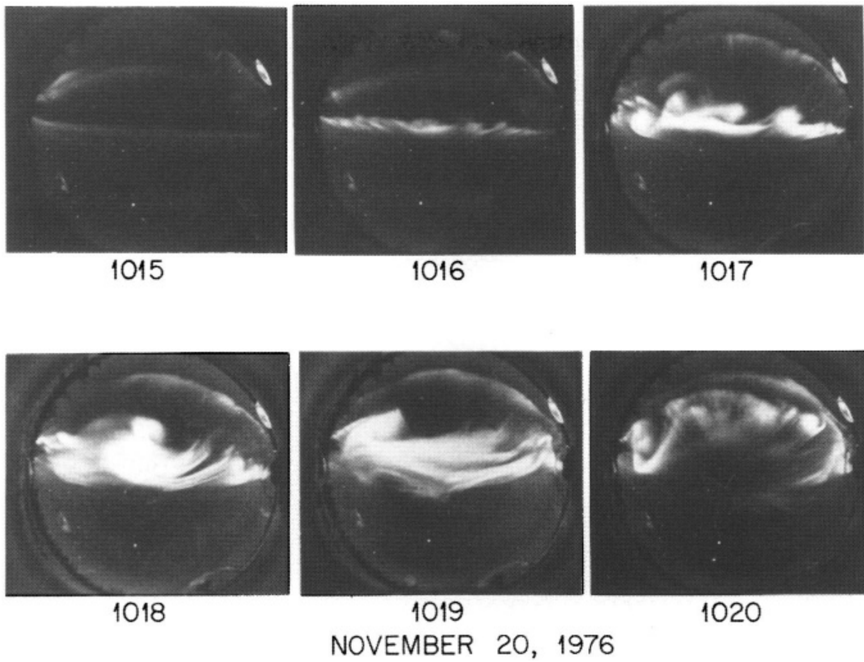


Figure 2.15 All-sky photographs showing a sudden brightening of an auroral arc at substorm onset: S.-I. Akasofu.

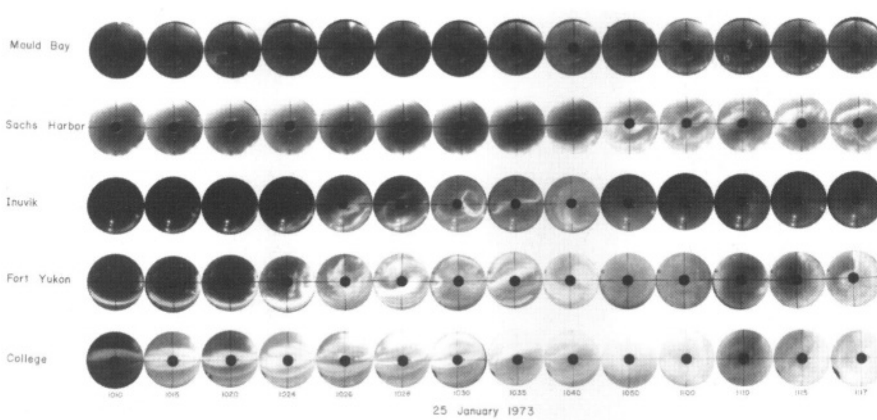


Figure 2.16a Intense poleward expansion of the auroral activity depicted by the Alaska meridian all-sky camera from the zenith of College (Fairbanks) to the northern sky of Sachs Harbor (see Figure 2.8a): S.-I. Akasofu (1974).

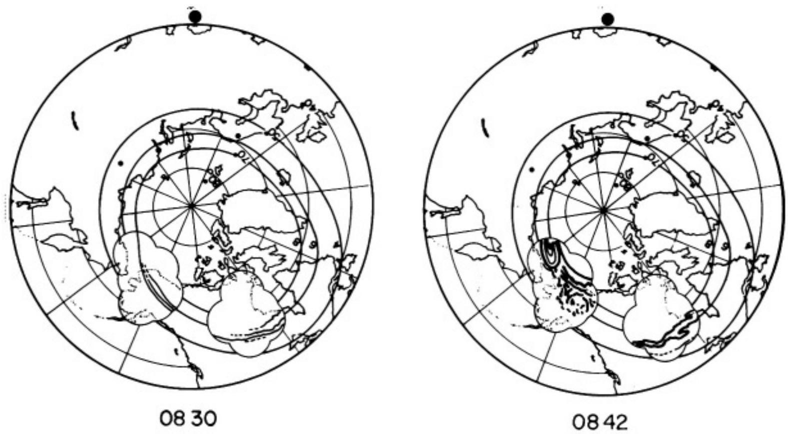


Figure 2.16b Simultaneous auroral activity observed in Alaska, Siberia, and Canada: S.-I. Akasofu (1963).

More important, when auroras over Alaska in the midnight sector became quiet again, in about two to three hours after an active period, auroras over Siberia and Canada also became quiet. Further, such activity often repeated two or three times during an active night. Chapman coined the term auroral substorm for this transient phenomenon. My paper on this subject initially had a title of *Auroral Activation.* Chapman refused to review my paper unless I changed the title to *Auroral Substorm* (Akasofu 1964). There was little mention of such auroral features in the current and most authoritative book, published by Joe Chamberlain (1961). Therefore, I sent a paper to the *Journal of Geophysical Research* reporting on our findings. The paper was rejected on the basis that there was nothing worth reporting, so I decided to analyze simultaneous all-sky films from a large number of stations. As I did, I became more convinced of the validity of my findings. A new paper was then sent to the late Sir David Bates, the editor of *The Journal of Planetary and Space Science,* who accepted it without review (Figure 2.17); I could assume this because I received his acceptance letter only about ten days after sending the paper to him.

However, I found it very difficult to convince my colleagues of my findings at first (although this paper was later recognized as one of the most cited papers by the *Science Citation Index* in 1979). This was particularly the case for those who were experienced in observing the aurora. This was because a single observer, standing at a point on the Earth, is carried by the Earth's rotation with a speed of 15° (in longitude) per hour, so that he gets an impression statistically that the fixed pattern was correct. Elvey was a firm believer in the fixed pattern concept. Many auroral scientists who have actually little experience in observing the aurora simply followed the experienced ones. Thus, it was hard to convince anyone about the validity of the concept of the auroral substorm. The only exception at that time was Feldstein, who strongly supported my findings.

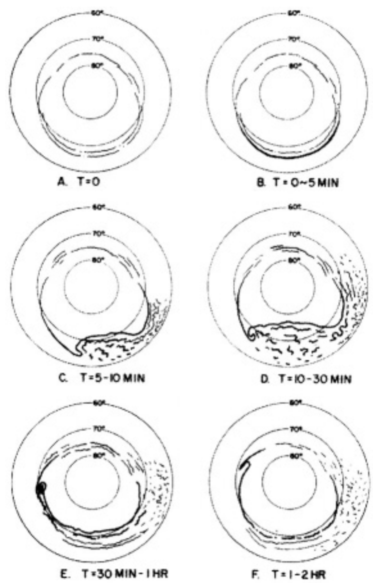


Figure 2.17 Schematic view of the auroral substorm as seen from above the geomagnetic pole: S.-I. Akasofu (1964).

I had to devise a scheme to prove the validity of the concept of the auroral substorm. The best way would have been to observe the aurora from a fixed point (with respect to the Sun) well above the North Pole for many hours, as the Dynamic Explorer satellite did in the 1980s. In the middle of the 1960s, this was nothing but a dream. One method I conceived was to fly westward under the aurora on a jet plane along the latitude circle of 65° or so. Because the speed of a jet plane is about the same as the rotation speed of the Earth at such a high latitude, it can cancel the effects of the Earth's rotation. Thus, a jet plane can stay in the midnight sector for about six hours by flying at midnight from the East Coast of the U.S. to Alaska. Both NASA and Air Force jet planes contributed to the so-called *constant local time (midnight) flights* many times for this study (Figures 2.18a, b, c).

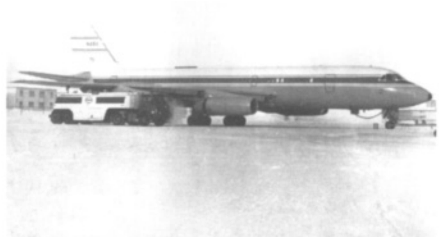


Figure 2.18a The NASA jet plane Galileo, which participated in auroral research: NASA Ames Research Center.

Figure 2.18b Operating instruments aboard the NASA jet Galileo: Geophysical Institute, University of Alaska.

Figure 2.18c Explaining instruments on the NASA Galileo to Sydney Chapman and Jerry Romick: Geophysical Institute, University of Alaska.

On my way back from one of my trips to Hanscom Air Force Base in Massachusetts, I learned that Elvey, who had since retired in Tucson, Arizona, was critically ill. I decided to visit him. Resting at his hospital bed, Elvey was waiting for my results. We sat together at his bedside to scan the all-sky film obtained by one of the constant local time (midnight) flights, which clearly registered intermittent auroral activities in the midnight sector. We shook hands firmly. He said, "Syun, you did a good job." I believe that I had finally convinced him of the validity of the concept of the auroral substorm. As I shook his hand, I noticed that his arms were just skin and bones. He died about ten days later.

In about 1966, I became convinced that a great variety of auroral and geomagnetic phenomena could be synthesized in terms of the substorm concept and I decided to write a book. It was published by D. Reidel Publishing Company in 1968 under the title *Polar and Magnetospheric Substorms* (Figure 2.19). It was dedicated as follows:

> *TO SYDNEY CHAPMAN, who unbeknown to most scientists, has encouraged and inspired the world's magnetic and auroral observatories to maintain the essential records upon which our understanding of geomagnetism and the aurora rests.*

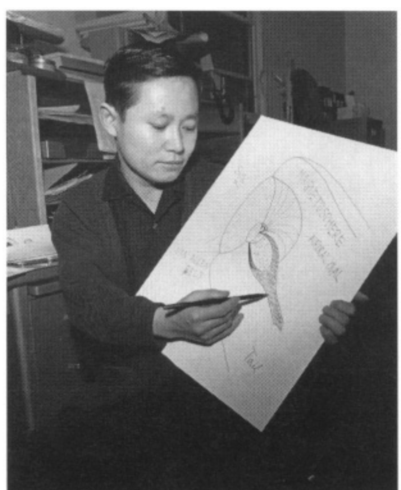


Figure 2.19 Announcing the publication of my first book *Polar and Magnetospheric Substorms*: Geophysical Institute, University of Alaska.

Chapman indeed had visited many magnetic observatories in the world and encouraged them to continue the recording. He told me that in the 1950s some prominent scientists were of the opinion that magnetic observatories were no longer necessary, because they thought that the geomagnetic daily variations and storms were already well understood. Some of them later regretted their premature judgment after noting the great development of geomagnetism, as it led to the discovery of the magnetosphere.

However, in spite of such an effort, the confirmation of the concept of the auroral substorm had to wait for images from satellites. The first images were obtained from the ISIS-2 satellite in 1971; they showed the clear pattern of the westward traveling surge (Figure 2.20), which is very similar to the auroral substorm pattern I constructed (Figures 2.16b an 2.17). I recall I was naturally excited in examining those images. This observation began to convince some researchers, both believers and nonbelievers alike.

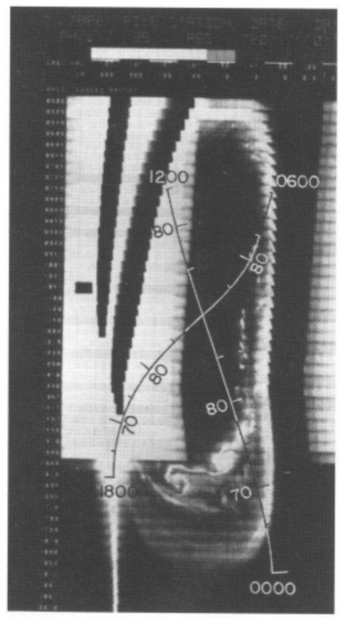


Figure 2.20 An auroral image depicting an auroral substorm: C. D. Anger, A.T.Y. Lui, and S.-I. Akasofu (1973).

The second set of convincing images came from the DMSP satellites (Figure 2.21) and showed different stages of the development of auroral substorms. Both polar-orbiting satellites scanned the polar region every 100 minutes or so, and were unable to obtain a sequence of images for a single substorm (as illustrated in Figure 2.22a, b). Nevertheless, many of my colleagues were surprised by the similarity of such images to each of the sequences of my substorm pattern (Figures 2.17 and 2.21). Ching Meng, Lee Snyder, Don Kimball, Jurgen Buchau, Jim Whalen, and many others joined me in a study of DMSP images.

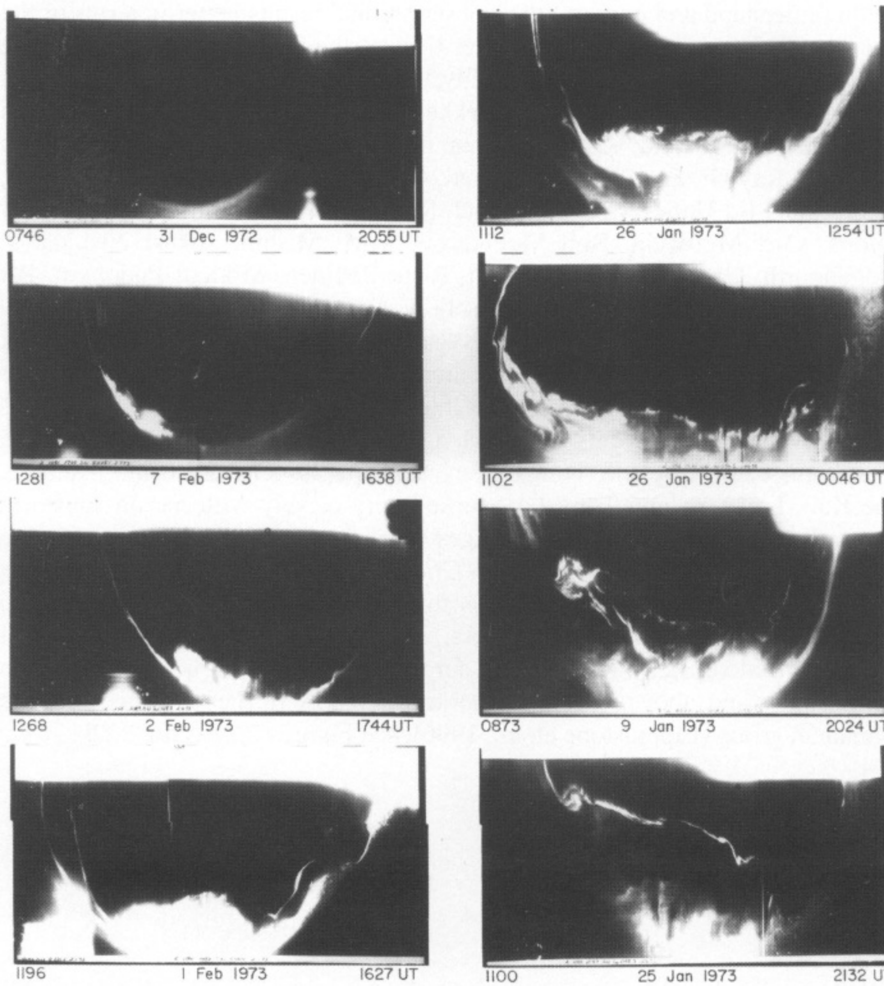


Figure 2.21 Snapshots of auroral substorm images by the DMSP satellite: S.-I. Akasofu (1975).

When the aurora was detected by the DMSP satellite for the first time, the images were classified as "Top Secret" data. I was asked to visit Hanscom Air Force Base to identify "strange lights" in the images. Not being a U.S. citizen then, they allowed me to see only pencil sketches of auroral images at that time. With Lee Snyder, we worked hard to declassify DMSP data.

During the next decade, I was fortunate in that many people realized they could understand and interpret their observational results better in terms of the concept of the auroral substorm rather than of the fixed pattern. Among the early participants were Roger Arnoldy, Dan Baker, Peter Banks, Wolfgang Baumjohann, J. Birn, Asgeier Brekke, Jim Burch, Ferd Coroniti, Stan Cowley, Don Fairfield, Carl-G. Fälthammer, Yasha Feldstein, Lou Frank, Yu Galperin, Ray Greenwald, Don Gurnett, Gerhard Haerendel, Walter Heikkila, H. Herman, Ed Hones, Charlie Kennel, Richard Lundin, W.B. Lyatsky, Larry Lyons, Carl McIlwain, Bob McPherron, V.M. Mishin, Atsusi Nishida, J. Opgenoorth, George Parks, R. Pellat, Risto Pellinen, Mikhail Pudovkin, Pat Reiff, Gordon Rostoker, Chris Russell, V.A. Sergeev, George Siscoe, Dan Swift, O.A. Troshichev, N.A. Tsyganenko, Vytenis Vasyliunas, Jack Winckler, and Dave Winningham. Later, waves of the new generation joined in our effort, particularly during the International Conference on Substorms (ICS). My former students Ching Meng, Koji Kawasaki, Lee Snyder, Fumi Yasuhara, Paul Perreault, Tom Berkey, and my associates Yosuke Kamide, Joe Kan, Lou Lee, and Tony Lui worked very closely with me on substorm research. Finally, long-awaited images from the Dynamic Explorer satellite began to arrive (Lou Frank and John Craven, 1988). I visited my colleagues at the University of Iowa to witness this event. I thanked Lou Frank and congratulated him on this great success. It was the ultimate test of the concept of the auroral substorm because the auroral substorm must be the same seen from below and above. Auroral morphology was further advanced by the Canadian group (Elphinstone et. al., 1996); see Figures 2.22a and 2.22b.

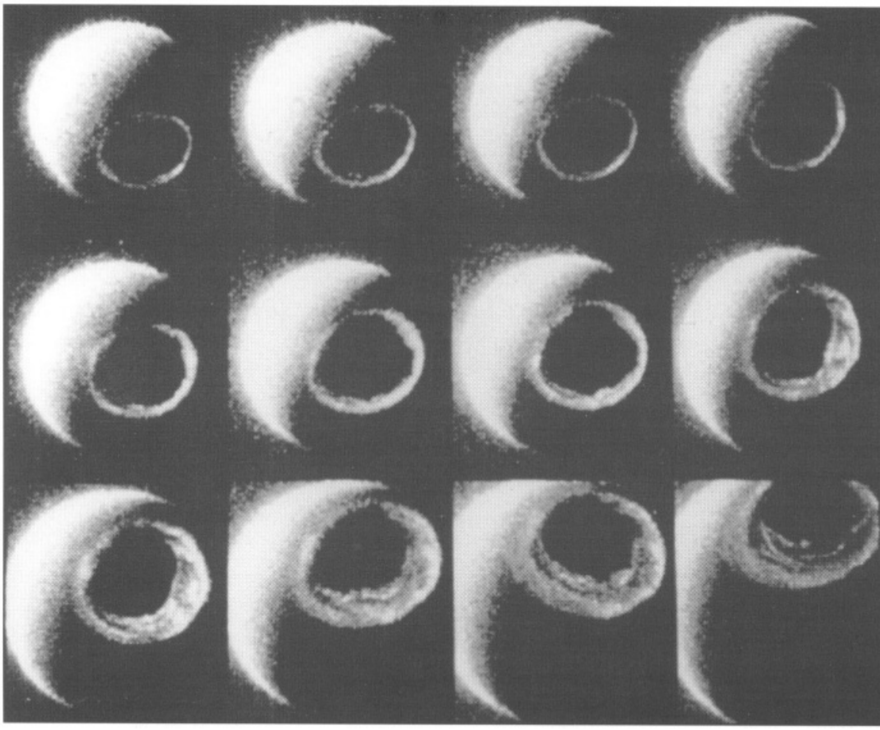


Figure 2.22a Auroral images taken from the DE satellite that depict the development of an auroral substorm: J.D. Craven, Y. Kamide, L.A. Frank, S.-I. Akasofu, and M. Sugiura (1983).

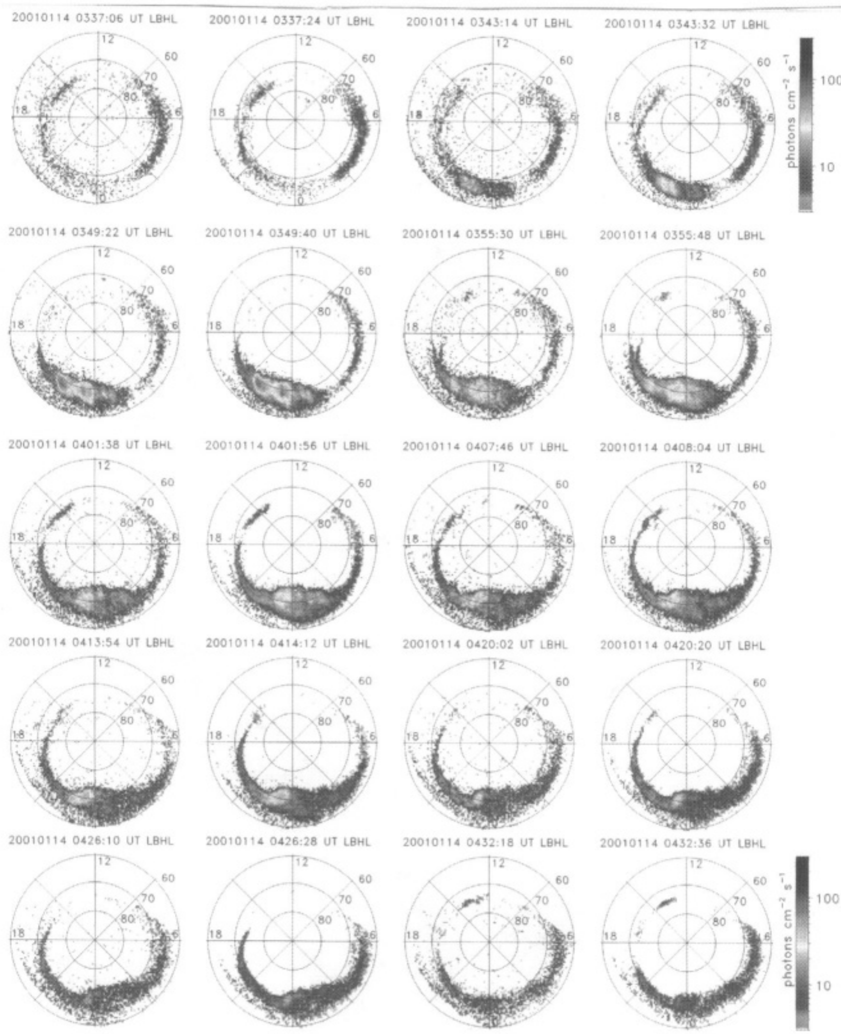


Figure 2.22b A typical substorm depicted by the POLAR satellite: G. Parks.

It is important to learn that it takes much more time than one thinks to convince colleagues if one's finding is radically different from what has been believed for years. Figure 2.23 shows schematically the auroral features at about the maximum epoch of a typical substorm. The visible feature consists of three parts, as shown on the left-hand side, the dayside part, the nightside part, and the diffuse aurora, which is located equatorward of the first two arc structures. Note the diffuse aurora evolves into many arcs that develop further complex folds.

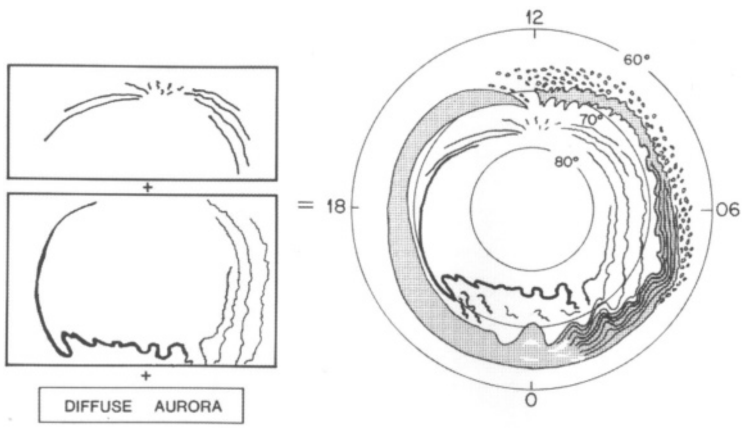


Figure 2.23 A schematic representation of auroral features at the maximum epoch of an auroral substorm: S.-I. Akasofu (1976).

As the study of substorms had progressed by the work of a large number of researchers in the 1980s, we thought that an organized effort was needed to advance it further. Joe Kan was instrumental in establishing the International Conference on Substorms (ICS). The first conference was held in 1992 under the leadership of Bengt Hultqvist at the Swedish Institute of Space Physics, in Kiruna, Sweden. The second conference was held in 1994 at the University of Alaska Fairbanks, commemorating the publication of my 1964 paper on auroral substorms. The conference was blessed by active auroral displays over Fairbanks. The ICS brought many younger researchers who have considerably advanced the study of magnetospheric substorms. I also express here great appreciation for the close interaction with a number of groups, the Swedish group in Stockholm, headed by Carl-Gunne Fälthammer; the Norwegian group in Oslo, headed by Alv Egeland; the Danish group, headed by Knud Lassen; the Canadian group headed by Cliff Anger; the Russian groups at Apatity, Moscow, Irkutsk, and Petersburg; and many U.S. groups, including Aerospace Corporation, Boston University, Johns Hopkins

University, University of New Hampshire, Rice University, Southwest Research Institute, University of Washington, UCLA, U.C. Berkeley, and UCSD. It is my sincere hope that this book will provide some historical background in substorm research and some new direction.

2.5 Auroral Storms

In 1962, when I began to study the great geomagnetic storm of February 11, 1958 (one of the most intense storms in the twentieth century, with the maximum Dst decrease being as large as 450 nT at 1000 UT; Figure 2.24a), I was greatly surprised to find that the aurora can be *very quiet* when the main phase of a great storm is reaching its maximum epoch; it is natural to assume that the intensity of auroral displays would also reach highest at the maximum epoch. In Figure 2.24b, the aurora at 1020 UT, on February 11, 1958, was quiet, in spite of the fact that the geomagnetic storm was at about the maximum epoch; the auroral oval (both the northern and southern boundaries) expanded greatly towards the equator. I thank Carl Gartlein who installed a number of all-sky cameras near the US-Canada border that could accurately locate the expanded oval on that day.

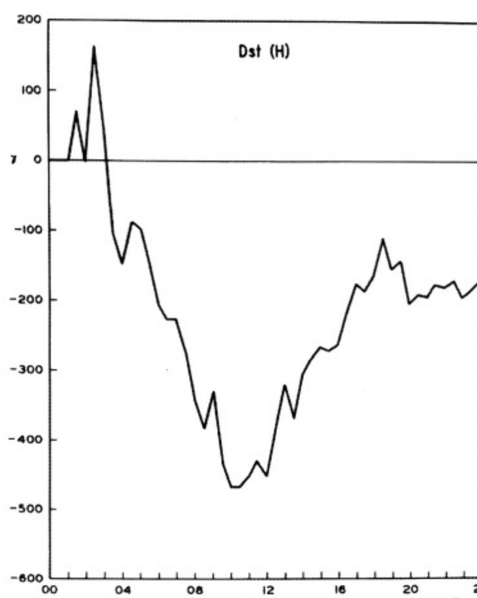


Figure 2.24a The Dst index during the geomagnetic storm of February 11, 1958: S.-I. Akasofu and S. Chapman (1962).

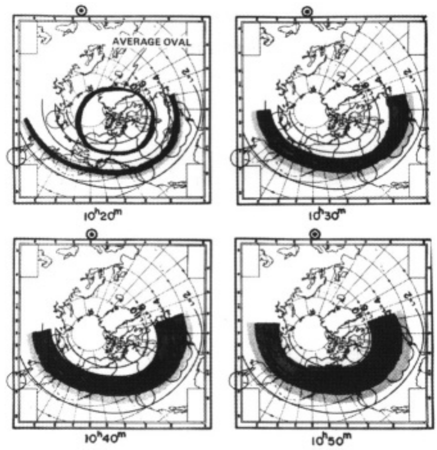


Figure 2.24b The distribution of the aurora at about the maximum epoch of the main phase of the geomagnetic storm of February 11, 1958: S.-I. Akasofu and S. Chapman (1962).

However, such a quiet condition did not last too long, and a great auroral substorm activity began soon afterward, resulting in one of the most spectacular poleward expansions of the auroral oval. There were similar displays a few hours before and after the event that began at 1020 UT. That is, similar auroral activity repeated several times during the great storm. The onset of the poleward expansion at 10:20 UT was observed at gm. lat. as low as 54°, L = 2.7 (Pullman, Washington).

This event became also the basis for Chapman and me to consider that an auroral storm consists of a number of distinct impulsive phenomena, namely the auroral substorms. That is to say, an auroral storm consists of a number of auroral substorms; an auroral substorm is the element of the auroral storm.

In my 1964, paper on auroral substorms I reported how individual auroral substorms develop (Figure 2.18). In high latitudes, a geomagnetic storm also consists of a number of impulsive changes (Figure II in the Prologue), coinciding with the auroral substorm, which is called the polar magnetic substorm. This idea developed into the concept of a magnetospheric storm that consists of a number of magnetospheric substorms. Thus, it was concluded that it is necessary to study magnetospheric substorms in order to understand a magnetospheric storm. It was in this way that the magnetospheric substorm became one of the main topics of research in magnetospheric physics.

Chapman used to tell me how lucky I was as a student of the aurora, since the IGY provided my generation with a great wealth of auroral data. In our precomputer days, the data analysis by hand was laborious, but there was enough time to consider what the data were trying to tell us. Compared with present data gathering, our days were almost like the Stone Age. Further, most of the data one needs, including real-time auroral images from the POLAR satellite, can now be obtained instantly by clicking on a computer; in the 1960s, 1970s, and even in the 1980s, it took several years to gather necessary data. My only hope is that the new generation of researchers would not have indigestion because of the present wealth of data.

2.6 Auroral Rays

The curtain-like structure of the aurora frequently develops vertical striations called the auroral rays. Tom Hallinan, of the Geophysical Institute, developed a high-sensitivity TV camera and with it captured these striations clearly for the first time (1983). He found that the rays are a sort of fine pleating of the curtain-like structure of the aurora.

At that time, I had a graduate student, John Wagner, who wanted to study the ray structure by a computer simulation method. The problem I faced was that the Geophysical Institute did not have a high-speed computer for such a project and I knew nothing about a plasma simulation method. This was one of the most difficult problems I faced as a professor, but I was determined to solve it. Thus, I contacted Walter Orr Roberts, the founder of the National Center for Atmospheric Research and my other mentor, to seek advice on how I might obtain a high-speed computer; he arranged to find $500,000 from the Fleishmann Foundation for this purpose. It was in this way that the Geophysical Institute obtained the first modern high-speed computer. Then, I sent Wagner to work with UCLA's John Dawson for a year to learn about a plasma simulation method. Wagner satisfied my expectations and simulated the ray structure that was successfully imaged by Hallinan. Wagner showed that the ray structure develops as a result of the counter-streaming plasma flow across an auroral curtain (Figure 2.25).

2.7 Thickness of an Auroral Curtain

As far as I am aware, I was the first to determine accurately the thickness of auroral curtains. The corona-type aurora is observed when an auroral curtain is located near the magnetic zenith. In some of the aurora photographs I took, the bottom edge of the curtain was clearly captured.

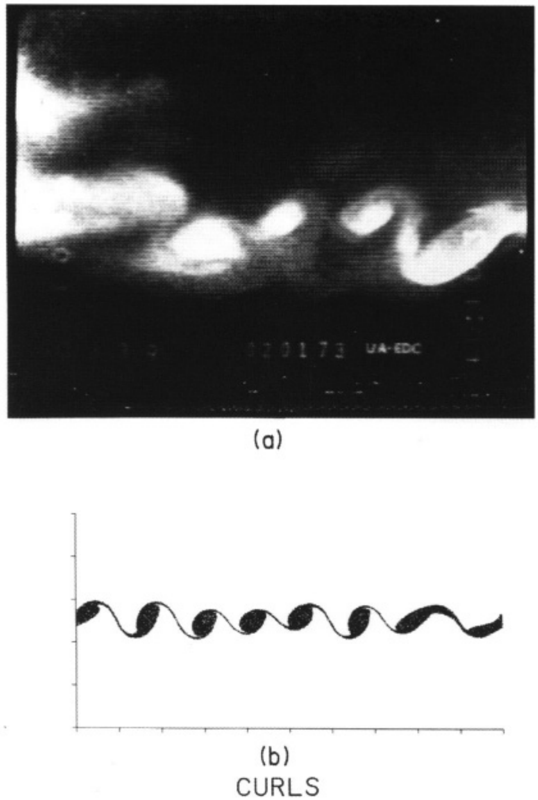


Figure 2.25 An auroral ray structure observed by a high-sensitivity TV camera) and a simulated one: J.S. Wagner, R.D. Sydora, T. Tajima, T. Hallinan, L.C. Lee, and S.-I. Akasofu (1983).

Using a star constellation map, I could measure the thickness to be about 500 meters. Chapman encouraged me to publish the result, and I wrote a short paper that appeared in the *Journal of Atmospheric and Terrestrial Physics*. It should be noted that the reason for the thin curtain-like form of the aurora is still one of the long-standing unsolved problems. The question remains why the field-aligned currents occur in the form of thin sheets.

John Wagner also simulated the auroral potential structure, which may be responsible for accelerating auroral electrons and for generating the radiation. However, there is still no agreed-upon mechanism for the acceleration process of auroral electrons. There have been a number of efforts to learn about the individual curtain-like structure of the aurora and precipitating electrons (Figure 2.26). The acceleration of charged particles in the magnetosphere and the solar atmosphere, perhaps even in galaxies, is one of the most fundamental issues in cosmic electrodynamics.

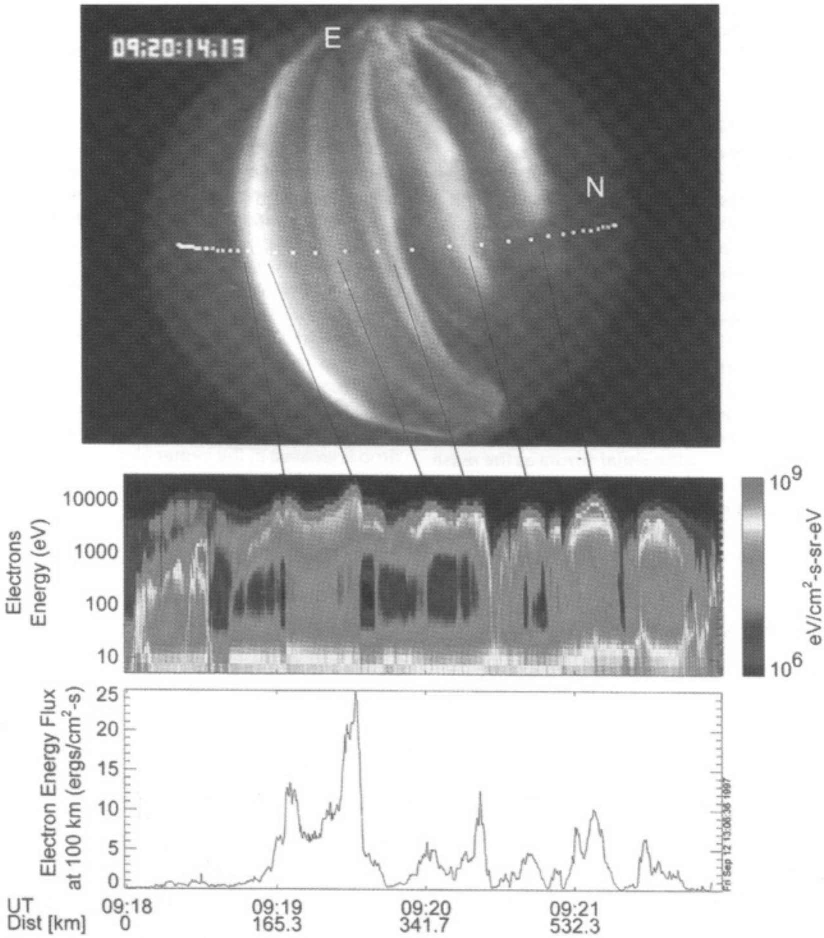


Figure 2.26 The Fast Auroral Snapshot Small Explorer (FAST) satellite passing through multiple auroral arcs at 0920:14 UT, February 6, 1997. The 110-km conjugate to the satellite is shown in the all-sky image at 10s intervals as FAST passed across from left (south) to right (north). The center panel is a "normal" format of the electron energy spectrum (integrated overall pitch angles) and shows a number of inverted-V structures. The bottom panel is the precipitated energy flux on a linear scale. The individual auroral arcs are clearly displayed here. The auroras, in particular the two arcs to the right (north), did change over the 4 minutes of the pass, so a detailed comparison between all-sky image and the particle data could not be attempted: H.C. Stenbaek-Nielsen, T.J. Hallinan, D.L. Osborne, J. Kimball, C. Chaston, J. Fadden, G. Delory, M. Temerin, and C.W. Carlson (1998).

I salute Duncan Bryant's devotion to this subject. He has pursued the processes associated with the acceleration of auroral electrons for more than twenty years, in spite of the fact that his idea was not widely accepted.

Bryant's efforts are well described in his recent book *Electron Acceleration in the Aurora and Beyond (1999).*

2.8 Auroral Kilometric Radiation

In 1973, I had an opportunity to visit the University of Iowa with the newly acquired DMSP images in hand. I also was able to talk to Don Gurnett who told me that his radio detector aboard the IMP-6 satellite had observed a new type of radio emission and was wondering if it was related to auroral activity. It was fortunate that the period his data covered coincided with the period that my DMSP images covered. It became immediately obvious to both of us that the radio emission occurred when DMSP images showed intense auroral activity. We published a joint paper on this subject (Figure 2.27). It is likely that the auroral kilometric radiation is emitted from the auroral electrons' acceleration region.

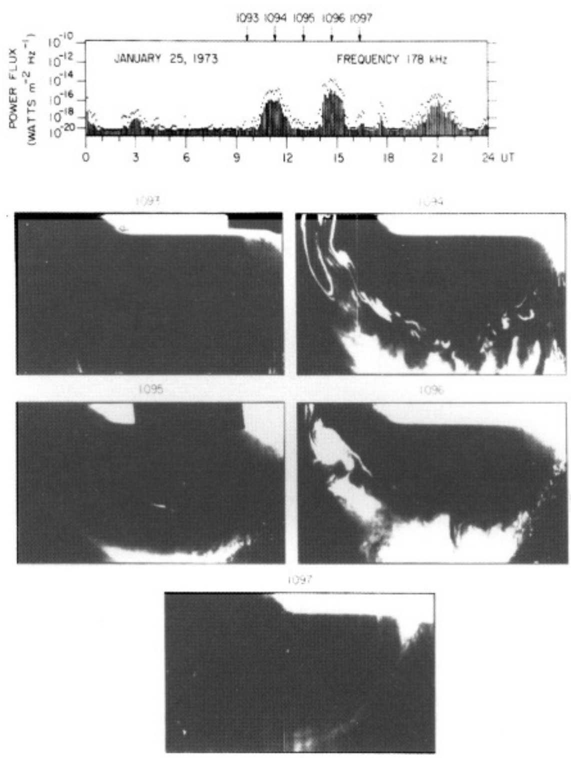


Figure 2.27 Auroral kilometric radiation observed by the IMP-6 satellite (D. Gurnett) and DMSP images: G.R. Voots, D.A. Gurnett, and S.-I. Akasofu (1977).

2.9 Auroral Observation at the South Pole

It is difficult to observe the midday part of the auroral oval in the northern hemisphere. Because the geomagnetic pole is located in the northwestern part of Greenland (instead of at the geographic pole), the midday aurora is observable at Svalbard and in the middle of the Arctic Ocean during a short period around the winter solstice. Even so, the upper part of the midday aurora is in sunlight. In 1970, we had a NASA airborne expedition over the Greenland Sea and confirmed the presence of the midday aurora in spite of the twilight. I clearly remember Bob Eather's excited voice on the intercom, describing that the spectrum was rich in the oxygen-red emission.

It so happens that the South Pole is located at gm lat. 78°, and the shadow height is highest on Earth. Therefore, one can clearly observe the entire midday part of the auroral oval. It is the most ideal location to observe the midday aurora. Merritt Helfferich was sent to install a Fairchild all-sky camera at the South Pole and high-resolution images of the midday aurora were obtained (Figure 2.28). The films provided us with many new results on the midday aurora.

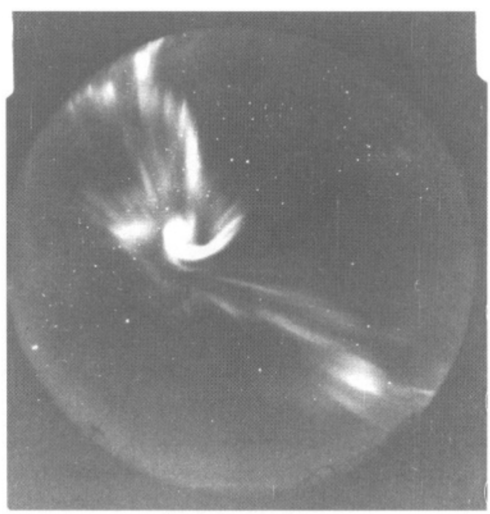


Figure 2.28 Midday aurora observed at the South Pole station: S.-I. Akasofu (1978).

2.10 Auroral Spectra as Tools for Detecting Extraterrestrial Life

One of the most prominent emissions from the aurora is the greenish-white light from oxygen atoms, while the Jovian aurora contains atomic hydrogen emissions (Clarke et al., 1989). Most of the processes leading to the production of oxygen atoms are directly or indirectly related to molecular oxygen produced near ground level. Thus, the oxygen emission, the so-called *green line* (557.7 nm) of the terrestrial aurora, arises mostly because plants release abundant free oxygen into the atmosphere by photosynthesis.

Thus, an intense green line emission suggests that plant life exists on Earth. It is expected that the green emission from oxygen atoms dissociated from CO_2 may also exist, but its contribution is very small. This is because even on the ground level, the amount of CO_2 is about 1/1000 that of O_2.

It was recently reported that Upsilon Andromedae, which is a solar-type star, has three planets. This discovery is significant because it shows the planetary system, like the solar system, is not quite unique. It is expected that a number of stars are accompanied by several planets, and it may not be too long before the aurora on such planets can be discovered.

One possible way to detect plant life on such planets is to examine their auroral emissions. If strong oxygen emissions can be detected among other emissions in the planetary auroral oval, the possibility of the presence of plant life is high. Further, if plant life exists, animal life, whether lower or higher, can also exist there.

The Earth-like auroral processes leading to the green light emission from the auroral oval require, in addition to plant life, both stellar wind and planetary magnetism. It is highly probable that solar-type stars have stellar wind. If such a planet does not have a strong dipole-like magnetic field, the stellar wind can cause atmospheric glow, in which oxygen emissions may be present. In any case, if *strong* emissions are detected in the planetary auroral spectra, the possibility of oxygen there is high.

There is no doubt that the detection of the green line is technically a very challenging problem, particularly from ground level. However, the planets expose their full dark sides to the Earth once during their revolution around their parent stars. Further, there are a number of prominent oxygen emissions in the infrared and far ultraviolet ranges that can be detected by satellites. In any case, this is only a technical problem to be solved.

Auroral science will evolve in a variety of ways in the future. It would be a great boon for auroral science if it could contribute to the search for extraterrestrial life, one of the ultimate human endeavors.

2.11 Emperor Showa and the Aurora

On October 3, 1985, I gave a special lecture on the aurora for the emperor of Japan in his palace in Tokyo. It seems the emperor, a marine biologist, had an unusual interest in the aurora and prepared a large number of questions before my lecture. After my slide presentation, he asked how we could confirm ancient sighting reports of the aurora in Japan. He was not satisfied with my response that anomalous events in the sky were well documented in an ancient publication titled *Japanese Meteorological Data,* and he asked further how one could confirm such sightings as auroral events. He was visibly pleased to learn that the dates of these sightings coincided with those recorded elsewhere in the world.

Many people still believe that the aurora occurs more frequently as we approach the north magnetic pole. The emperor was not an exception. Thus, when I told him that this is not the case, he was very puzzled. However, he was delighted when I showed him an image of the ring-shaped aurora taken from the Dynamics Explorer satellite (provided to me by Lou Frank, University of Iowa) and I explained that at the geomagnetic pole, which is located near the center of the auroral oval, the aurora is located well beyond the southern horizon most of the time. He was pleased to learn that the aurora appears along an annular ring. This is because on his flight from Anchorage to London, he observed the aurora above Alaska and expected to see more at higher latitudes. However, he could not see the aurora farther north. Obviously, he was very puzzled by this experience, but the satellite image I presented solved his puzzle. He also wanted to know about the auroral spectral composition. "I want to make sure that the auroral green line comes from atomic oxygen, not from molecular nitrogen," he told me.

A videotape of a spectacular auroral display that was recorded at our Poker Flat Research Range, Geophysical Institute, University of Alaska Fairbanks, fascinated the emperor. It was projected on a 150-inch screen, which was kindly provided by Panasonic. The emperor was pleased to know that a Japanese company produced the high-sensitivity video camera. He asked how often we could observe such a display in Fairbanks (how many days per week and then how many hours on active nights, and so on). He was interested in astrophysical implications and practical applications of auroral studies, and he asked me a few basic and technical questions about them. I also recall that on the occasion when I received the Japan Academy Award in 1977, he asked me if there is any relationship between the solar corona and the aurora. I will never forget my meeting with the emperor.

2.12 An Exciting New Development

The auroral dynamo requires both the solar wind and a planetary magnetic field to be available. Thus, it was expected that both Jupiter and Saturn would have the auroral oval, while Venus and Mars would not. The Hubble Space Telescope depicted a clear image of the auroral oval on Jupiter and Saturn (Figure 2.29), while both the Venus orbiter and the Mars orbiter did not find any indication of an auroral oval. These results confirm that the solar wind-planetary magnetic field interaction is essential in providing the power for auroral discharge, as we learned was the case for the Earth.

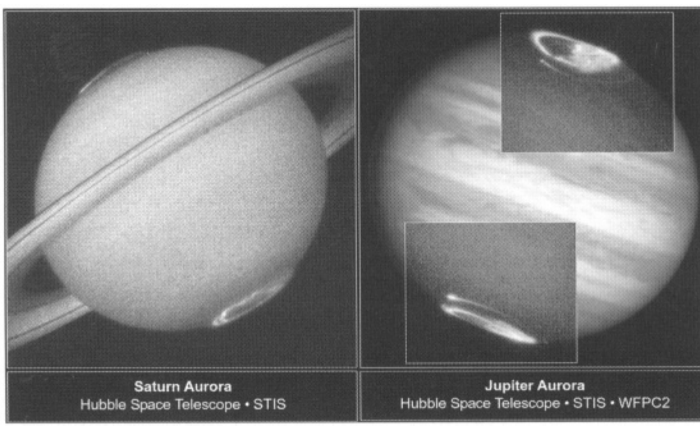
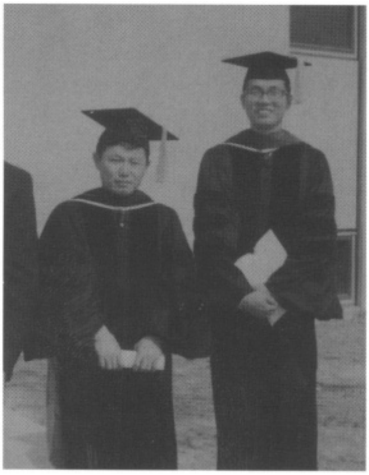


Figure 2.29 The aurora on Saturn and Jupiter: A. Bhardwaj and G.R. Gladstone (2000), NASA Hubble Space Telescope Project.

The aurora on Earth: L.A. Frank.

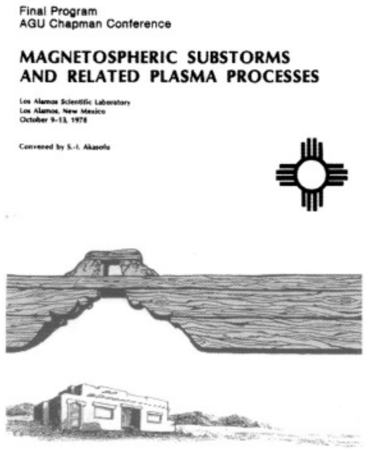
The solar wind causes both the aurora and the comet's tail: Geophysical Institute, University of Alaska.



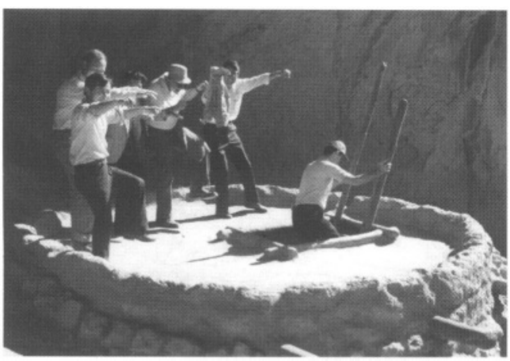
With Ching Meng, my first Ph.D. graduate (1968).

With Charlie Kennel and Vytenis Vasyliunas, at the Kiruna Geophysical University.

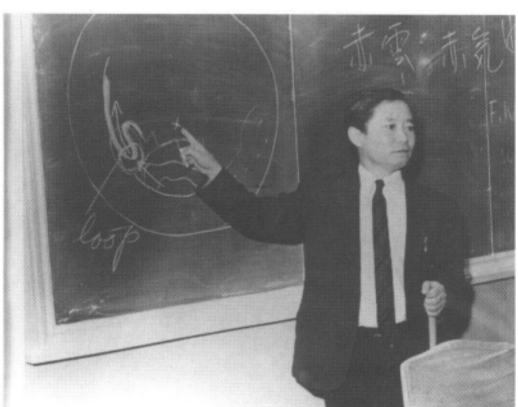
From left to right, Mrs. Olson, Mrs. Kathy McPherron, Hiroshi Fukunishi, Bob McPherron, Atsushi Nishida, Gordon Rostoker, and Syun Akasofu at the Grenoble, France, IUGG General Assembly meeting (1975).



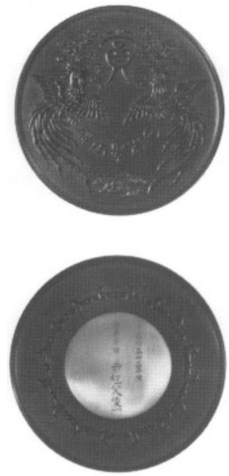
The cover of the program of the AGU Chapman Conference on *Magnetosphere Substorms and Related Plasma Processes*, held at Los Alamos Scientific Laboratory, October 9-13, 1978.



On the occasion of the Los Alamos Conference (1978), several attendees went to Pueblo to dance, from left Syun Akasofu, Carl McIlwain, unidentified, Jack Winckler, and Ching Meng (1978).

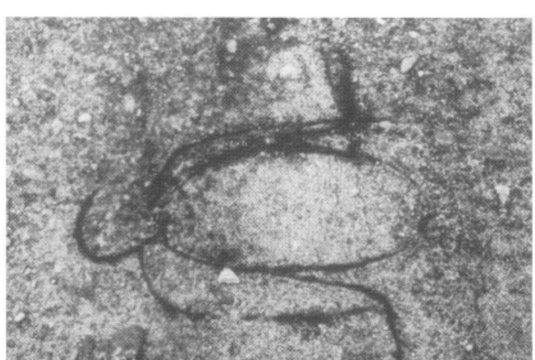


Lecturing at the Institute of Space Physics as an invited guest by the People's Republic of China.



The Japanese Academy Award (front and back of award).

Mike Schultz and Syun Akasofu discussing magnetospheric processes: J. Olson.



A drawing we scratched into the ground: J. Olson.

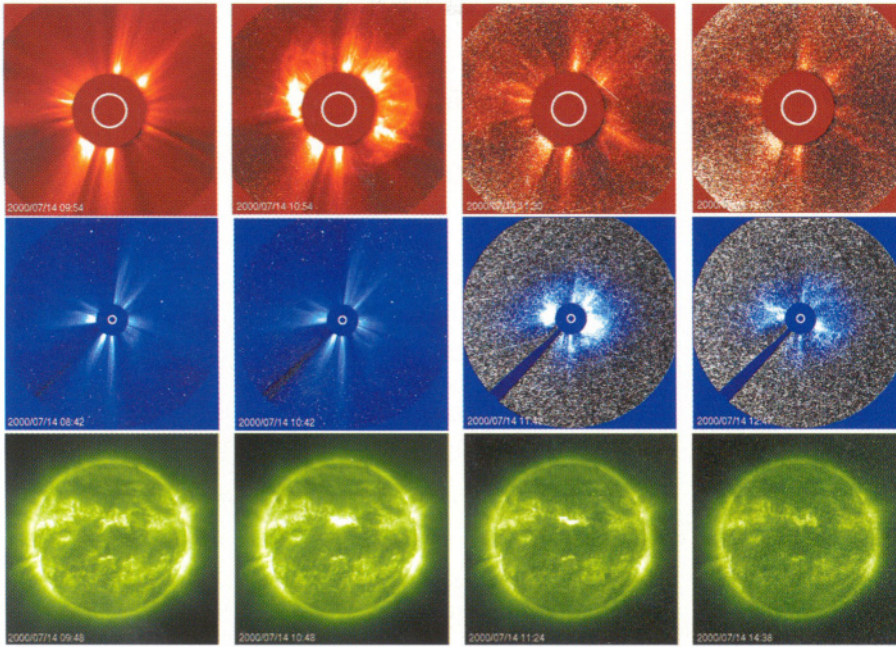


Plate 5: Solar activity observed by the Solar and Heliospheric Observatory (SOHO) on July 14, 2000. The instrument was bombarded by solar protons in the last two images: NASA/ESA SOHO Project.

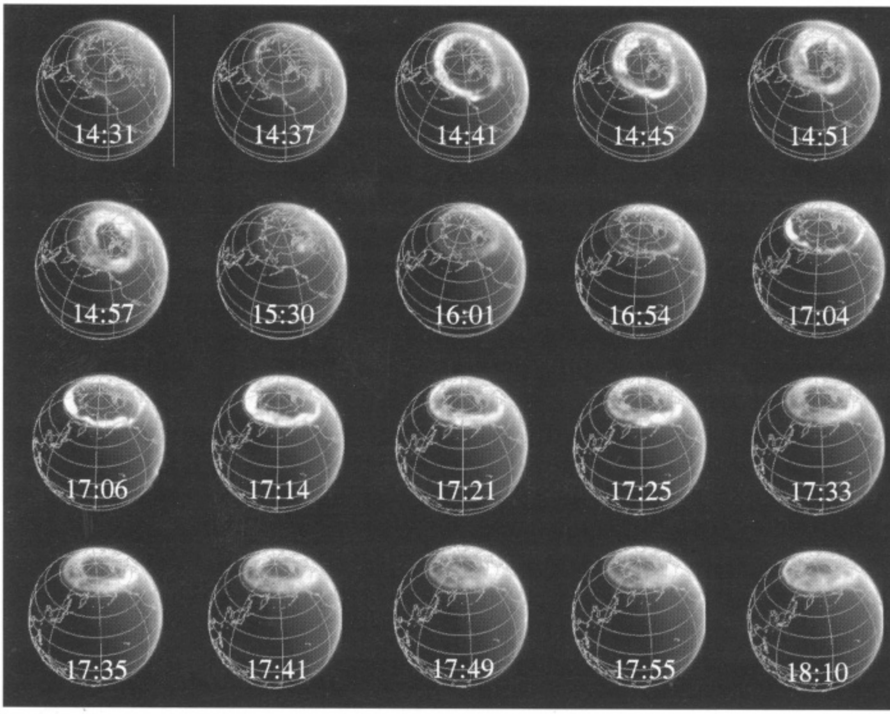


Plate 6: Auroral displays during July 15, 2000: NASA/IMAGE FUV Team.

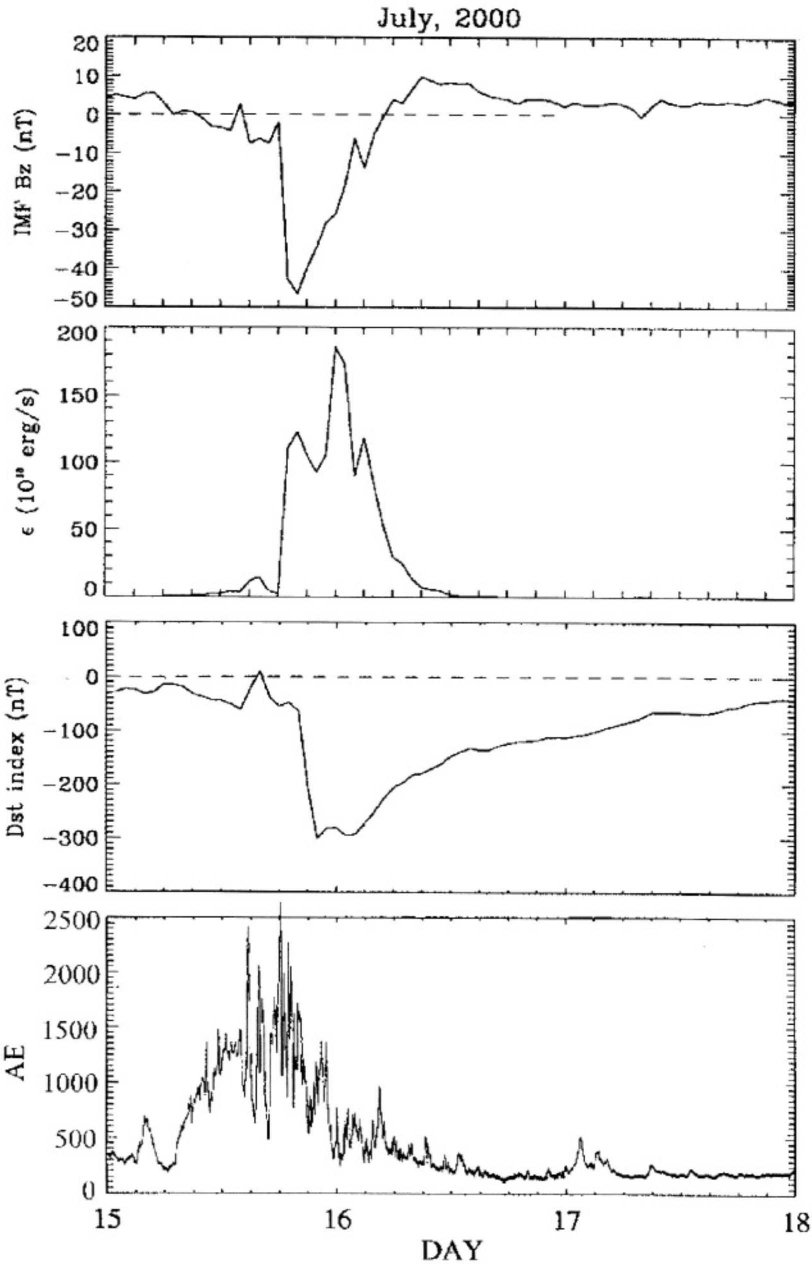


Plate 7: From the top, the IMF *Bz*, epsilon function (ε), and the geomagnetic indices Dst and AE on July 15, 16, and 17, 2000: World Data Center for Geomagnetism, Kyoto University.

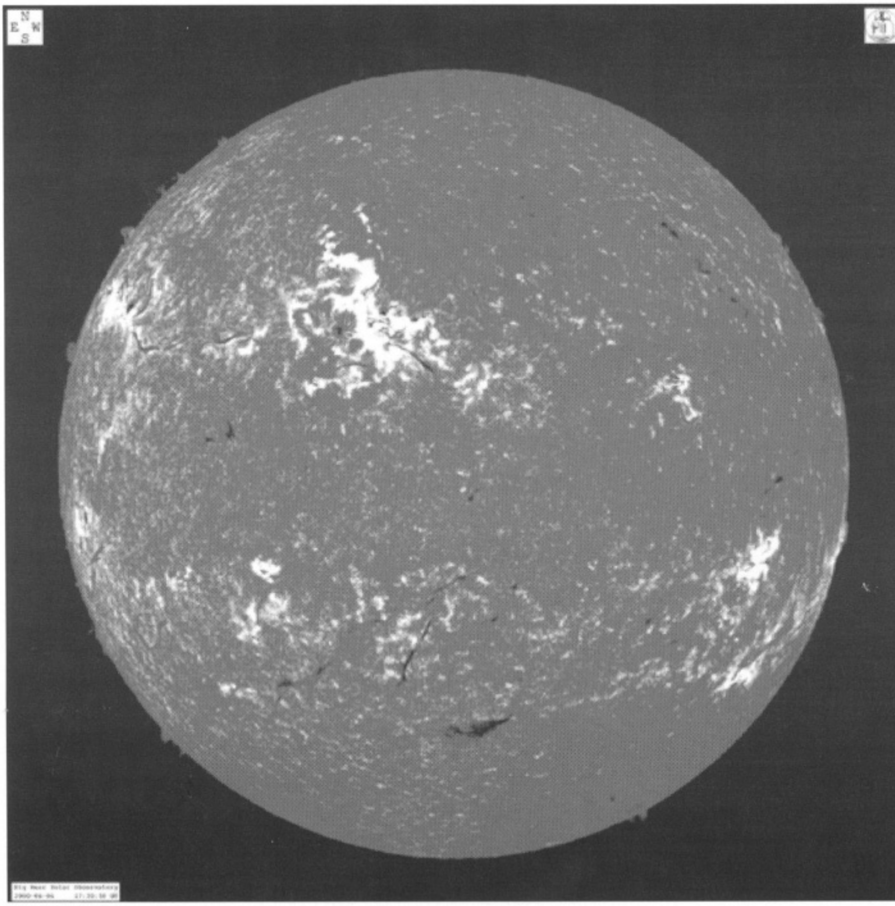


Plate 8: Solar flare on June 6, 2000: Big Bear Solar Observatory.

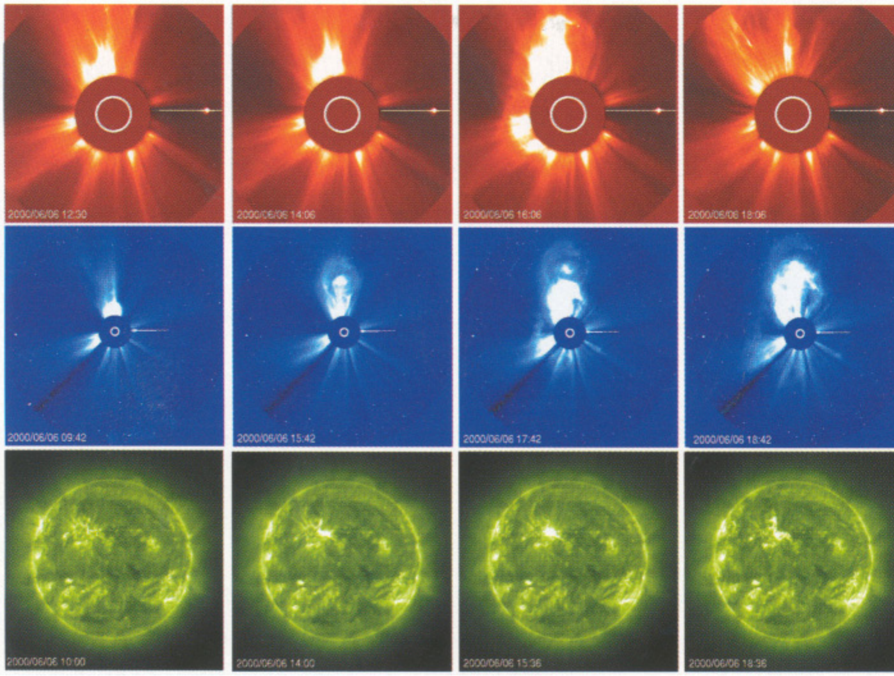


Plate 9: Solar activity on June 6, 2000: NASA/ESA SOHO Project.

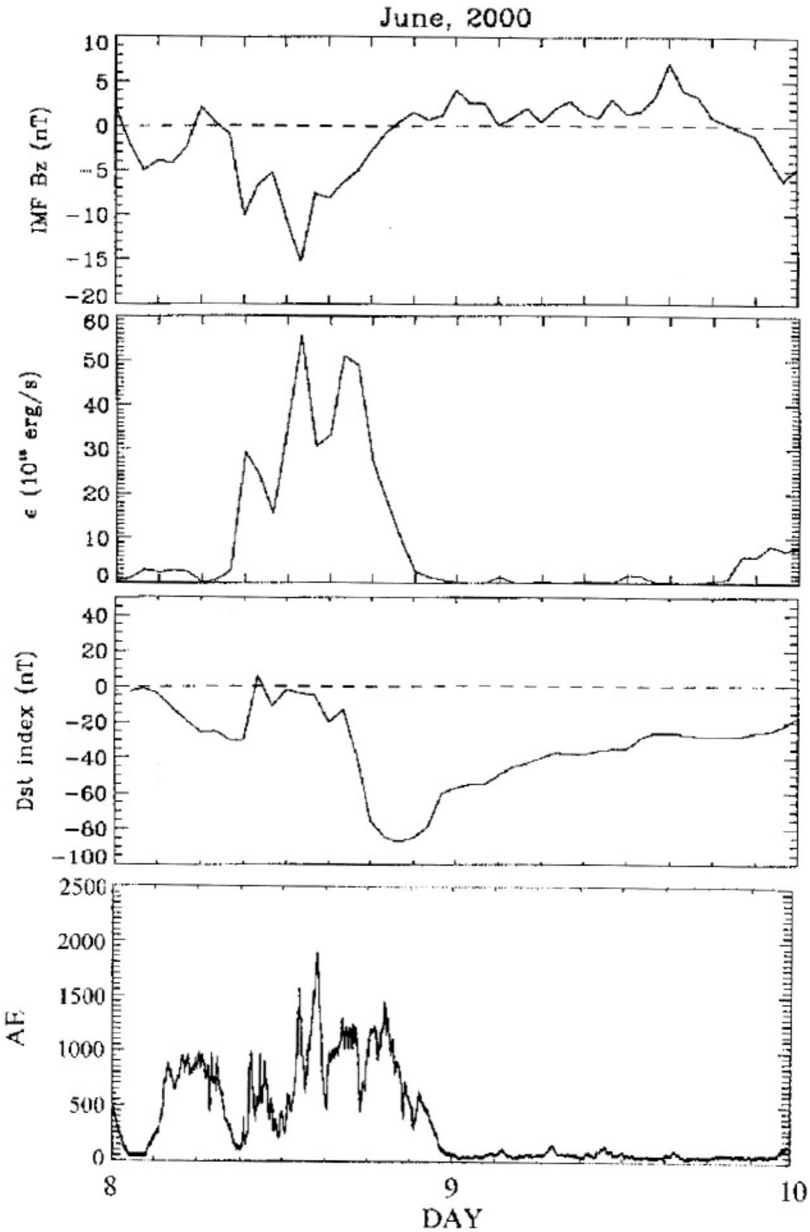


Plate 10: From the top, the IMF *Bz*, the epsilon function (ε), and the geomagnetic indices Dst and AE on June 8, and 9, 2000: World Data Center for Geomagnetism, Kyoto University.

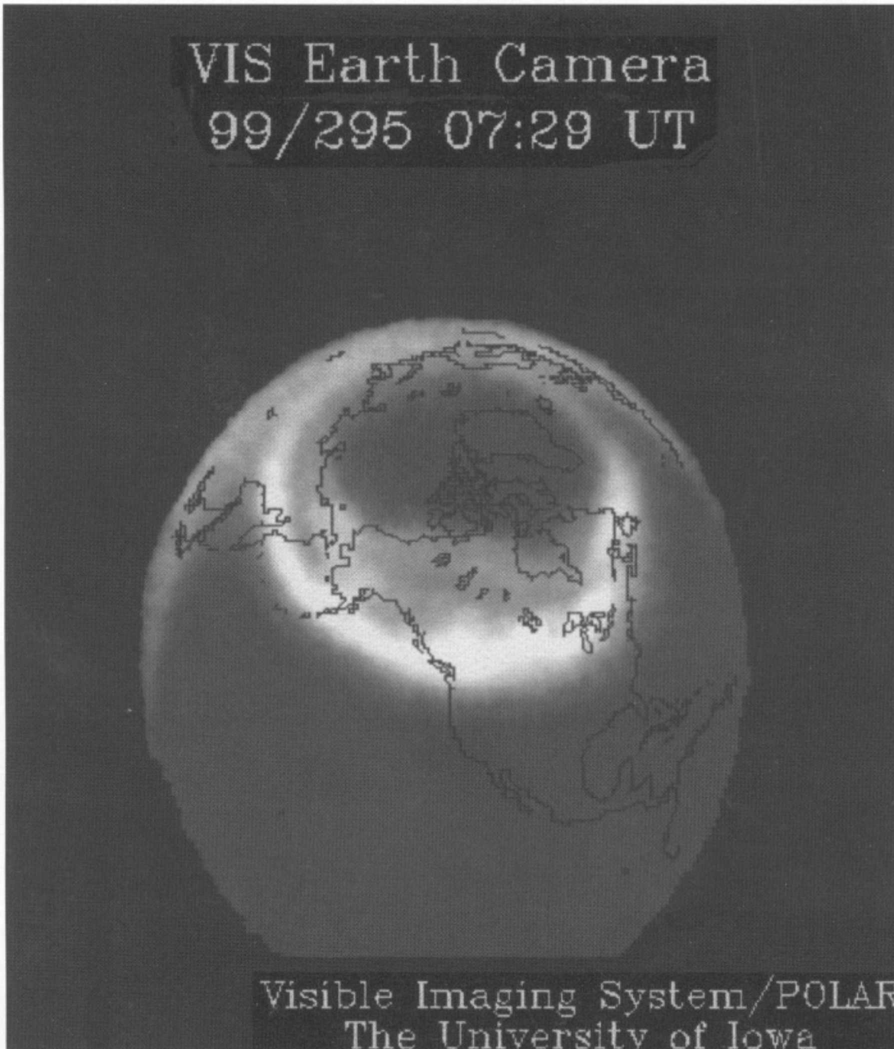


Plate 11: The auroral oval during the great geomagnetic storm of October 22, 1999. Imaged by the POLAR satellite: NASA/IMAGE FUV Team.

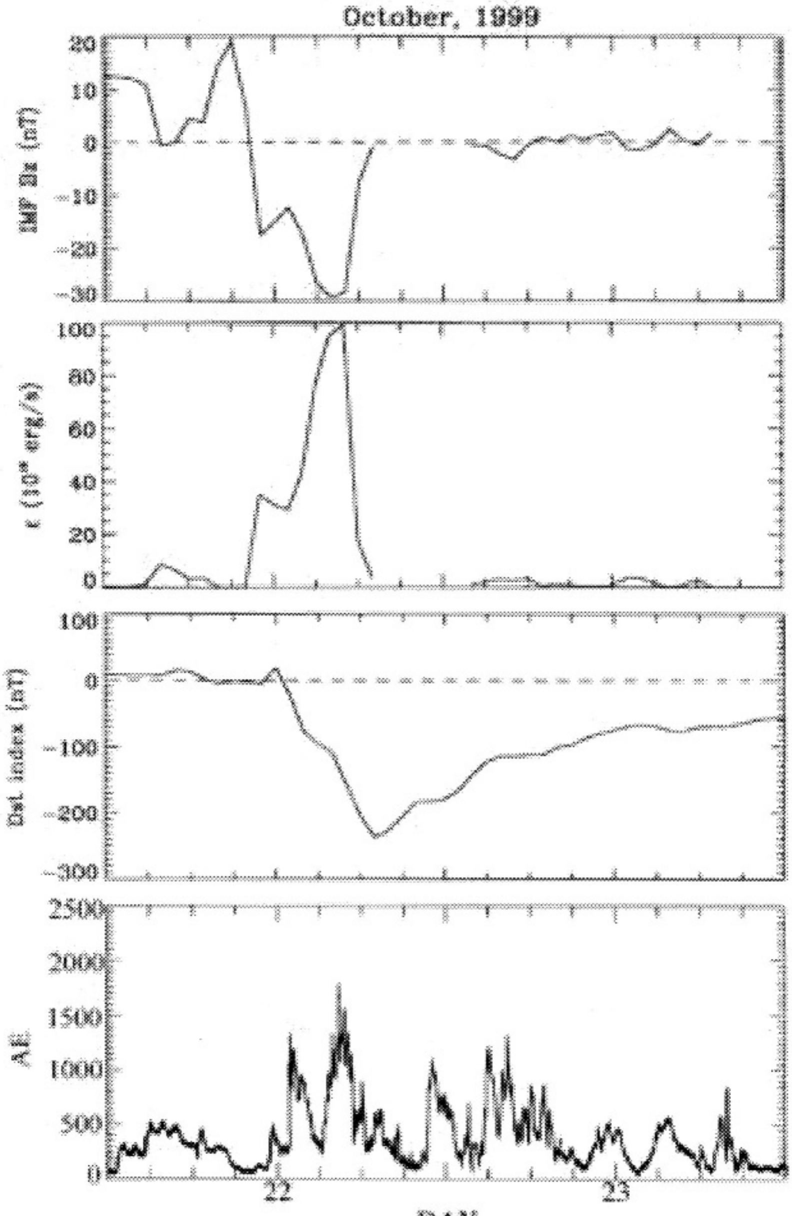


Plate 12: From the top, the IMF *Bz*, the epsilon function (ε), and the geomagnetic indices Dst and AE on October 21, 22, and 23, 2000: World Data Center for Geomagnetism, Kyoto University.

Chapter 3

REALIZING THE DREAM OF OUR PIONEERS: POLAR MAGNETIC SUBSTORMS AND THE ASSOCIATED CURRENT SYSTEM

In modern terms, the Earth and its magnetic field are enclosed in a comet-shaped cavity, called the magnetosphere, which is carved in the solar wind, a high-temperature plasma flow from the Sun. Geomagnetic disturbances can be defined as the magnetic manifestation of an increased level of the solar wind-magnetosphere interaction, resulting in an increased electric power output from the solar wind-magnetosphere dynamo process. Magnetic fields generated by the resulting increased electric currents are defined as geomagnetic disturbance fields $\Delta\boldsymbol{B}$. The disturbance magnetic fields are recorded as variations superimposed on the main field $\boldsymbol{B}_o$, namely $\boldsymbol{B} = \boldsymbol{B}_o + \Delta\boldsymbol{B}$. More precisely, $\Delta\boldsymbol{B}$ includes the solar quiet-day daily variation that is caused by solar tidal effects on the ionosphere and the Chapman-Ferraro current on the magnetopause.

One of the important purposes of the discipline of geomagnetism is to examine the configuration of the electric current systems that cause geomagnetic disturbances $\Delta\boldsymbol{B}$, and to elucidate their driving processes in terms of the solar wind-magnetosphere interaction. This long-term effort began with the pioneering effort of scientists such as K. Birkeland, C. Störmer, Sydney Chapman and Hannes Alfvén at the beginning of the twentieth century. Before satellite observations gave views of the three-dimensional space around the Earth, the early studies were limited to inference of the current systems in space around the Earth $\boldsymbol{J}$ (r, θ, λ, t) on the basis of records of magnetic variations $\Delta\boldsymbol{B}$ (r= a, θ, λ, t) on the Earth's surface (r=a), where a, θ, and λ, denote the Earth's radius, latitude, and longitude, respectively. One of the early debates in this discipline was whether one could determine uniquely $\boldsymbol{J}$ (r, θ, λ, t) on the basis of observed variations $\Delta\boldsymbol{B}$ (r= a, θ, λ, t) at a number of places on the Earth's surface. As physicists, Birkeland and Alfvén attempted to determine a unique three-dimensional current configuration. Meanwhile Chapman, as a mathematical physicist, limited his study to a mathematical equivalent current on a spherical shell concentric to the Earth, avoiding the uniqueness issue.

3.1 The Three-Dimensional Current System

3.1.1 The Uniqueness Problem

The nonuniqueness of the solution to the problem of determining the distribution of $\boldsymbol{J}$ was recognized by both Birkeland and Chapman independently. Chapman (1935) noted:

> *It is, of course, in principle, impossible to infer uniquely, purely from observations of a magnetic field (of external origin) at the earth's surface, the location of the current system which is the source of the field.*

It is interesting to note that by facing this nonuniqueness problem, Birkeland and Chapman took contrasting approaches. Being a physicist, Birkeland (1908) attempted to construct a three-dimensional current system by stating:

> *If we assume, as from a physical point of view we might legitimately do, that the current is of a cosmic nature, and consists of negatively and positively charged corpuscles, the trajectories of the separate corpuscles must, as already stated, more or less approximately, follow the magnetic lines of force, moving in spirals about them.*

On the other hand, as a mathematical physicist, Chapman limited himself to a two-dimensional equivalent current system on a spherical shell. Chapman (1935) wrote:

> *The current distribution over a spherical sheet can easily be represented by a diagram using any projection of the sphere upon a plane. This is one method of representing the potential of the field graphically.*

Figure 3.1 shows schematically the equatorial view and the polar view of the current systems considered by Birkeland and Chapman.

3.1.2 Chapman's Equivalent Current System

As a mathematical physicist, Chapman considered that the magnitude of the storm field $|\Delta \boldsymbol{B}|$ = D consists of two components, the component independent of longitude or local time (Dst) and the component dependent on longitude or local time (DS). Both depend on the storm time t,

$$D(\theta, \lambda, t) = Dst(\theta, t) + DS\,(\theta, \lambda, t).$$

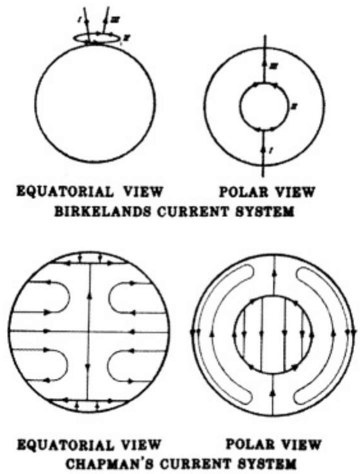


Figure 3.1 Birkeland's 3-D current system and Chapman's 2-D equivalent current system, both equatorial and polar views: S. Chapman (1935).

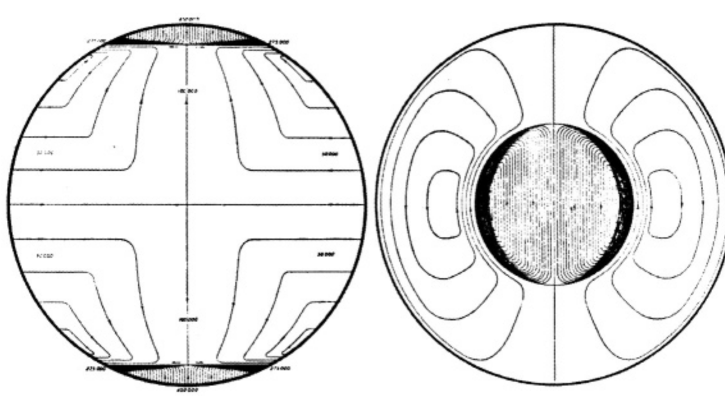


Figure 3.2 Details of Chapman's 2-D equivalent current system for the DS component, the equatorial (left) and polar (right) views for the DS component: S. Chapman (1935).

The equivalent current for the Dst component is a zonal current on a spherical surface, since it does not depend on longitude. It can be considered as equivalent to the ring current. The ring current index Dst originates from this definition. The equivalent current for the DS component has a more complicated distribution (Figure 3.2). The polar view of the DS component shows two concentrated currents of equal intensity, but of opposite directions, along the classical auroral zone. Chapman coined the terms the *eastward electrojet* for the current in the evening sector, and the *westward electrojet* for the current in the morning sector. Each electrojet has a return current in the polar cap and the lower latitude. Note that this analysis of D is basically Fourier analysis for a given latitude θ; Dst is the first constant term and DS is in sinusoidal terms, namely $DS(\lambda) = a_1 \sin\lambda + a_2 \sin 2\lambda \ldots$ for a given θ.Unfortunately, after the ionosphere was discovered, Chapman's equivalent current, or more physically the two-dimensional current, was accepted as the real ionospheric current (H.C. Silsbee and Harry Vestine, 1942; Naoshi Fukushima, 1953), becoming the leading paradigm in the study of magnetospheric current systems. Thus, field-aligned currents were not considered by most researchers for a few decades.

Chapman told me that he thought that there were an infinite number of possible current systems for a given distribution of magnetic disturbance fields observed on the ground, choosing just one arbitrarily did not make sense. Instead, he thought that he could remain accurate so long as he dealt with the equivalent (two-dimensional) current system. Although Chapman had many deep insights into physical processes, he tended to become an applied mathematician when he encountered mathematical uniqueness issues. Mathematical rigor was his life, and it was part of the reason for his friction with Hannes Alfvén, who tended to be intuitive in interpreting physical phenomena. Later, Chapman told me that he went a little too far in avoiding the non-uniqueness.

Chapman's DS component has a return current from each electrojet in the polar cap, constituting two current cells in the polar ionosphere, one located in the morning sector and the other in the evening sector. Ian Axford and Colin Hines (1961) thought that Chapman's two current cells are the Hall current and thus that they are a manifestation of large-scale convective (***E***x***B***) motions of plasma in the magnetosphere. In the E region of the ionosphere, the Hall current arises from the flow of electrons along the streamlines of the convection flow; positive ions cannot participate in the (***E***x***B***) convective motion because of friction with the neutral component. They suggested that various polar upper-atmospheric phenomena could be understood in terms of manifestations of the convective motion (Figure 3.3a). They concluded their paper by stating:

> *We are led to believe that the convective system is of major importance to these phenomena, and we expect it to provide a new basis on which theories of detail may, in time, be based.*

During the last three decades, a number of methods were developed to observe, directly or indirectly, the convective motion of magnetospheric and ionospheric plasmas by incoherent scatter radars, chemical releases, electric field measurements by satellites, and balloons. As Axford and Hines predicted, the convection of magnetospheric plasma has become one of the most important processes and paradigms in our present understanding of magnetospheric processes as a whole.

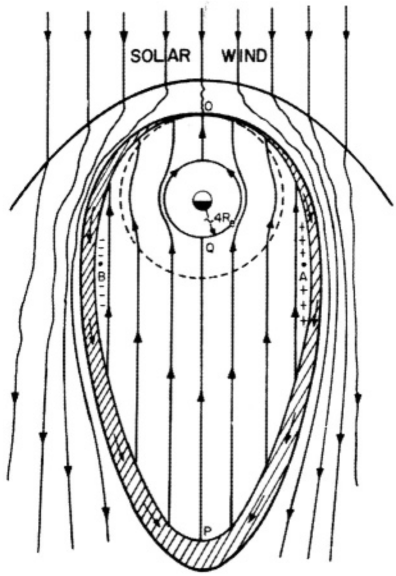


Figure 3.3a The (***E***x***B***) convective motion of magnetospheric plasma on the equatorial plane inferred by Ian Axford and Colin Hines who projected Chapman two-dimensional flow pattern onto the equatorial plane of the magnetosphere. They suggested that a viscous-like interaction between the solar wind plasma and the magnetospheric plasma drives the convective motion. It is more likely that the electric field generated by the solar wind-magnetosphere dynamo process is the direct source: W.I. Axford and C.O. Hines (1961).

Thus, Chapman's equivalent current system contributed to magnetospheric physics in this interesting way, akin to how we learned earlier that Störmer's study of motions of electrons from the Sun is not applicable to magnetosphere formation, but has helped us understand motions of trapped particles in the radiation belt and the ring current belt.

It is now possible to observe the convection on a real-time basis, by a network of HF radars, the SuperDARN network. Figure 3.3b shows an example of the convection pattern about 10 minutes after the dayside reconnection (Ray Greenwald, 1999). It remains to be seen whether or not the corresponding convection exists on the equatorial plane. It is also important to examine what the SuperDARN network really detects.

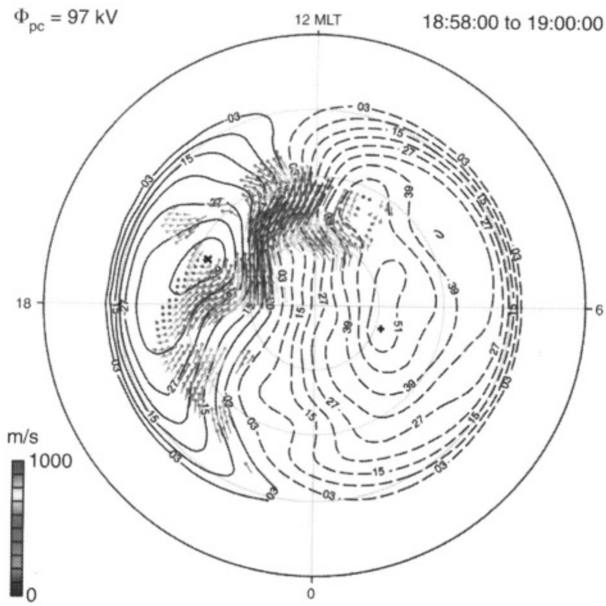


Figure 3.3b The high-latitude electrical potential pattern showing the global response to a significant increase in dayside reconnection on 13 November 1996. Color bars show speed of plasma controlled by the electrical potential pattern: R.A. Greenwald (2001).

3.1.3 Birkeland-Alfvén Model

Most researchers took Chapman's equivalent current system as the true current system for many decades. Meanwhile, Hannes Alfvén (1950) demonstrated with a wire model and a search coil that the local time-dependent part of the geomagnetic storm field (DS) defined by Chapman can be reproduced by a combination of field-aligned currents, the auroral electrojet and the connecting equatorial currents. However, this work did not receive the attention it deserved. It was unfortunate that Alfvén could not attract the attention of the scientific community to the merit of his three-dimensional model against the two-dimensional equivalent current model.

Koji Kawasaki, Ching Meng, and I decided to examine whether the observed distribution of magnetic disturbance vectors can be reproduced by a model three-dimensional current system, which was developed by Alfvén and modeled by C.B. Kirkpatrick (1952). Kirkpatrick's three-dimensional model is not too much different from the currently accepted one. It was a great surprise to us that Kirkpatrick's model reproduced the observations very well (compare Figure 3.4 and 3.5). When I showed my results to Alfvén during one of my visits to Stockholm, he was almost irritated. He said that I was too slow to recognize the validity of his three-dimensional current system. I could well understand his impatience.

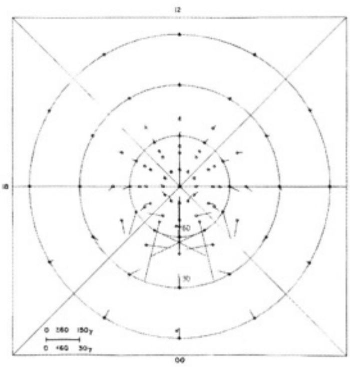


Figure 3.4 Distribution of the magnetic disturbance vectors for the Kirkpatrick-Alfvén model: K. Kawasaki and S.-I. Akasofu (1971).

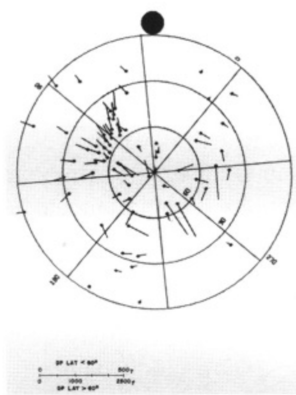


Figure 3.5 The distribution of the observed magnetic disturbance vectors at 1400 UT, December 16, 1964. The vector distribution may be compared with that in Figure 3.4: S.-I. Akasofu and C.-I. Meng (1969).

Incidentally, I object to the use of the term Birkeland current for the field-aligned currents in magnetospheric physics, because Birkeland's currents are far from what we define as the field-aligned currents today. The field-aligned currents flow between the magnetosphere and the ionosphere as a result of the magnetosphere-ionosphere coupling, and they are not *extraterrestrial* currents from the Sun, as Birkeland thought. I also object to statements that imply that Chapman was wrong in rejecting Alfvén's paper on magnetic storms. Note that neither Birkeland nor Alfvén could conceive of the magnetosphere in the way we envision it today. In their first paper on the formation of what we now call the magnetosphere, Chapman and Ferraro (1931) obtained an equation similar to the Debye length and concluded that the solar gas flow must be treated as what we now call *plasma.* Birkeland, Störmer, and Alfvén treated the solar gas as if it were composed of solitary particles. Alfvén's magnetosphere is quite different from what we know today.

Since this point is so basic in the Chapman-Ferraro theory, it was difficult for Chapman to entertain Alfvén's theory, in which protons and electrons in the stream are semi-independent. This point became one of the most serious controversies of the 1960s. During many international conferences, Alfvén (1954) confronted Chapman and emphasized the importance of the electric field in the solar plasma by stating, "The solar plasma flow must satisfy $\boldsymbol{E} + \boldsymbol{V} \times \boldsymbol{B} = 0$." Chapman thought that Alfvén's theory was not appropriate for dealing with the solar plasma as Alfvén ignored the strong electrostatic coupling between protons and electrons in the solar plasma. Chapman was correct in emphasizing that the solar gas be treated as plasma, but his theory could not account for the solar wind-magnetosphere coupling because his plasma was diamagnetic. Alfvén was correct in introducing the concept of the interplanetary magnetic field $\boldsymbol{B}$ and electric field $\boldsymbol{E} = -\boldsymbol{V} \times \boldsymbol{B}$, but failed to treat the solar gas as plasma. Both Chapman and Alfvén were emphasizing the importance of two different electric fields in the solar wind. This should be the way by which science develops constructively, even though controversies tend to flare up from time to time, although it was unfortunate that Chapman and Alfvén could not communicate better.

I am not trying to criticize the monumental pioneering work of the Scandinavian school of Birkeland, Störmer, and Alfvén. What I emphasize here is that we must be cautious in carelessly commenting on the works by our great pioneers. We must give credit where it belongs. We have to be careful also in assigning the nomenclature with the name of the appropriate scientist.

I used to tell my students and colleagues that when two groups have very different views on the same phenomenon, it is very likely that they are observing two different aspects on the same phenomenon. I used to use a pencil to explain such a situation. One end of a pencil is sharp and hard, while the other end, the eraser, is round and soft. Each of the two groups observes only one end, and thus a controversy erupts. Eventually, a third group finds that the two groups are looking at the different ends of the same object, a pencil. An irony may be that the third group gets most of the credit, at least in some cases.

One lesson here is that when a serious controversy erupts, there may be a way to integrate two controversial views and make an epoch-making advance. That is to say, in such a situation, the problem may not be that one is right and the other is wrong, but how to integrate two seemingly contradicting views (see Epilogue).

3.2 Alaska Meridian Chain of Magnetic Observatories

The early study of the current systems in the polar region suffered from insufficient magnetic records, since there were very few possible locations for the observatories in high latitudes. As a first step to improve the situation, I equipped the entire Alaska meridian chain of all-sky camera stations with magnetometers. Like the meridian chain of all-sky cameras, the meridian chain of magnetometers scans the polar magnetic variations once a day. This was the first attempt in the history of geomagnetism to obtain geomagnetic data systematically as a function of latitude. The College observatory, operated by Jack Townshend of the USGS, was the key observatory in this operation.

When the Alaska meridian chain of magnetometers became operational, I was very surprised that we could obtain a fairly systematic magnetic vector distribution over the entire polar region by averaging the data for only a few months (Figure 3.6a). Yosuke Kamide from the University of Tokyo joined my group at that time as a post-doctoral fellow. I suggested to him that there may be a way to obtain the true current system, not the equivalent current system, using such a systematic data set by knowing that the currents can flow only in the ionosphere and along the geomagnetic field lines. After moving to the National Center for Atmospheric Research, in Boulder, Colorado, he developed a computer algorithm with Art Richmond and Sadami Matsushita for this purpose. This code is called the KRM code and has been most useful in studying the high-latitude current configuration.

The early study of the current systems in the polar region suffered from insufficient magnetic records, since there were very few possible locations for the observatories in high latitudes. As a first step to improve the situation, I equipped all the Alaska meridian chain of all-sky camera stations with magnetometers. Like the meridian chain of all-sky cameras, the meridian chain of magnetometers scans the polar magnetic variations once a day. This was the first attempt in the history of geomagnetism to obtain geomagnetic data systematically as a function of latitude. The College observatory, operated by Jack Townshend of the USGS, was the key observatory in this operation.

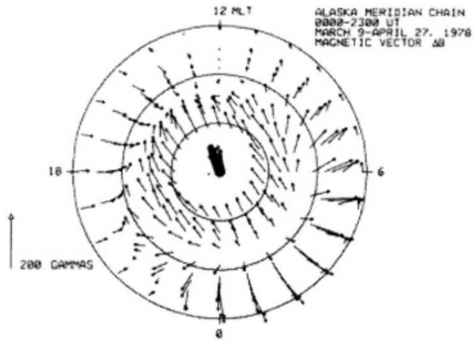


Figure 3.6a

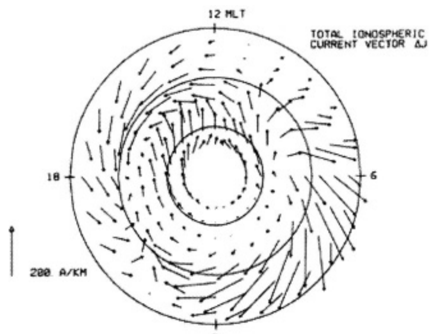


Figure 3.6b

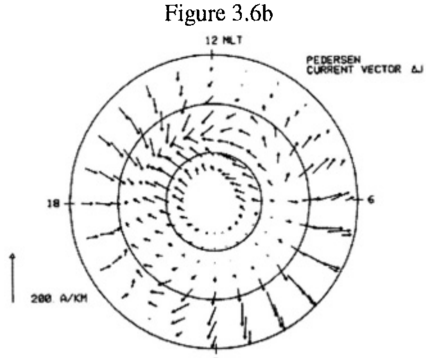


Figure 3.6c

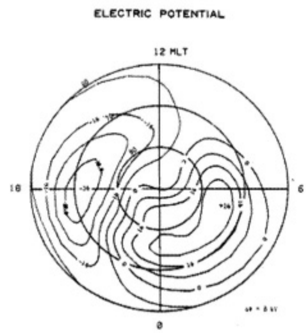


Figure 3.6d

Figure 3.6a, b, c, d (a) Hourly average horizontal magnetic disturbance vectors obtained by the Alaska meridian chain between March 9 and April 27, 1978; (b) The total current vectors are obtained from Figure 3.6a by applying the KRM algorithm; (c) The Pedersen component of Figure 36b; (d) The potential distribution obtained from Figure 3.6a: S.-I. Akasofu and Y. Kamide (1985).

The KRM code was successfully applied to the data obtained from the Alaska meridian chain, although it is necessary to assume the distribution of the conductivity for both the Pedersen and Hall currents. Byung-Ho Ahn, one of my graduate students, developed an excellent conductivity model for this purpose. Thus, we could obtain the daily average current pattern over the entire polar region. The pattern of both the westward electrojet and the eastward electrojet is clearly elucidated (Figure 3.6b). Figure 3.6c shows the distribution of the Pedersen component. Figure 3.6d shows the distribution of the electric potential. All the results are self-consistent.

3.3 The IMS Meridian Chains of Observatories

When I was operating the Alaska meridian chain of observatories in the 1970s, Gordon Rostoker was also establishing a meridian chain of magnetometers in Canada. We agreed to operate the Inuvik, Northwest Territories, station jointly. We flew to Inuvik by small plane to install a magnetometer there. Our operation of the meridian chains inspired our colleagues in Europe and Russia to establish four other chains during the International Magnetosphere Study (IMS). Thus, six meridian chains of magnetometers, consisting of 71 magnetometer stations, became operational during the IMS (Figure 3.7). The KRM computer code has further been developed to deal with instantaneous current patterns based on the simultaneous magnetic records from all six IMS meridian chains of magnetometers. Thus, both the data from the six meridian chains and the powerful KRM code enabled Kamide and his colleagues (1982) to study the development of the three-dimensional substorm current system with a time resolution of 5 minutes. A great wealth of knowledge on the ionospheric currents, electric fields, potential field, field-aligned currents, and the Joule heating rate was obtained over the entire polar region (Figure 3.8). This was a great joint effort by many of our colleagues, which made the dream of our pioneers a reality after more than half a century. They were Yosuke Kamide, Byung-Ho Ahn, Wolfgang Baumjohann, Eigil Friss-Christensen, Herb Kroehl, H. Maurer, Art Richmond, Gordon Rostoker, R.W. Spiro, John Walker, and Alex Zaitzev.

3.4 Boström's Current Loops

In spite of such progress, we encountered a serious problem in confirming the validity of our results. An independent method may be found by satellite observations. However, most satellite-based observations of electric current $\boldsymbol{J}$ must be based on measurements of $\Delta \boldsymbol{B}$ (r, θ, λ) at single points and at particular times. Thus, except for a rather simple geometrical configuration of

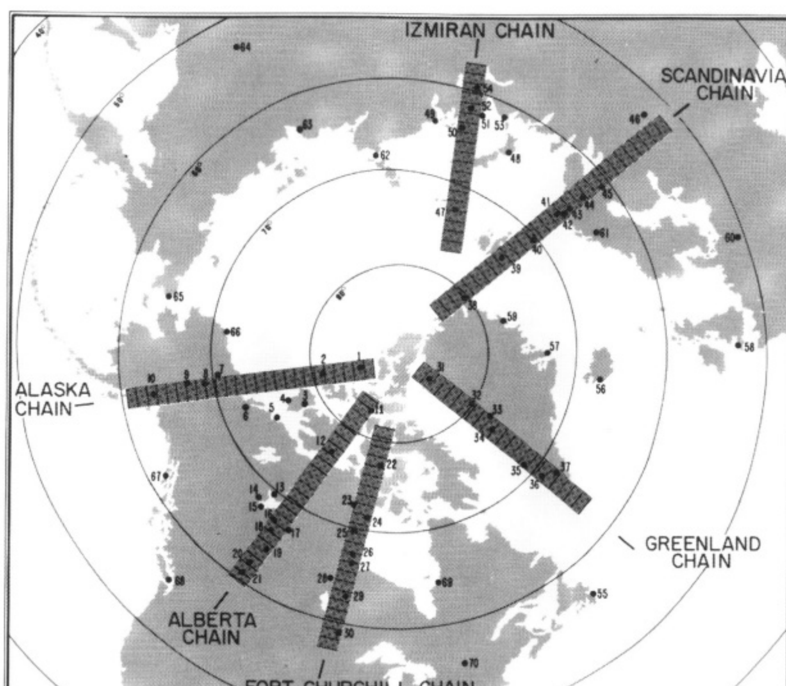


Figure 3.7 Six IMS meridian chains of magnetometers: B.-H. Ahn, Y. Kamide, and S.-I. Akasofu (1984).

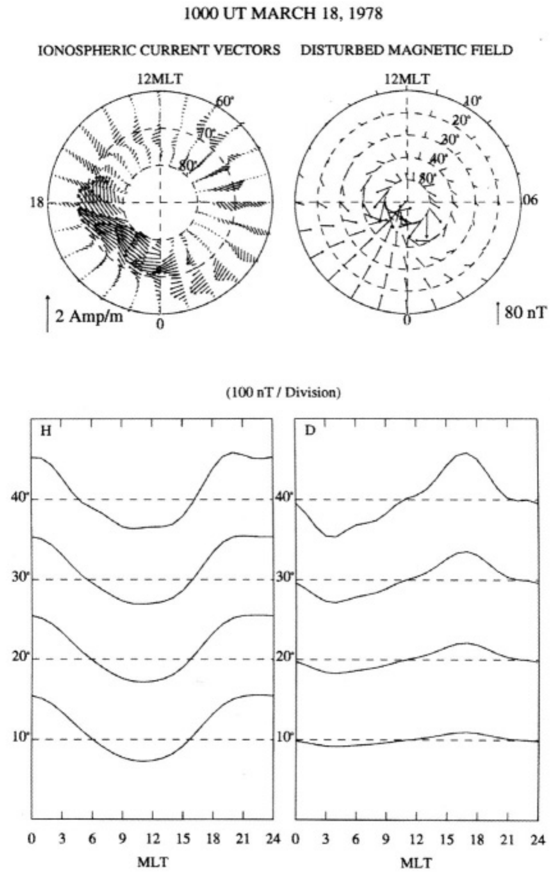


Figure 3.8 Example of the results obtained by the KRM method on magnetic data that was obtained from the IMS meridian chains: S.-I. Akasofu, W. Sun and B.-H. Ahn (2002).

the field-aligned currents just above the ionosphere, it is not possible to determine $\nabla \times \Delta \boldsymbol{B} = \boldsymbol{J}$. This fact requires that it takes at least one year of data to determine the average distribution of currents on the entire polar region or the equatorial plane by a satellite (Iijima et al., 1990). For this very reason, it is not possible to obtain the distribution of electric currents in the magnetosphere at a particular time, $\boldsymbol{J}$ (r, θ, λ, t) by a satellite or two. Thus, our problem is that neither ground-based nor satellite-based observation would allow us to confirm each other's results.

After much thought, an interesting idea emerged based on one theoretical insight by Rolf Boström (1964). As early as 1964, Boström had suggested that ionospheric currents and equatorial currents are connected by two types of field-aligned loop currents, the meridional loops and the azimuthal loops (Figure 3.9). Thus, one way to test both the ground-based and satellite-based results, as well as Boström's loops, is to project the *average* ionospheric currents onto the equatorial plane and compare the results with the satellite results obtained by Takeshi Iijima and his colleagues (1990). The left-hand side of Figure 3.10 shows the projected *average* Pedersen component of the ionospheric current onto the equatorial plane (the insert shows the geometry of a meridian loop). Note that the ionospheric Pedersen current during substorms is basically the north-south current (Type II in Figure 3.9). This projected distribution may be compared with the distribution of the satellite-based average radial currents on the equatorial plane on the right-hand side of Figure 3.10 (Iijima et al., 1990). I was greatly surprised by this unexpected agreement between the two.

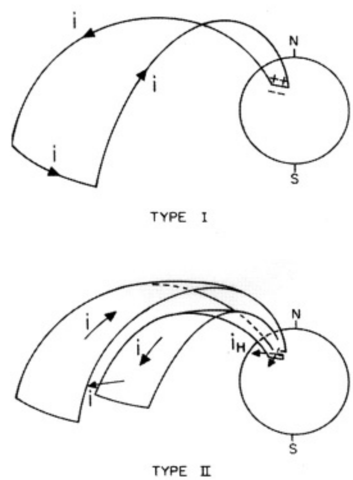


Figure 3.9 Two types of the magnetospheric current system proposed by Boström: R. Boström (1964).

In spite of the fact that both results are obtained by entirely different methods and at different times, the agreement shown in Figure 3.10 is quite satisfactory, testifying that both the KRM method and satellite observation can obtain the *average* ionospheric and equatorial currents reasonably well. Further, the results indicate the general validity of Boström's meridional loop currents. Unfortunately, it took several more years to identify the azimuthal loops. This was because satellites measure both the ring current and the cross-tail current (which do not close in the ionosphere), in addition to the azimuthal loop currents. Here, we had to rely on modeling efforts to extract the azimuthal loop currents from the satellite observations (Wei Sun et al., 1996, 2000).

One of my colleagues had never trusted our ground-based study of inferring the current system. It was his strong opinion that satellite measurements ($\nabla \times \Delta \boldsymbol{B} = \boldsymbol{J}$) are the only way (although satellites do not carry a current meter!). Thus, I asked him to compare the two diagrams in Figure 3.10. After this comparison, I have not heard him remark on our ground-based study of the magnetospheric currents.

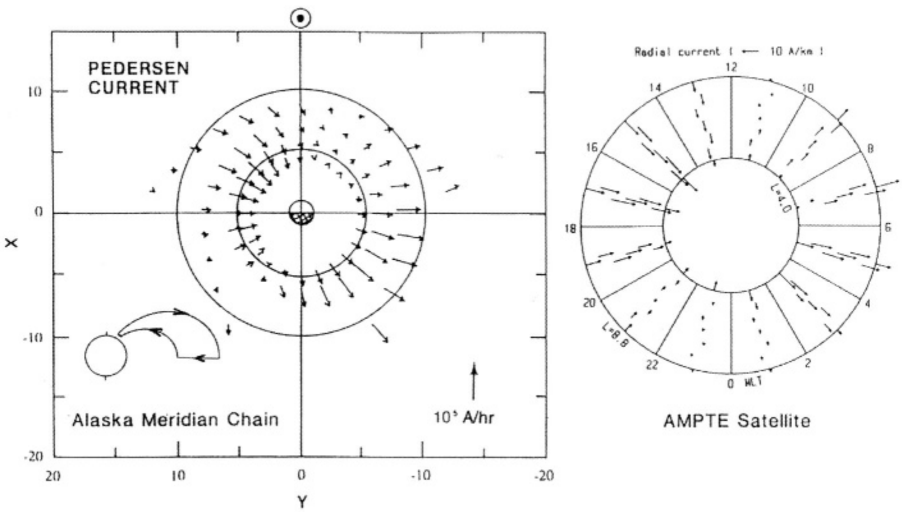


Figure 3.10 Left: Distribution of the projection of the Pedersen currents (Figure 3.6c) on the equatorial plane by considering Boström's Type II current. Right: Distribution of the radial currents on the equatorial plane deduced from satellite observatories: S.-I. Akasofu (1992).

Our success has encouraged us to extend our method to infer the distribution of substorm electric currents on the equatorial plane with a high time resolution by projecting the ionospheric currents using Boström's two elementary loops, since the distribution of the ionospheric currents can be obtained with a time resolution of 5 minutes. There is no way to achieve this without hundreds of satellites.

Figure 3.11a shows the distribution of ionospheric and magnetospheric currents from 11:30 to 12:10 UT on March 19, 1978; see the AE index in the insert. It is a substorm event. The top row shows the distribution of ionospheric currents; magnetic records from 71 high-latitude observatories were used for this particular analysis. As can be seen from the AE index in the insert, the current pattern at 11:30 UT shows the distribution just prior to the onset, at 11:35 UT, of a substorm. One can see a dramatic increase of the westward current in the late evening sector at 11:40 UT, five minutes after the onset, while the corresponding increase of the eastward current was less prominent. The substorm reached its maximum epoch at 12:10 UT.

The second row shows the corresponding distribution of the field-aligned currents (upward in red, downward in blue). One can see that the tip of the westward current is associated with a pair of downward- and upward-directed current areas; the former is located poleward of the latter. The third row shows the equatorial current distribution that is obtained by projecting the ionospheric currents onto the equatorial plane. One can see a major change of the distribution in the late evening sector at substorm onset.

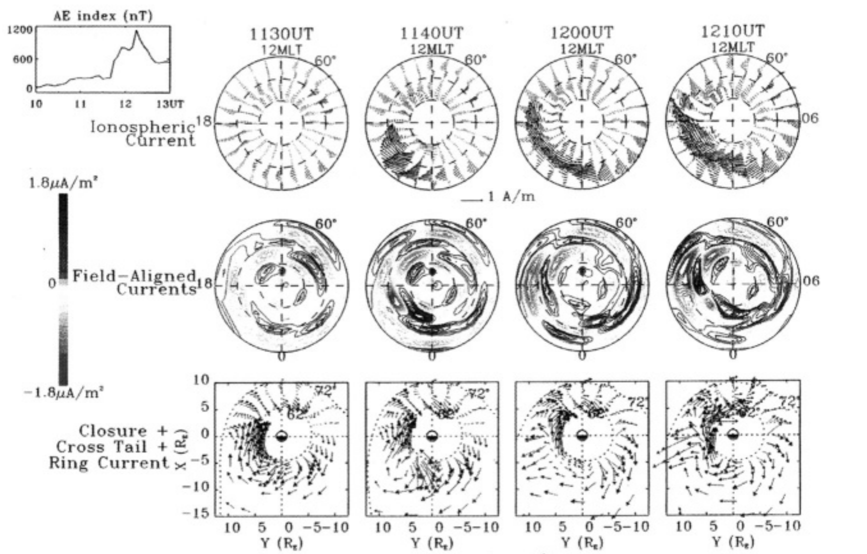


Figure 3.11a Distribution of the ionospheric currents (blue arrows for westward currents and red arrows for eastward currents), the field-aligned currents positive (blue) and negative (red) values, corresponding to downward and upward field-aligned currents, and the inferred currents on the equatorial plane during a substorm: S.-I. Akasofu and W. Sun (2000).

It was unfortunate that there was no satellite to measure $\Delta \boldsymbol{B}$ on the equatorial plane at that particular instant. On the other hand, since satellite measurements cannot provide $\boldsymbol{J}$, but instead $\Delta \boldsymbol{B}$, we can examine the validity of our results by comparing the well-established observation result of $\Delta \boldsymbol{B}$ and the computed $\Delta \boldsymbol{B}$ based on our results. Figure 3.11b shows the results. One can see that the stretched field lines contract and become dipolar after substorm onset at 11:30 UT. Thus, the inferred currents on the equatorial plane can reproduce the known change of $\Delta \boldsymbol{B}$. This phenomenon is often referred to as *dipolarization* and is a common feature at substorm onset, so that our results are consistent with the satellite observations.

Examining changes of the equatorial currents shown in Figure 3.11a, one can see that the dipolarization resulted from the current that has a strong eastward component, which counters the westward cross-tail current, which grew during the growth phase (10:30-11:30 UT). In fact, the total current direction had an eastward component at 11:40 UT. Magnetic reconnection is not needed as the cause of the dipolarization in this case.

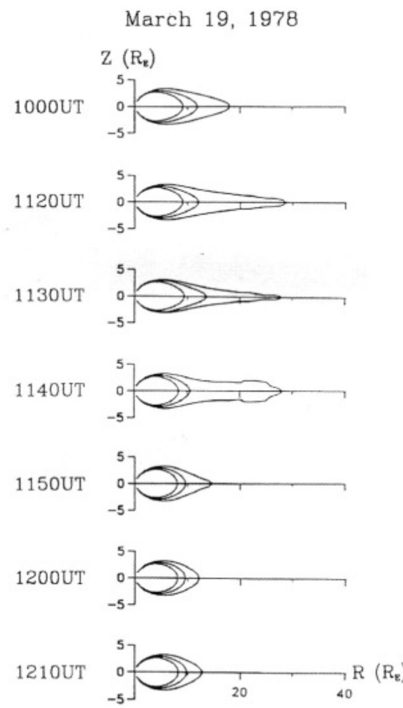


Figure 3.11b Computed configuration of the magnetic field lines at the time of the substorm in Figure 3.11a: S.-I. Akasofu and W. Sun (2000).

It should be noted that the so-called "dipolarization" is not necessarily the recovery of the stretched field lines to become the dipole field lines. The field magnitude of the dipolarization often far exceed the magnitude of the dipole field. At the geosynchronous distance, the dipole field is about 100 nT, but the observed field can be as large as 150 nT (see Figure 3.11c); considering that the electron pressure is comparable with $B^2/8\pi$ and that the tail current can reduce the dipole field from 110 nT to 80-90 nT, this large value of 150 nT is surprising. Thus, the dipolarization is not a matter of diverting the westward tail current (namely, the current wedge concept), but an eastward current of Boström Type I current must grow. The ionosphere must thus be the driving force for such an eastward current.

This fact indicates that the dipolarization is not just reducing or diverting the cross-tail current, indicating the growth of currents against the normal tail current. The ionosphere must be playing an active role in the dipolarization and perhaps even the over-dipolarization. Thus, the ionosphere actively participates in substorm processes, rather than passively reacting to magnetospheric processes. As mentioned in Section 1.10, regardless how interesting the magnetotail is in terms of plasma physics, the magnetotail is simply the "tail" of the ionosphere (dog). Substorm onset will be discussed in Section 3.6.

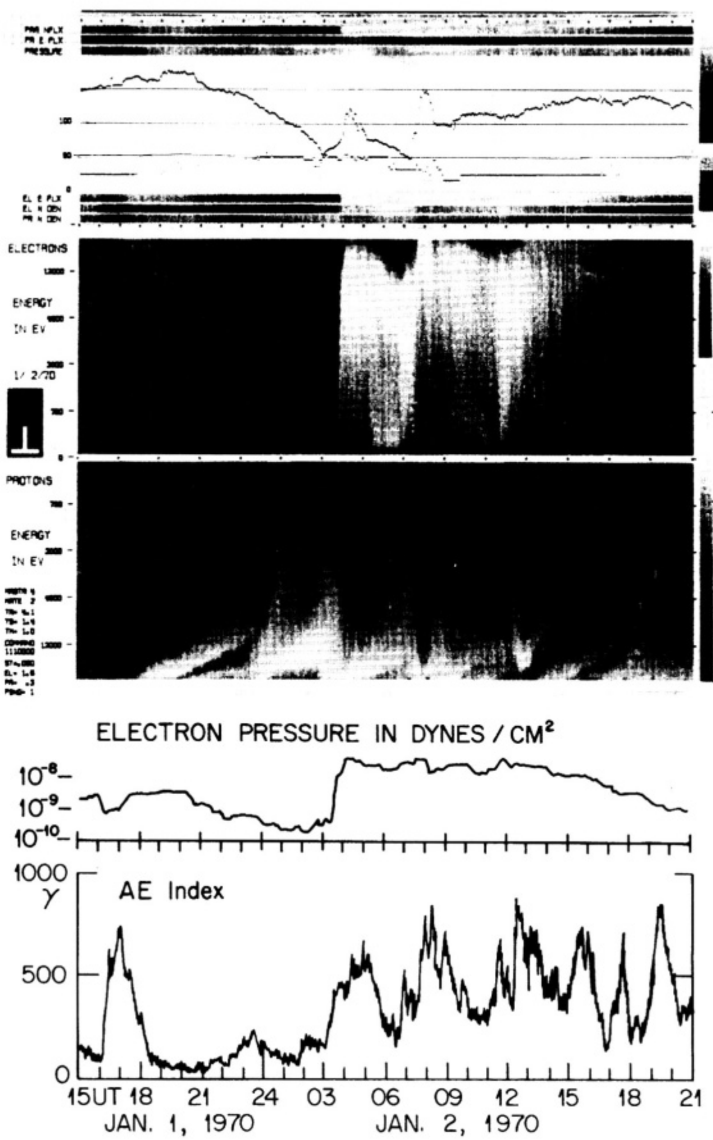


Figure 3.11c The dipolarization and injection of electrons and protons, observed at the geosynchronous satellite ATS-5 on January 2, 1970. From the top, the H component magnetic variations, the electron and proton spectrograms, plasma pressure and the AE index: S.E. DeForrest and C.E. McIlwain (1971).

It is in this way that the ground-based observation $\Delta \boldsymbol{B}$, together with satellite observations and modeling efforts, has finally enabled us to determine $\boldsymbol{J}$ (r, θ, λ, t), realizing the dream of our pioneers, although our results are still only the first approximation.

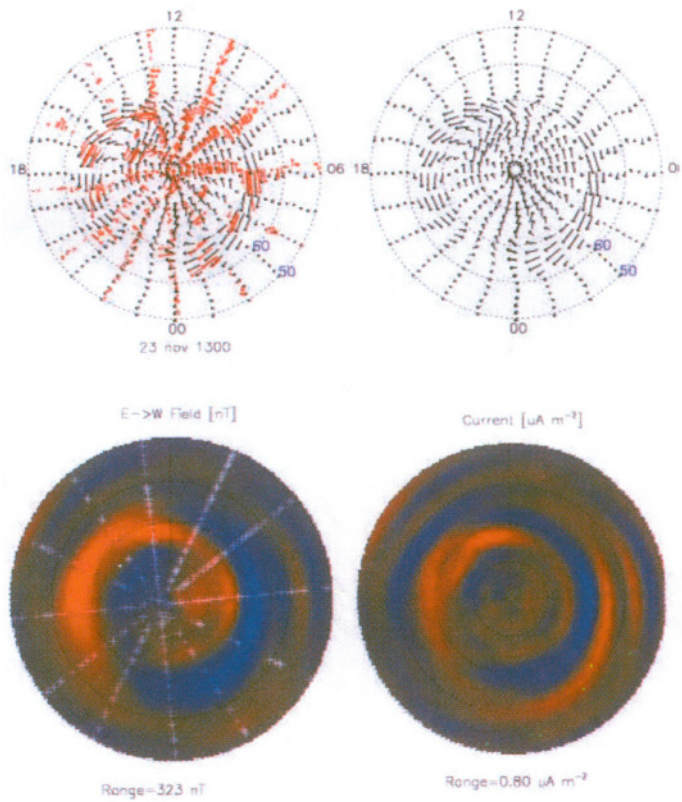


Figure 3.12 The distribution of $\Delta \boldsymbol{B}$ and of field-aligned currents determined by the Iridium satellites: C.L. Waters, B.J. Anderson, and K. Liou (2001).

Ideally, hundreds of satellites are needed to obtain $\Delta \boldsymbol{B}$ (r, θ, λ, t) in space around the Earth. It so happened that the Iridium satellite communication system launched 66 satellites that carried magnetometers. The system can determine the distribution of $\Delta \boldsymbol{B}$ and of the field-aligned currents on a few-hour average basis. Figure 3.12 may be compared with Figure 3.11a, which was obtained by the Alaska meridian chain (Section 3.2).

3.5 Nikolsky's Spiral

This is one of the early stories of the study of the current distribution in the polar region. The concept of the auroral *zone* had greatly influenced the study of geomagnetic disturbances; the auroral zone is a circular belt in the geomagnetic coordinate system, centered around the geomagnetic pole. The SD current system obtained by Chapman (1935) was an example of this influence. He suggested that the auroral currents consist of a pair of concentrated currents along the auroral *zone* (not the auroral *oval,* which was not known then). There were the westward auroral electrojet in the morning sector and the eastward electrojet in the afternoon sector, and their return currents in the polar cap and in lower latitudes (Figure 3.2). It was thought that a magnetic observatory rotates under such a fixed current system, registering the daily magnetic variations. Under the eastward electrojet (in the afternoon sector), there occurs positive (poleward) magnetic disturbances in the horizontal (H) component, while the westward electrojet produces negative (equatorward) magnetic disturbances in the morning sector. The SD current system had become the standard model and thus a major paradigm for a few decades.

However, Nikolsky (1947) found an interesting phenomenon: geomagnetic disturbances recorded at high-latitude stations have three activity peaks during a day. He denoted three peaks: A (afternoon), N (night) and M (morning) (Figure 3.13a). Further, he found that the peak tends to occur earlier in time at higher latitudes for the A and N peaks, later for the M peaks. Thus, in a polar plot, the peak occurrence times for A, N and M delineate three spiral curves. I was fascinated by Nikolsky's results, but had no idea as to how to interpret them. This is particularly because it is not possible to understand his results in terms of the SD current system. However, one day I recognized that the combination of the N and M spirals delineates the auroral oval; the A peak spiral indicates the eastward electrojet. The results suggested to me that the westward electrojet does not stop in the midnight sector (as indicated by the SD current), but continues to flow westward, with the westward-traveling surge along the auroral oval in the evening sector where auroral arcs actually lie. Thus, the westward electrojet is located at latitudes higher than 65-70° in the evening sector, not along the auroral zone. When I reported this result in Moscow in the 1960s, Nikolsky was very happy. He wrapped me in a typical Russian bear hug, saying that I was his son's age.

I thought that my interpretation on Nikolsky’s results was reasonable, because the auroral ionization takes place along the auroral oval, not along the auroral zone. Therefore, there is no reason why the westward electrojet must stop in the midnight sector as the SD current system indicates. Fortunately, the Alaska meridian chain of all-sky camera stations was also equipped with magnetometers. An examination of both all-sky photographs and magnetic records indicated clearly that the westward electrojet extends into the evening sector with the westward-traveling surge. During the passage of westward-traveling surges to the north (Figure 3.13b), an auroral zone station (point A) registered positive changes in the H component, while at a station of gm lat. 70-75° (Point B in Figure 3.14), negative changes with greater magnitudes were observed. Therefore, it became obvious that the westward electrojet flows along the oval, not the auroral zone.

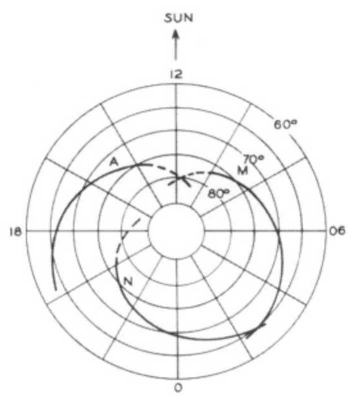


Figure 3.13a Nikolsky's A, N, and M spirals. Both N and M constitute the auroral oval: S.-I. Akasofu and C.-I. Meng (1967).

Thus, Meng and I found that Chapman's SD current in the polar region must be revised. As Figure 3.14b shows, the westward electrojet during substorms forms a single vortex, not a double vortex as the SD current system suggests. As described in Section 1.10, the double vortex pattern is prominent during the growth phase. A strong single vortex pattern appears at substorm onset. It was hard for Chapman to realize that his SD current system was not correct. However, after examining my analysis, he was convinced about its validity, becoming a strong supporter of the revision of his SD current system. On the other hand, it was difficult to convince many of my colleagues of the results in the 1960s and 1970s. I recall that there were even emotional objections to the revision. Again, Yasha Feldstein was one of the strong supporters of my work (Figure 3.15).

Indeed, such a flapping may be one of the important causes of geomagnetic storms, because it is associated with changes of the IMF *B*z component; note that the interplanetary magnetic field lines are expected to be nearly parallel with respect to the current sheet, particularly near the current sheet (Akasofu 1979).

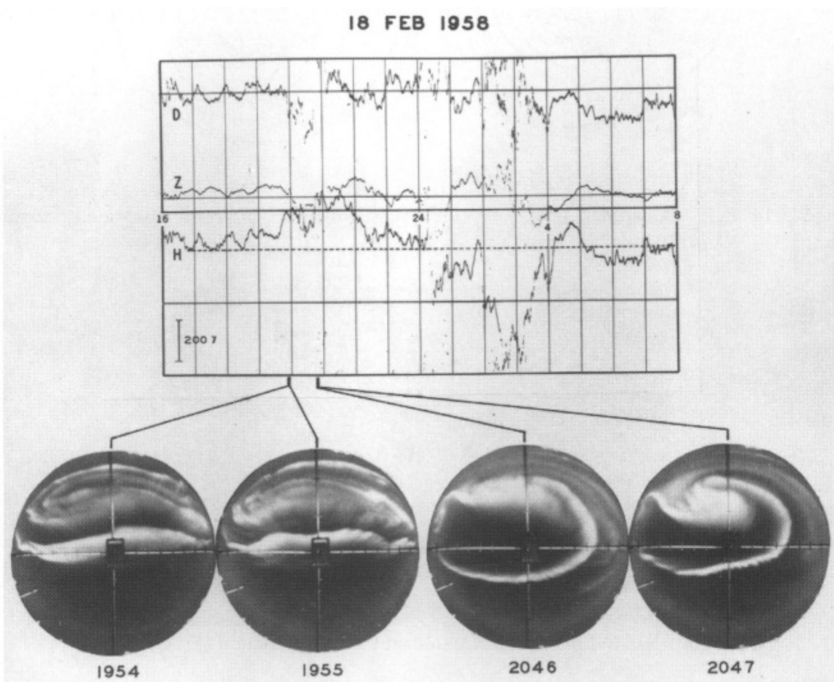


Figure 3.13b College magnetic record and all-sky photographs from Fort Yukon. When positive changes in the H component were observed at College, an intense westward-traveling surge was passing over Fort Yukon; the surge is the leading edge of the westward electrojet: S.-I. Akasofu and C.-I. Meng (1967).

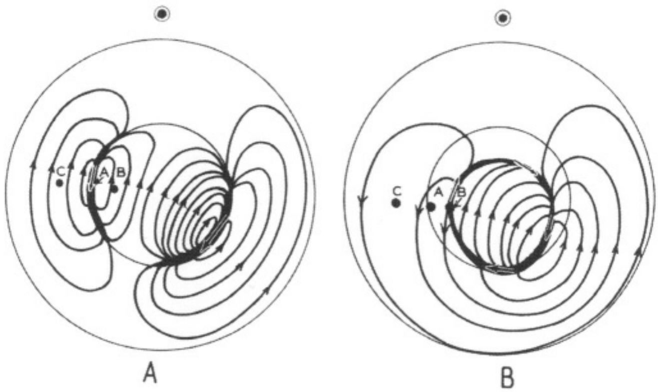


Figure 3.14 Left: Chapman's SD current system. Right: Revised equivalent current system: S.-I. Akasofu and C.-I. Meng (1968).

Figure 3.15 With Yosuke Kamide and Yasha Feldstein (Prague).

3.6 Substorm Onset

When theoretical progress stagnates, it is best to go back to the fundamental observed facts. There are three distinct and well-established phenomena at substorm onset, as well as an enhanced convective flow after the so-called *southward turning* of the IMF or an enhancement of the solar wind-magnetosphere dynamo power ε:

1. A sudden brightening of an auroral arc over a distance of 1000 km at the poleward boundary of the diffuse aurora in the late evening or the midnight sector; the arc is located just the poleward side (~gm lat. 65°) of the diffuse aurora (caused by energetic electrons from the outer radiation belt).

2. The westward electrojet develops suddenly along the brightening arc.
3. The magnetic field structure changes suddenly from a tail-like configuration to the dipolar structure in the magnetotail. This phenomenon is often referred to as dipolarization and propagates outward.

The three phenomena have the following physical meanings:

Sudden Brightening

This phenomenon must be associated with an increase of the energetic electron flux into the existing arc, carrying the upward field-aligned current over an east-west extent of the order of 1000 km within the narrow width of an auroral arc.

Westward Electrojet

The westward electrojet is essentially the Hall current carried by an eastward flow of ionospheric electrons. For this to happen, a southward electric field has to develop along a narrow east-west strip in the ionosphere. An increased ionospheric conductivity caused by the electron precipitation will also enhance the current.

Dipolarization

Prior to substorm onset, the dipolar field lines at five Earth radii and beyond are stretched greatly by an intensification of the cross-tail current. The sudden dipolarization indicates that the cross-tail current is suddenly reduced. In some instances, an over-dipolarization can take place, indicating the reversal of the westward tail current.

Any successful theory of substorm onset must thus explain at least these three processes, which follow after the magnetosphere is driven for 30-40 minutes. The arc, which brightens first at substorm onset, is located just on the poleward side of the diffuse aurora in the late evening or the midnight sector. The diffuse aurora is caused by the precipitation of high-energy electrons from the outer trapping (Van Allen) belt. It is certain that the processes associated with substorm onset must thus be found at the region of transition from a dipolar field regime to the stretched field regime, not deep in the magnetotail. It is likely that this particular region is located at a distance of 5-10 Earth radii (Frank and Sigwarth, 2000), not as far as 20-30 Earth radii.

My scenario of substorm onset: As the power ε increases and drives the magnetosphere-ionosphere coupling system, the two convection cells are intensified. The morning cell advances into the afternoon cell in the late evening sector. It is where the vorticity increases and thus where the upward field-aligned current is generated (Hasegawa and Sato, 1980). This process establishes Boström's Type II current which reduces the cross-tail current, resulting in substorm onset.

In the past, proponents of magnetic reconnection assumed that magnetic reconnection could form at a distance of 5-6 Earth radii by calling it the near-Earth neutral line. Realizing that such a possibility is unlikely and that there is no definitive observational evidence, they put the near-Earth neutral line to as far as 20-30 Earth radii (and still call it near-Earth). Now, they are trying to connect it to a distance of 5-10 Earth radii by an earthward ($\boldsymbol{E}$x$\boldsymbol{B}$) plasma flow that has not been confirmed by observations; most of the observed flows are flows along the magnetic field lines.

I am afraid that it will not be possible to make substantial progress in understanding substorm onset so long as we cling to elusive magnetic reconnection as the primary process. Looking back through the history of magnetic reconnection, I was of the opinion that the original concept of magnetic reconnection in the solar atmosphere was not realistic (see Section 5.6). I doubted that a stable antiparallel field condition could even be set up in the dynamical and turbulent solar atmosphere and that it could be explosively destroyed, although it is a sort of a problem theorists love to deal with, regardless if it is realistic, and thus worthwhile. In fact, theorists assumed the initial condition of the antiparallel field (the so-called "Harris solution"), which may not even exist in the realistic solar atmosphere and in the magnetotail, and soon found that the assumed antiparallel field configuration is hard to destroy explosively. If it were easy to do so, the antiparallel field configuration would not form to begin with as such an initial condition. The only possibility would be that magnetic reconnection could be driven, namely, if two magnetic configurations are forcefully brought together to produce an antiparallel configuration. In such a situation, what magnetic reconnection could generate in terms of energy may be the same amount of energy that is needed in driving magnetic reconnection, not energy resulting from annihilation of the original antiparallel field.

As the concept of magnetic reconnection has become explosively popular, some of us have been left as an almost invisible minority. A fanatic believer in magnetic reconnection told me that I am not qualified to be a magnetospheric physicist unless I believe in such a fundamental process as explosive magnetic reconnection. On the other hand, Tony Lui and I were one of the small groups who very seriously attempted to find definitive indications of magnetic reconnection as the source of substorm energy in satellite data during the 1970s and 1980s. We failed to find it. A series of our papers on this subject was titled *Search for the Magnetic Neutral Line in the near-Earth Plasma Sheet.* Indeed, to date, no one has conclusively found the line within about ten Earth radii. I am still not convinced that magnetic reconnection is the primary cause for substorms as the energy supply process; if any takes place, it may occur as a secondary process. Magnetic reconnection will be discussed further in Section 5.6 in connection with solar activities.

3.7 Changes of Magnetic Energy in the Magnetotail

A proponent of the magnetic reconnection hypothesis has shown that the magnitude of the magnetic field *B* in the magnetotail decreases sharply at substorm onset. However, this is not necessarily the case. Figures 3.16 shows that the ε function, magnetic energy density ($B_T^2/8\pi$), and AE index had similar time variations during two successive substorms. Wolfgang Baumjohann (1996) showed also that there is no distinct change in $B_T^2/8\pi$ for isolated substorms.

Actually, a better parameter to examine changes of magnetic energy in the magnetotail is the size of the open region, which should be approximately proportional to magnetic flux in the magnetotail, instead of the field magnitude *B* at a single point in the magnetotail.

Here, the open region is defined as the highest latitude region that is free from auroral electrons except for the *polar rain* that consists of the high-energy tail of electrons in the solar wind. (The term "polar rain" was coined by Walter Heikkila of the University of Texas, where snow does not fall!)

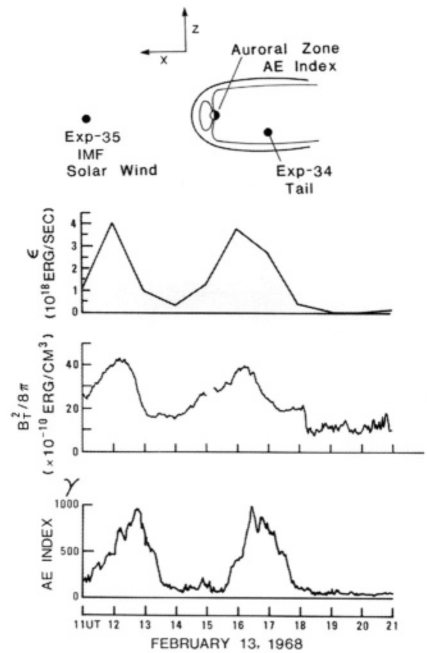


Figure 3.16 Changes of ε, $B_T^2/8\pi$ and the AE index for two successive substorms: S.-I. Akasofu (1980).

Figure 3.17 is an example in which the dimension of the polar cap thus defined was monitored by both the noon-midnight and evening-morning sector satellites for a four-day period. The auroral oval (the belt of visible auroras) is located in the cross-hatched region, while the belt of soft electrons (<500eV) lies in the dot-shaded region. During the first 10 hours or so, until about 10:00 U.T. on 30 October 1984, the substorm activity was subsiding, and the noon-midnight and the evening-morning dimensions of the polar cap were decreasing. This resulted from an increase in the width of the precipitation belt of soft electrons (<500eV) during this period, except in the night sector. These soft electrons must be a trapped population; otherwise, they will escape into interplanetary space in less than one second. The second substorm activity began at about 10:00 UT on 31 October. Both the noon-midnight and evening-morning dimensions increased. Substorm activity intensified at about 06:00 U.T. on November 1, and lasted until about the end of that day.

The open region expanded during this substorm activity. First, the dayside precipitation (the cusp) shifted equatorward, as much as 10° as reported by a large number of researchers. On the night side, there was a poleward shift of the precipitation boundary and a large expansion of the hard electron precipitation region, indicating a large poleward expanding and bulge. After 10:00 UT on November 1, the width of the soft precipitation belt began to increase rapidly, although the hard precipitation region remained in the same latitude range; the AE index shows that the substorm activity was subsiding at that time. The open region contracted gradually until about 08:00 UT on November 2, but began to expand rapidly afterwards. This expansion was associated with the last substorm activity. Figure 3.18 shows also changes of the open flux, estimated on the basis of the size of area surrounded by the oval during a substorm. There is no indication of the decrease of the flux at substorm onset observed by a satellite (Frank et al., 1998).

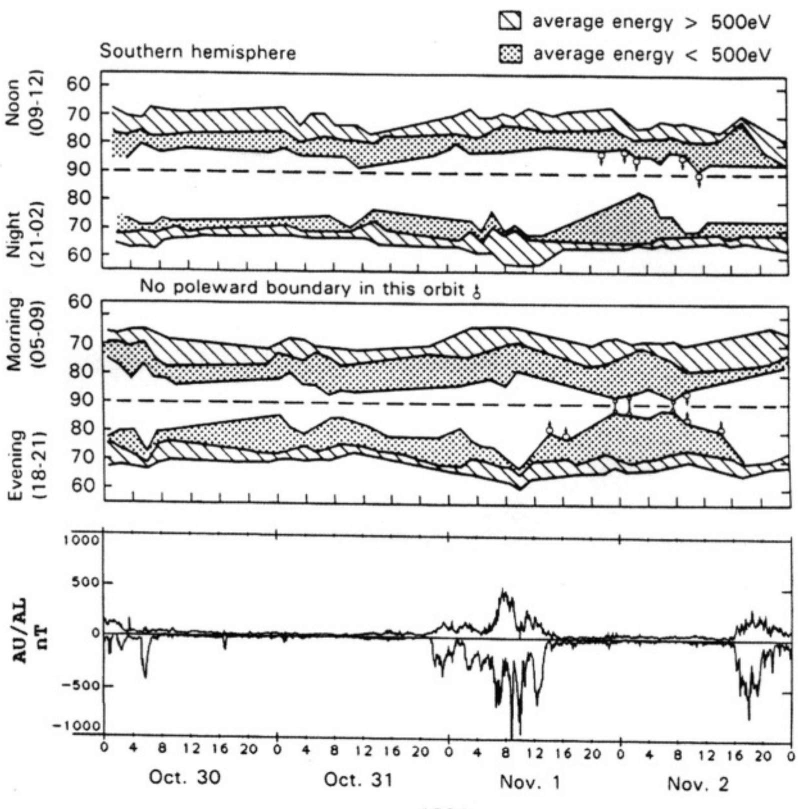


Figure 3.17 Changes of the size of the polar cap, the precipitation belt of auroral electrons of low and high energies (the noon-midnight and the evening-morning orbits) and the AE index: S.-I. Akasofu, C.-I. Meng,, and K. Makita (1992).

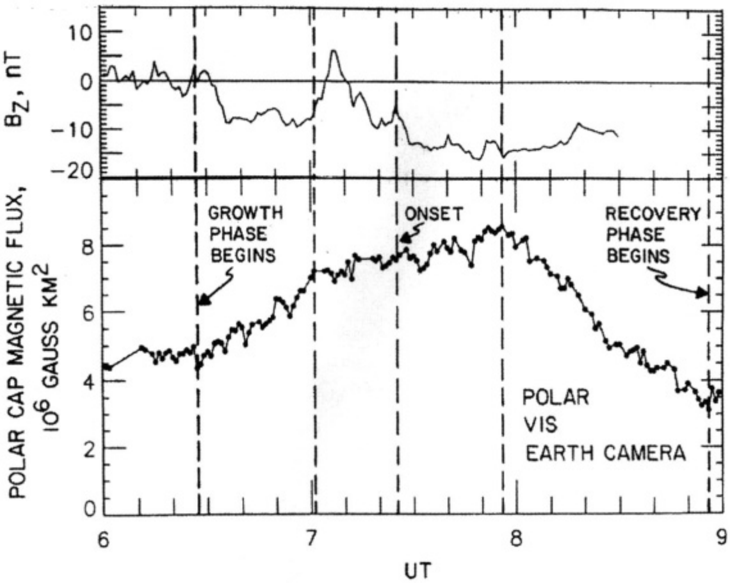


Figure 3.18 Upper: The IMF *B*z component. Lower: The magnetic flux in the polar cap (modified): L.A. Frank, J.B. Sigwarth, and W.R. Paterson (1998).

This and many other examples show that the dimension of the open region varies roughly in harmony with the AE index. Both the noon-midnight and evening-morning dimensions of the open region are substantially larger during substorm activity than those during a quiet period, by a factor of about two to three. Thus, the open magnetic flux is at least four times greater during substorm activity than those during a quiet period. Actually, the expansion begins soon after the IMF *B*z component becomes negative and the contraction proceeds gradually well after the IMF *B*z component becomes positive.

Thus, there is more magnetic flux in the magnetotail during substorms than during a quiet period. This situation is illustrated in Figure 1.13c; if the input is large enough, the water level of the pitcher can be above the upper spout. If the sudden conversion were solely responsible for the expansive phase, the dimension of the open region should decrease suddenly at onset. This is not necessarily the case. The fact that the dimension of the open region is greater during an active period than during a quiet period shows that the magnetosphere is highly driven throughout substorm activity. We shall see in Section 5.5 that magnetic energy in an active region of the Sun tends to increase at flare onset. Such observational facts mean nothing to believers in magnetic reconnection.

3.8 Storm-Substorm Relationship

A geomagnetic storm tends to occur in association with a series of magnetospheric substorms (Figures 1.8a and 3.19). It is quite clear that an intense geomagnetic storm tends to develop when intense substorms occur rapidly in succession. This is because the injection rate of ring current particles must overcome a heavy loss rate in the magnetosphere. I suggested in 1968 that substorms are elements of a geomagnetic storm and that each substorm contributes to the formation of the ring current belt by injecting ring current particles. Section 2.5 described how our study of the great geomagnetic storm of February 11, 1958 led us to this conclusion.

One of the most interesting findings in this connection in recent years is that O^+ ions, instead of protons from the solar wind, are often the dominant ions in the storm-time ring current belt (Yannis Daglis, 1997). Since oxygen atoms in the solar wind are highly ionized, say O^{+7}, ring current ions O^+ must be of ionospheric origin.

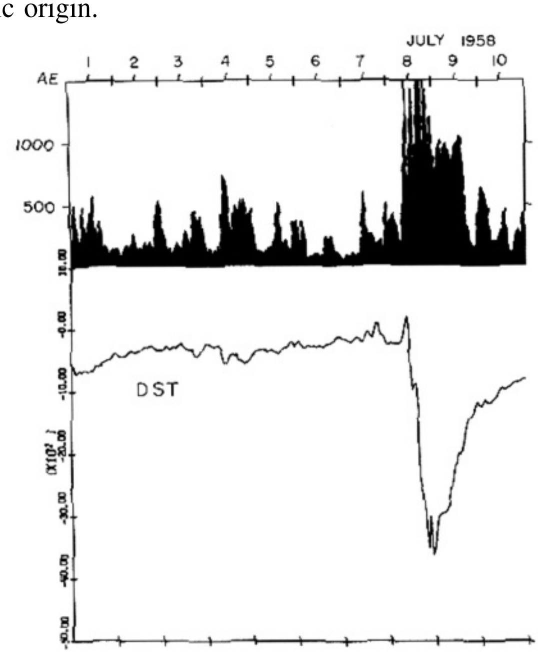


Figure 3.19 The relationship between substorms (indicated by the AE index) and a storm (indicated by the Dst index): S.-I. Akasofu (1970).

Indeed, energetic O^+ beams are observed to stream out from the auroral ionosphere. Hence, it is reasonable to think that the development of the ring current belt is caused by O^+ ions from the ionosphere associated with the upward field-aligned currents during substorms. Further, it appears that they are accelerated along the geomagnetic field lines from the ionosphere into the plasma sheet and are subsequently injected into the trapped region during substorms after being convected toward the Earth.

It appears that a number of researchers question this simple role of substorms in the formation of the ring current, in spite of the fact that Daglis (1997) and others showed that the flux of O^+ ions in the ring current correlates well with the AE index. Further, a recent result showed conclusively that the flux of O^+ ions in the magnetotail increases at substorm onset. Most geomagnetic storms are associated with substorms. The formation of the ring current belt is a two-stage process; the first is the injection of O^+ from the ionosphere into the magnetotail, and the second is the injection of O^+ from the magnetotail to the ring current belt. Thus, there can be a variety of complication before O^+ ions can form the belt under a heavy loss caused by the charge exchange process. For example, unless the injection from the ionosphere to the magnetotail can occur relatively close to the ring current belt, O^+ ions may not be able to form the ring current belt, they may drift way from the magnetosphere. The injection from the magnetotail to the ring current belt has to last long enough (note also that a strong convection alone cannot produce an intense ring current belt and an intense main phase without O^+ ions resulting from prior substorm activity). Therefore, we have need to examine why different substorms contribute differently to the ring current formation, rather than deny the storm-substorm relationship, as some suggested.

As exciting new development in this particular subject is a possibility to study visually the formation of the ring current belt by observing energetic neutral atoms (ENA) by the High Energy Neutral Atom (HENA) images. An example is shown in Figure 3.20.

3.9 Geomagnetic Indices

Many researchers attempt to use the geomagnetic indices in *quantitative* studies. Unfortunately, serious mistakes can arise as many authors attempt to find the relationship between substorms and a geomagnetic storm by examining the relationship between the Dst index and the AE/AL index without knowing their accuracy. The Dst and AE indices were devised in the early 1960s for individual geomagnetic storms (Akasofu and Chapman, 1961; Neil Davis and Masahisa Sugiura, 1966) and are only a very rough measure of the ring current intensity and of the electrojet activity, respectively.

If we base our quantitative study on the present AE and Dst indices, our results will be greatly limited by the accuracy of these indices. In any scientific field (or in economics) an index is a very rough measure intended to show a trend. One can make a serious mistake without knowing what the index indicates, how it is derived, and how rough it is. To begin with, *both indices are not really the physical quantities* we seek. They are not the total ring current intensity and the total electrojet current intensity. Further, for example, as shown in Figures 3.4 and 3.5, the substorm current system causes a positive change in the H component in low latitudes. This positive change can reduce the magnitude of the Dst index. Therefore, unless this positive effect can be removed, one could conclude that substorms reduce the ring current intensity. Therefore, the present AE and Dst indices should be calibrated before the storm-substorm relationship can be studied quantitatively.

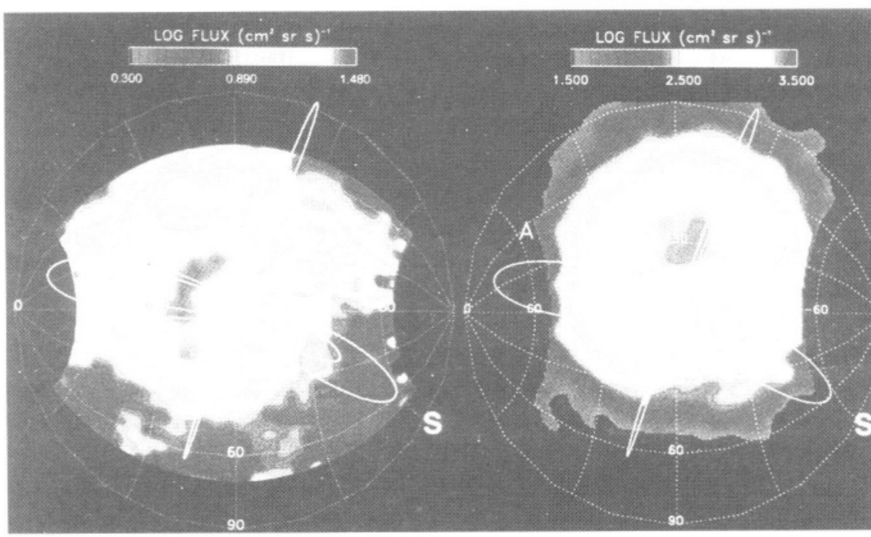


Figure 3.20 MENA image at 8.6 keV (left) and HENA image at 16-27 keV (right) during a magnetospheric substorm on June 10, 2000, at 1100 UT. The white circle in the center of each image represents the Earth. Field lines at L=4 and L=8 are drawn at noon (denoted by S), midnight, dawn, and dusk: J. Burch, J.L. Green, and S.A. Fuselier (2001).

3.10 Publication of Solar-Terrestrial Physics from Oxford University Press

Geomagnetism, published by Chapman and Bartels in 1940, was the classic treatise and served the development of the field of geomagnetism. I still remember the great excitement I felt when I bought it, after long saving money for it; it was a very expensive book for a poor student. In 1968 or so, Chapman and I felt that a new comprehensive treatise was needed because geomagnetism had developed into magnetospheric physics. It was the period when magnetospheric physics was developing rapidly. As a result, the manuscript had to be revised many times. I regretted that I could not complete it before Chapman's untimely death in 1970. Sir Edward Bullard gave me many valuable pieces of advice in completing it. It was finally published under the title *Solar-Terrestrial Physics*, from Oxford University Press, in 1972. I dedicated it to Katherine Chapman. I felt that Chapman would have agreed to do so. Thirty years have already passed after the publication of the book. This means that it was published before many of the present active researchers were born. Although I myself am not very inclined to read any papers published before my birth (the Chapman-Ferraro paper was published in 1931, so I had no choice but to read it!), it is my hope that the young generation might at least flip through it to find that many unsolved problems today were present before their birth. Our book might prevent their rediscovery of well-established facts.

It may be appropriate here to relate how Chapman described his boyhood:

When I was 14 my father wondered what I should become, and he took me first of all to a builder's merchant, who said that plumbing was quite a good trade. I recall a story of an American plumber who had an only daughter of who he was very fond and proud. He sent her to college and she got a degree. And then she got a secretarial job. He said, "Yes, she gets $3,000 a year. Not bad for an educated person, is it?" Perhaps I might have been rich if I'd emigrated to America and became a plumber. But no such thought ever crossed our minds.

And then my father took me to an engineer, a man who with his brother had built up a very successful gas engine manufacturing firm. He had been to Manchester University and taken a degree, and was science-minded. He said he would take me into his works at that age, 14, but he said it would be better if I went for two years to a technical school, and I might even, from there, go on to the University. But at any rate, he said he would take me if I wanted to go.

My father took his advice, and I went to a technical institute a few miles away it was there that I first set eyes on an eminent scientist...

...This chemist–a Scotsman–took a kindly interest in me, and suggested I should sit for a county scholarship examination in hope of going to Manchester University.

...The county of Lancashire in which I lived offered university scholarships.

...In England there's a very good scholarship system, much better now than then. But even then the county of Lancashire in which I lived offered 15 university scholarships. I took the examinations and I was 15th on the list. At any rate, being 15th on the list, I was in. At the age of 16, I went to Manchester University to continue the studies in engineering...

With Emiko Akasofu and Katherine and Sydney Chapman at the University of Michigan campus in 1962. Our Oxford book *Solar Terrestrial Physics* was dedicated to Katherine.

3.11 Summary of Chapters 1, 2, and 3

It may be useful to summarize in a few paragraphs what our generation has learned from the IGY period to the end of the 1980s and what are described in Chapters 1, 2, and 3.

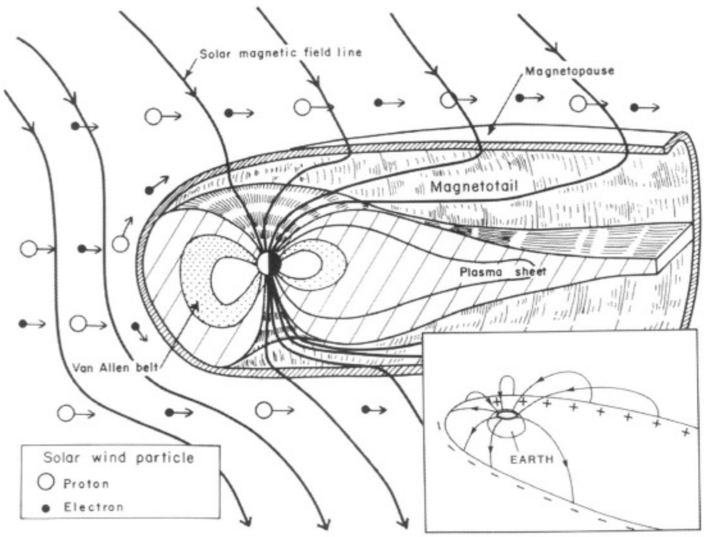
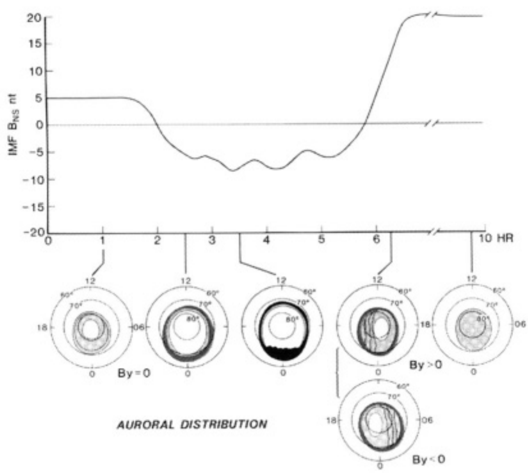
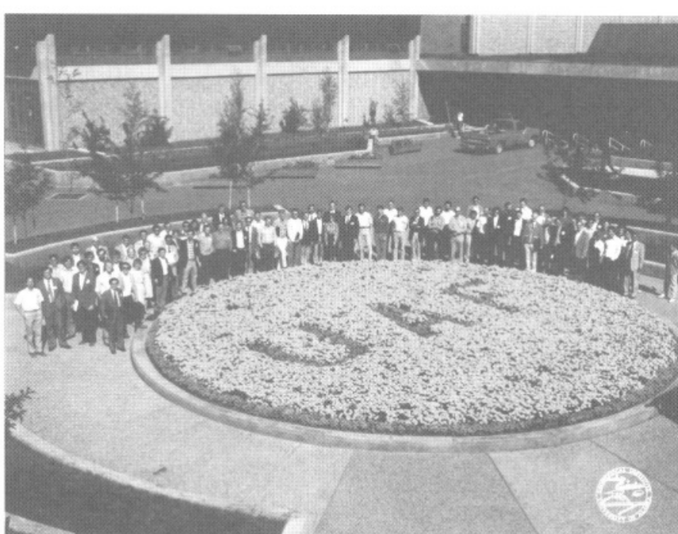


Figure 3.21 The solar wind-magnetosphere dynamo and its primary discharge circuit: S.-I. Akasofu (1992).

The solar wind and the magnetosphere constitute a dynamo (Figure 3.21). Its power is given by $\varepsilon = VB^2 \sin^4(\theta/2)\ell_0^2$. The discharge powered by the dynamo manifests itself in various phenomena, such as geomagnetic disturbances and auroral displays. When the power exceeds 10^{18} erg/sec or so for a few hours, the magnetosphere exhibits a specific response called the magnetospheric substorm. Its magnetic manifestation is the polar magnetic substorm and its auroral manifestation is the auroral substorm. Figure 3.22 shows the resulting response of the aurora including the auroral substorm. The discharge takes place between the magnetosphere and the ionosphere. The upward currents from the ionosphere are carried by downward flowing electrons that cause optical emissions by colliding with atoms and molecules in the ionosphere, which we identify as the aurora. The upward currents are also associated with outflow of O^+ ions. When O^+ ions are injected into the inner magnetosphere, they form the ring current belt. When intense substorms occur frequently, a large number of O^+ ions accumulate in the trapping region and its effect is mainly responsible for a large depression of the Earth's magnetic field in low latitudes. This phenomenon is prominent during the main phase of the geomagnetic storm. It is my hope that the new generation of researchers will advance our understanding of the solar wind-magnetosphere interaction beyond this summary, if necessary by revising it completely.



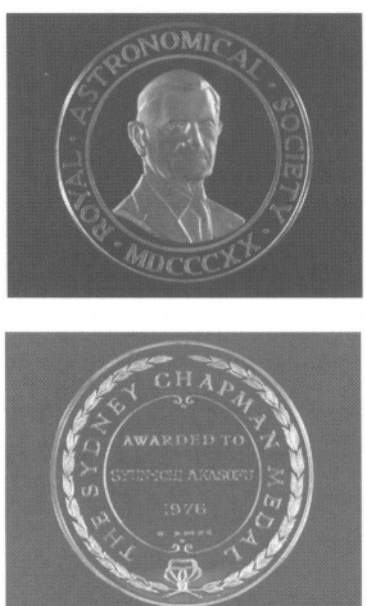
3.22 Schematic illustration to show how the aurora responds to a specific change of the IMF: S.-I. Akasofu (1989).



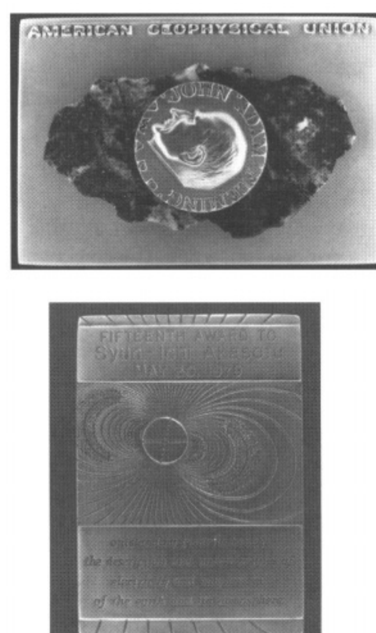
At the Geophysical Institute, we hosted many international conferences. Previous page: AGU Chapman Conference "Magnetospheric Polar Cap," August 6-9, 1984. Above: AGU Chapman Conference "The Formation of Auroral Arcs," July 21-25, 1980: Geophysical Institute, University of Alaska.

The opening ceremony of the International Conference on Substorms (ICS-2) at the University of Alaska Fairbanks. Joe Kan is introducing me as the first speaker, March 7-11, 1994: Geophysical Institute, University of Alaska.

Some of the attendees of the ICS-2, March 7-11, 1994.

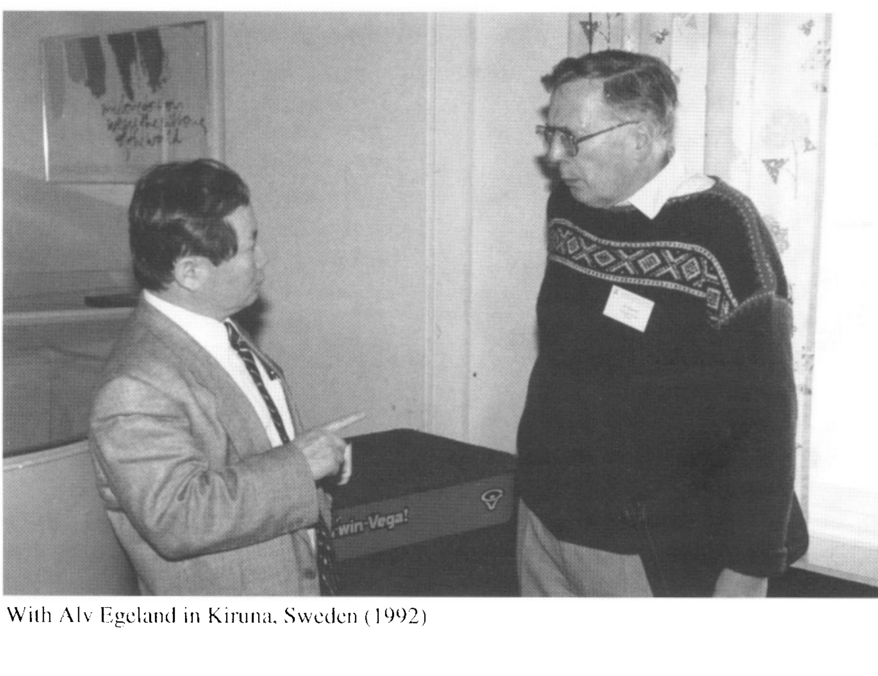


Chapman Medal from the Royal Astronomical Society, London.

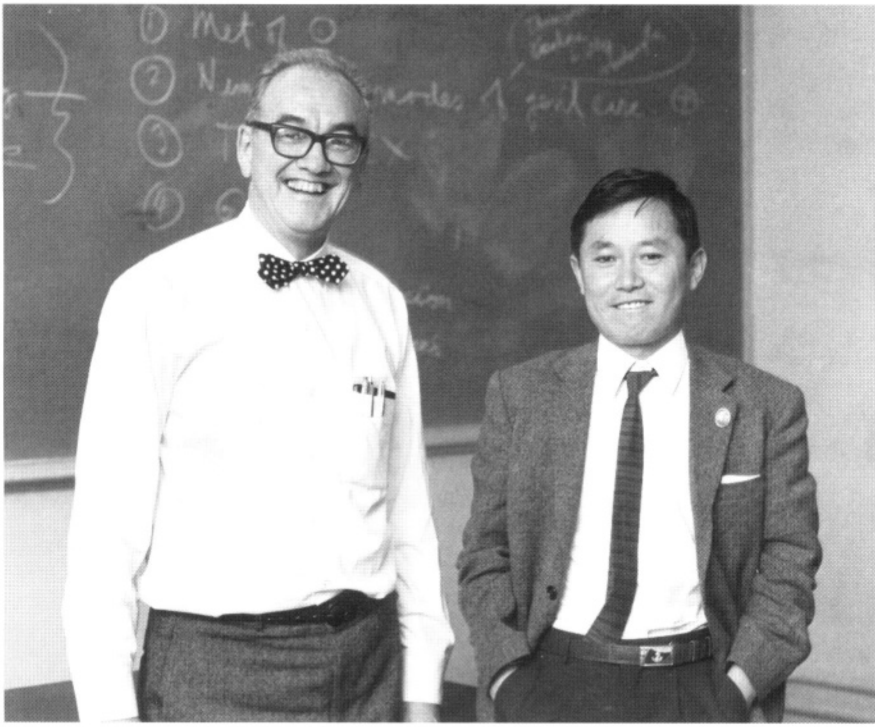


Fleming Medal from the American Geophysical Union.

At the occasion of the opening of the Van Allen Hall, University of Iowa. From left: D. Venkatesan, Harold Taylor, Mary McIlwain, Carl McIlwain, Guido Pizzella, Bruce Randall, James Van Allen Thomas Armstrong, Theodore Fritz, and Syun Akasofu.

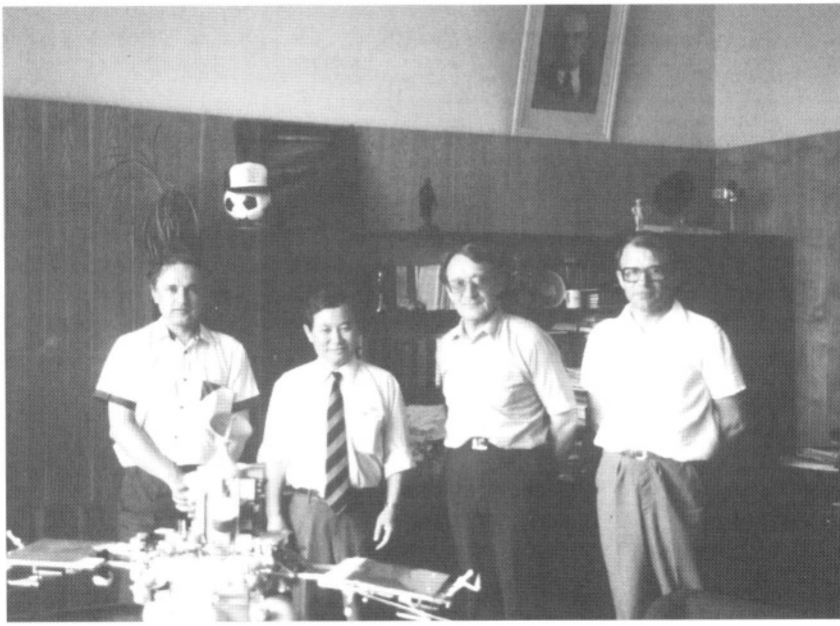


With Alv Egeland in Kiruna, Sweden (1992)

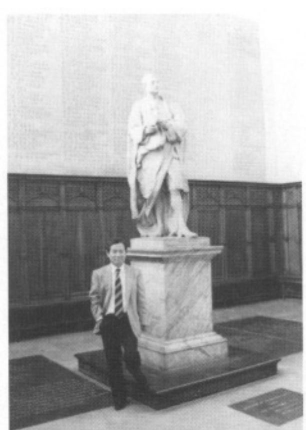


With Walter Orr Roberts (1979)

Kiruna Geophysical Observatory (presently the Swedish Institute of Space Physics).

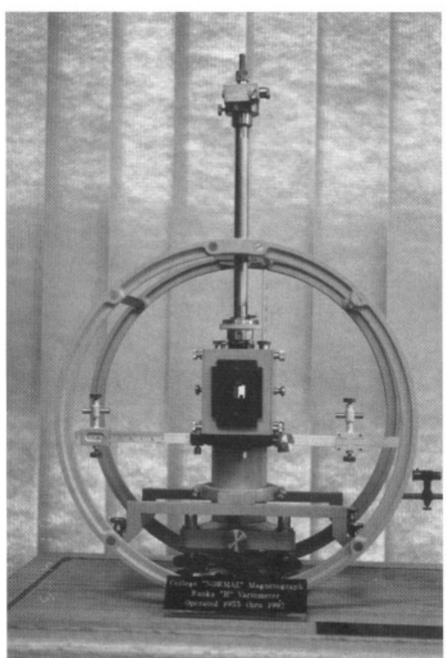


From left: Albert Galeev, Syun-Ichi Akasofu, Renat Zinnurovich Sagdeev, and an unidentified person at the Space Research Institute, Russian Academy of Sciences.



With Newton's statue at Trinity College, Cambridge University (1988), Chapman's alma mater.

With Yosuke Kamide and Tony Lui at a volcano in Hokkaido (1994).



Normal-run magnetometer (the H component) operated at the College magnetic observatory from 1955 to 1992. Presented to me by J. Townshend.

Chapter 4:

IS THE EARTH'S DIPOLE REALLY OFF-CENTERED AND INCLINED?: PLANETARY MAGNETIC FIELDS

4.1 Introduction

Spherical harmonic analysis has been considered the most powerful method for studying planetary magnetic fields. However, it is a mathematical tool, like the Fourier analysis method, and its results must be examined in the light of physics.

Based on the spherical harmonic analysis method, it has long been believed that the Earth's main dipole is off-centered by 0.08 Earth radii and is inclined by 11.5° with respect to the rotation axis (cf. Chapman and Bartels, 1940). These deviations have not received much attention in the past.

However, we have to face them now, because the spherical harmonic analysis method shows that the main dipole of Uranus is off-centered by 0.3 Uranus radii and is inclined by 60° with respect to the rotation axis; Neptune's dipole is off-centered by as much as 0.55 Neptune radii and is inclined by 47°. The problem is that since all the dynamo theories of planetary magnetism rely on planetary rotation, it is unlikely that they can explain easily how the main dipole of the planet can be greatly off-centered or inclined as inferred. Thus, how can we understand the greatly off-centered or inclined main dipole of Uranus and Neptune?

In this chapter I intend to show that it is possible to learn a great deal about planetary magnetism from the solar magnetic fields. (Note that both solar and planetary magnetisms rely on the same dynamo theory.) For this purpose, I ask the reader to consider solar magnetic fields on an imagery spherical surface of 2.5 solar radii over the Sun. This surface is called the *source surface*. In spite of the great complexity of the photospheric magnetic field, the field at a distance of two to three solar radii is much simpler and is approximately dipolar. In the upper and lower parts of the left-hand side of Figure 4.1, the magnetic equator on the source surface inferred from the photospheric magnetic field during Carrington rotation 1720 (the Carrington rotation is a sort of solar day number) is shown both in a regular rectangular map and on a source (spherical) surface. In this particular case, the first spherical harmonic term provides an inclined dipole at the center of the Sun; the equator is an inclined circle on the source surface and its rectangular projection is a sinusoidal curve (the right-hand side of Figure 4.1). In this situation, the common practice is to state that the solar dipole for Carrington rotation 1720 is inclined from the rotation axis by about 45°.

However, if the dynamo theories demand that the dipole axis should be parallel or antiparallel with respect to the rotation axis in this example, additional dipoles are required to reproduce the observed equator (the right-hand side of Figure 4.2). So long as such a combined field of the dipoles can reproduce the observed field (expressed mathematically by the spherical harmonic analysis method), such an inference is at least a possibility. Although it is difficult to determine uniquely the characteristics of the additional dipole, our interpretation may be physically more meaningful than the spherical harmonic analysis results.

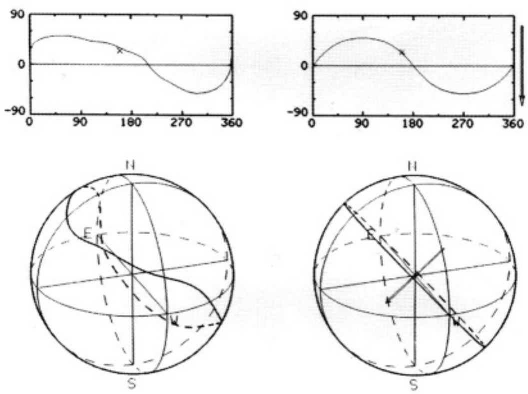


Figure 4.1 Left: Observed magnetic equator on the source surface. Right: Result of the spherical harmonic analysis: S.-I. Akasofu.

4.2 Rotation of the Solar Magnetic Dipolar Field on the Source Surface

A number of researchers have shown that the magnetic fields on the source surface can be approximated by a dipole field and that the polarity (towards/away from the sun) of the source surface field is fairly well correlated with that of the interplanetary magnetic field (IMF) observed near the Earth.

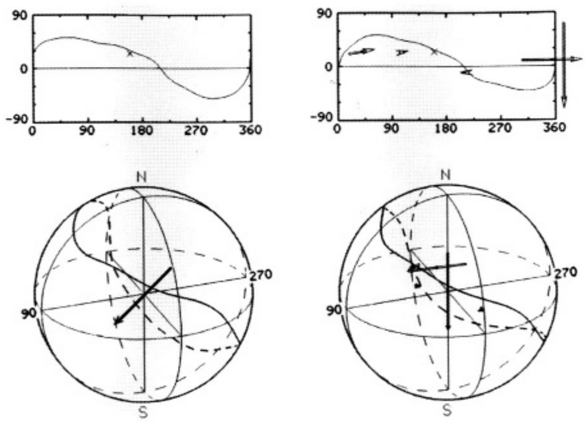


Figure 4.2 Left: Magnetic equator with the central dipole determined by the spherical harmonic analysis. Right: The computed magnetic equator for the central dipole (assumed to be parallel or antiparallel) to the rotation axis, together with auxiliary dipoles on the photosphere: S.-I. Akasofu.

It is for this reason that this particular surface is called the *source surface*. Further, Todd Hoeksema and P.H. Scherrer (1984) and Takao Saito (1987) examined sunspot cycle variations of the magnetic equator (which is usually referred to as the neutral line) on the source surface and demonstrated that the neutral line varied fairly systematically during Sunspot Cycle 21 and earlier cycles.

The neutral line lies near the ecliptic plane at the beginning of the cycle and tilts gradually as the cycle advances, standing almost vertically (with respect to the equatorial plane) during the maximum epoch of the cycle. The neutral line tilts further during the declining epoch of the cycle and lies near the ecliptic plane at the end of the cycle. If one approximates the magnetic field on the source surface by a central dipole, this change can be represented by a gradual rotation of the dipole by 180°, so it changes from pointing northward to pointing southward. Figure 4.3 shows the dipole and the neutral line at different epochs during Sunspot Cycle 21.

On the other hand, it is important to realize that there is no indication that the main dipole of the *photospheric* magnetic field shifts in such a way to produce the rotation from one polarity to the other and also the rotation of the neutral line on the source surface. Indeed, it is well known that the so-called *unipolar region* is located near each pole throughout the sunspot cycle. It is believed that the reversal of the dipolar field occurs as a result of the *migration* of a large-scale unipolar field (say, positive) from low latitudes to the polar region, canceling the pre-existing unipolar (negative) field there. Meanwhile, a cancellation of the opposite polarity occurs in the opposite hemisphere at about the same time. Thus, the reversal of the polarity does not involve a gradual shift of the dipole pole from one hemisphere to the other across the equator. These observations show that one must look for other causes of the rotation of the dipole field on the source surface, namely causes that do not rely on the rotation of the main dipolar field on the photosphere. It is difficult to comprehend why researchers in planetary magnetism do not accept such important solar information. It appears that their theorists decide that the planetary dipole should rotate and try to find a way for it to do so, regardless if such an effort is realistic or worthwhile.

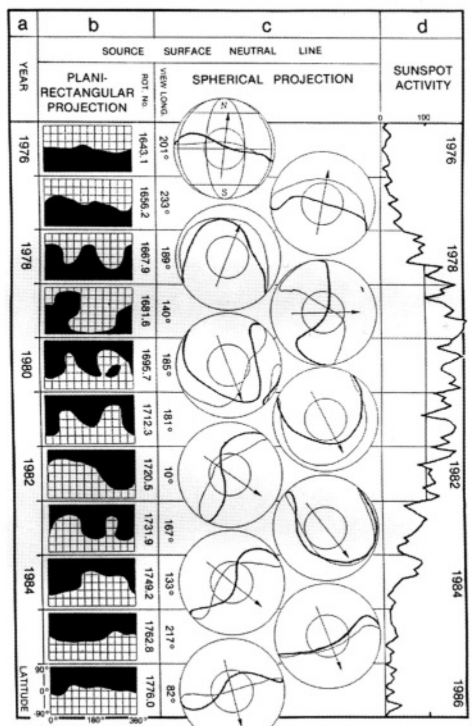


Figure 4.3 Rotation of the assumed central dipole and the observed magnetic equator during Sunspot Cycle 21, which peaked in about 1979-1980: T. Saito, T. Oki, S.-I. Akasofu, and C. Olmsted (1989).

As a solution to this puzzle, Takao Saito and I suggested that the major changes of the neutral line on the source surface during a sunspot cycle can be well represented by a combined effect of changes of an axial dipole located at the center of the Sun and of two nearly antipodal (equivalent) dipoles near the equatorial plane on the photosphere. Specifically, the observed variations of the neutral line during Sunspot Cycle 21 can be expressed by assuming that:

1. The magnetic moment of the central dipole (parallel to the rotation axis, say, directed northward) decreases as a new sunspot cycle advances and becomes null at about the sunspot maximum;
2. Subsequently, a small central dipole of the opposite polarity (directed southward) appears and its moment reaches maximum intensity near the sunspot minimum; and
3. A pair of dipoles on the photospheric surface, located at low latitudes, increases its magnetic moments from the beginning of a sunspot cycle until about the sunspot maximum and then decreases during the declining phase.

These variations are illustrated in Figure 4.4. The top row in the figure shows the sunspot number during Sunspot Cycle 21. The second row shows the neutral line determined by the Wilcox Observatory. The next two rows show the axial and the auxiliary photospheric equivalent dipoles, respectively, for eight different epochs. The second row shows the observed neutral line and the last row, the neutral lines computed based on our three-dipole model. The combined magnetic fields of these dipoles result in the neutral lines shown in the bottom row. The agreement between the observed and computed neutral lines is quite reasonable.

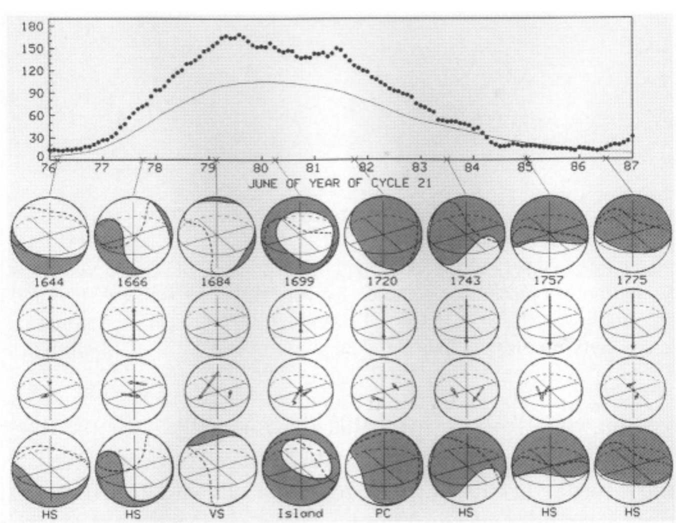


Figure 4.4 Sunspot number, the observed neutral line (Wilcox Observatory), the central dipole parallel or antiparallel to the rotation axis, the auxiliary photospheric dipoles and the neutral line determined by the combination of the central dipole and auxiliary dipoles: T. Saito, T. Oki, S.-. I Akasofu, and C. Olmsted (1991).

The two hypothetical dipoles reveal their existence when we examine the distribution of the photospheric magnetic field. Figure 4.5 shows such an example. The source surface field is modeled in the upper left diagram by a combination of an axially parallel dipole and two photospheric dipoles (the upper right diagram and lower right diagram). The two dipoles, thus determined, are then transferred to the magnetic field map in the lower left diagram. One can see that each of the two dipoles is fairly well co-located with the observed large-scale dipolar field; this is a sort of blind experiment. Therefore, the two dipoles inferred from our modeling method do actually exist. These dipolar fields are not individual sunspot pairs, and are larger scale fields in active regions.

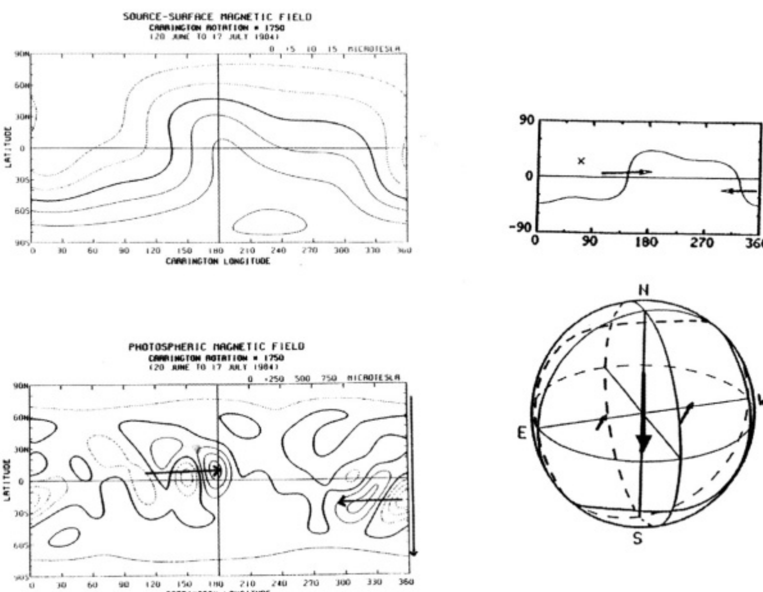


Figure 4.5 Upper left: Observed magnetic equator. Upper right: Observed magnetic equator and the auxiliary dipoles on the photosphere. Lower left: Photospheric field and the location of the auxiliary dipoles on the photosphere. Lower right: Spherical representation of the axial dipole and the auxiliary dipoles: T. Saito, Y. Kozuka, and T. Oki, S.-I. Akasofu (1991).

It is possible, therefore, to infer that the main dipole is axially parallel or antiparallel and that the inclination (with respect to the rotation axis) of the dipole on the source surface is produced by a combined effect of the axially parallel (or antiparallel) field and the two photospheric dipoles. The apparent rotation of the dipole throughout the sunspot cycle is produced by a relative change of the strength of the axially parallel dipole and the photospheric dipoles, together with the reversal of the axial dipole as a result of the migration of a low-latitude unipolar field to the polar region in each hemisphere.

4.3 Large Inclination and Eccentricity of the Dipole-like Field of Uranus and Neptune

It is interesting to speculate that the photospheric surface corresponds to the surface of the core of magnetized planets and that the source surface of the Sun corresponds to the surface of the magnetized planets. An assumption is that the mantle of the planets is inactive in generating the main dipole field. It is expected that the dynamo process is the same or similar for the Sun and magnetized planets. In fact, all the dynamo theories treat them in the same way.

If this is the case, one should be able to assume that the main dipole of the magnetized planets is located at the center of the planets and is parallel (or antiparallel) with respect to the rotation axis and also that additional dipolar fields, together with the main dipole field, give the result of the off-centered and inclined dipole by the spherical analysis method. This is physically a more plausible situation than what the spherical harmonic analysis can provide.

Norman Ness and his colleagues (1986, 1989) suggested that the magnetic fields of Uranus and Neptune indicate that the main field can be represented, as a first approximation, by an eccentric dipole and that the dipole is greatly inclined with respect to the rotation axis; see Table 4.1. Their model is often referred to as the offset tilted dipole (OTD) model. Their results are based on the spherical harmonic analysis of the magnetic field observed along the flyby trajectory of the Voyager spacecraft.

The large inclination and eccentricity of the dipole-like field of Uranus and Neptune can be described, as a first approximation, by the combined field of an axial dipole and a single auxiliary dipole.

The upper left diagram of Figure 4.6 shows the offset tilted dipole (OTD) located near the surface of the core of Uranus, as proposed by Ness et al. (1986). The lower left diagram shows some magnetic field lines in the plane that contains the OTD. We determine the magnitude and the orientation of both an axial dipole and an auxiliary dipole in the same way as we examined the solar source surface field. The results are presented in the upper right diagram of Figure 4.6. For simplicity, we assume only one auxiliary dipole, which is located at the position calculated for the single offset dipole. The parameters for the two dipoles are given in Table 4.2.

Some magnetic field lines of the one- and two-dipole models are shown in Figure 4.6. Comparing the lower left and right diagrams, one can see that the simple two-dipole model can reproduce reasonably well the observed field, which is represented by a single off-centered dipole.

TABLE 4.1

	Inclination Angle	Location from the Center	Reference
Earth	11.5°	0.08 R_E	Chapman and Bartels, 1940
Uranus	~60°	0.3 R_U	Ness et al., 1986
Neptune	~47°	0.55 R_N	Ness et al., 1989

TABLE 4.2 Two-dipole Model of Uranus

	M (Gauss R_U3)	Location	Orientation
Axial dipole	0.143	x = 0 y = 0 z = 0	θ = 0°
Auxiliary dipole	0.157	x = 0 y = 0 z = -0.3 R_U	θ = 90° θ = 0°

TABLE 4.3 Two-dipole Model of Neptune

	M (Gauss R_N3)	Location	Orientation
Main dipole	0.0641	x = 0 y = 0 z = 0	θ = 0° φ = 0°
Auxiliary dipole	0.0769	x = 0.14 y = 0.42 z = 0.24	θ = 60° φ = 0°

There is no doubt that one or two additional dipoles in the model can better reproduce the observed field. However, the main point of this section is to illustrate the basic idea that even a simple two-dipole model could reproduce the observed field fairly well and thus may be able to remove the great puzzle of the large inclination angle and the large eccentricity of the main field of Uranus.

The upper left diagram of Figure 4.7 shows the offset tilted dipole (OTD) of Neptune, as proposed by Ness et al. (1989). The lower left diagram shows some magnetic field lines in the plane that contains the OTD. The magnitude and the orientation of an axial dipole and an auxiliary dipole giving a similar magnetic field are shown in the upper right diagram. The parameters for the two dipoles are given in Table 4.3. Some magnetic field lines of the one- and two-dipole models are shown in the lower part of Figure 4.7.

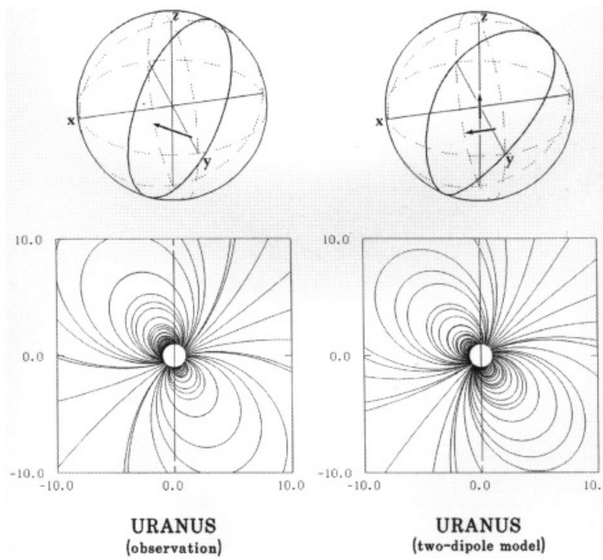


Figure 4.6 Left: Uranus' main dipole and the field lines based on the spherical harmonic analysis. Right: Uranus' magnetic field based on the axial and an auxiliary dipole: S.-I. Akasofu, L.-H. Lee, and T. Saito (1991).

The discovery of the large inclination angle and the eccentricity of the main field of Uranus and Neptune provided a great puzzle. However, it is important to realize that the finding is based on spherical harmonic analysis of the planetary fields observed by a spacecraft flyby. Certainly, the dipole representation based on spherical harmonic analysis provides us with the unique mathematical description of the planetary magnetic field. However, the result obtained does not indicate that the field inside the planets is physically given by such an analysis. Indeed, since the dynamo process is thought to rely so strongly on the rotation of the magnetized planets, it is possible that the observed dipole field consists of the combined field of an axial dipole (parallel or antiparallel to the rotation axis) and a few auxiliary dipoles. This possibility is physically more plausible than the mathematical representation and is further supported because our interpretation was tested in Section 4.1.

It is generally believed that the magnetic axis of neutron stars is also inclined significantly from the rotation axis, so that the same issue may be raised with them.

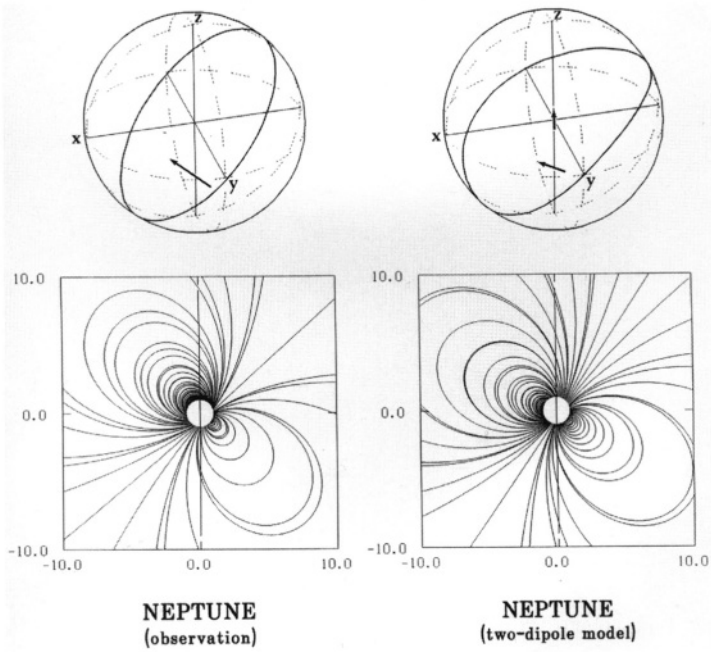


Figure 4.7: Left: Neptune's main dipole and the field lines based on the spherical harmonic analysis. Right: Neptune's magnetic field based on the axial and an auxiliary dipole: S.-I. Akasofu, L.-H. Lee, and T. Saito (1991).

4.4 Is the Earth's Dipole Actually Inclined with Respect to the Rotation Axis?

Spherical harmonic analysis of the Earth's magnetic field indicates that the main field can be represented, as a first approximation, by an off-centered dipole, and that the dipole axis is inclined with respect to the rotation axis by about 11.5°. Since the present dynamo theory for generation of the Earth's magnetic field relies heavily on the planet's rotation, it may be worthwhile to examine whether the Earth's magnetic field could consist of an axially antiparallel dipole and a few dipoles on the surface of the core.

It is our finding that three dipoles near the core surface, together with the axially antiparallel dipole, can reproduce fairly well the magnetic equator (Figure 4.8). The three dipoles are located at longitudes ~105°, ~210° and ~330°, respectively; thus, they are located southeast of Hawaii, at the Atlantic

Ocean between Africa and South America, and at the southern part of Thailand, respectively. It is suggested that the main dipole is aligned with respect to the rotation axis and that the combined effect of the three dipoles provides the tilted and off-centered main dipole.

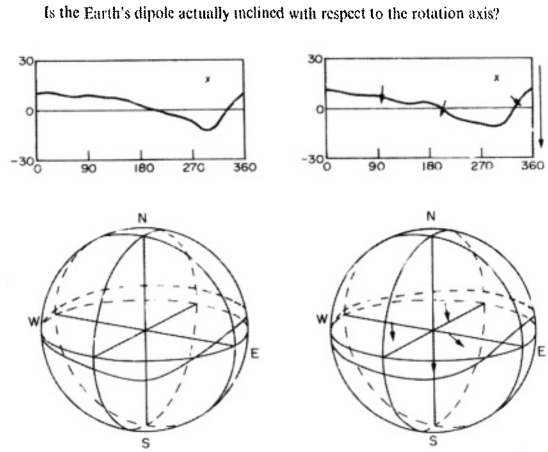


Figure 4.8 The upper- and lower-left presentations show the observed magnetic equator of the Earth in the standard and spherical projections, respectively. The upper- and lower-right presentations show the magnetic equator reproduced by the axially antiparallel dipole and three dipoles on the core surface, which are indicated in both presentations: S.-I. Akasofu., and T. Saito (1990).

It is unfortunate that researchers in this discipline believe so firmly that the spherical harmonic analysis method can provide the physical solution. The late Sir David Bates, editor of the *Journal of Planetary and Space Science,* accepted my paper on this subject against the recommendation of a referee. I learned from him later that he received an angry letter from the referee who stated that he would not submit his future papers to the *Journal* A referee should not forget that the editor of a scientific journal makes the final decision, not the referee. This is not to say that I have an accurate solution to this difficult problem. We have to be open-minded in facing it.

4.5 Does the Main Dipole of the Geomagnetic Field Rotate during the Reversals?

After a long debate, the reality of reversals of polarity of the geomagnetic field through geologic time has been established. The phenomenon of reversals is considered by many researchers to be the *rotation* of the dipole axis, either from the normal to reversed (N → R) or from the reversed to normal (R → N). It is generally accepted that:

(1) The dipole pole shifts often along a restricted sector of longitude.
(2) There occurs a significant reduction of the field.
(3) The transition stage is relatively short, ~4500 ±100 years.
(4) The frequency of reversals has been increasing from 0.5 Ma to 0.15 Ma during the last 70 Ma.
(5) The field becomes highly nondipolar during a transition, although the importance of the higher order terms (g2/g1, g3/g1) during the reversals is not well established; item (2) above is considered by some to be an indication of the growth of the higher-order terms.

Figures 4.3 and 4.4 also show that at the beginning of Sunspot Cycle 21, the magnetic equator lay nearly parallel to the heliographic equator and also the axial dipole was large and was pointing approximately northward. During the ascending phase of Sunspot Cycle 21, the axial component decreased rapidly. The magnitude of the reversed dipole grew steadily during the descending phase of the cycle.

On the other hand, the equatorial dipoles grew rapidly during the ascending phase. When the axial dipole was weakest, one of the equatorial dipoles was very large. Obviously, this particular dipole had the largest influence on the source surface. In fact, the main dipole component on the source surface was almost perpendicular to the rotation axis during Carrington rotations 1681-1685 (see Figure 4.3). After the reversal of the axial dipole, the magnitude of the equatorial dipoles gradually decreased.

Thus, there appears to be some similarity between the polarity reversals observed on the Earth's surface and on the solar source surface. Both reversals occur during a relatively short period compared with the period of one polarity; a significant decrease of the main dipole field occurs, there is some indication of the growth of higher order fields. The major difference is that the solar reversals are quite regular compared with those of the geomagnetic field. In fact, there is no physical reason why one cannot assume as a first approximation that the source surface corresponds to the surface of a magnetized planet, and that the photosphere corresponds to the core surface. The most important point here is that the photospheric magnetic field does not show any indication of the rotation of the dipole axis, in spite of the fact that the dipole field on the source surface rotates; the polar regions remain as the magnetic poles.

As mentioned in Section 4.2, the reversal of the solar dipole field on the photosphere occurs as a result of the migration of a large-scale unipolar field, not by the rotation of the dipole axis. Perhaps a similar process is responsible for the reversal of the Earth's dipole field. Without the reversal of the direction of the Earth's rotation, it is very difficult to explain the reversal of the dipole field.

4.6 Heliospheric Current Sheet

As the solar wind stretches the dipolar field on the source surface, an extensive current sheet is formed, dividing the magnetic regimes into two, the northern and southern hemispheres. However, as shown in Figures 4.3 and 4.4, the magnetic equator has a complicated wavy character. As a result, the heliospheric current sheet also has a complex configuration. It may be possible to test the inferred configuration of the heliospheric current sheet near the Sun by observing the outer solar corona. Kazuyuki Hakamada, Ghee Fry, and I developed a method to construct the heliospheric current sheet near the Sun (1986). We decided to predict the shape of the outer solar corona on the basis of the magnetic equator inferred by the Wilcox Observatory (Figures 4.9a). The predicted coronal configuration is shown in Figure 4.9b. Then, Takao Saito and his colleagues successfully photographed the outer solar corona during the 1991 eclipse (Figure 4.9c). The agreement between the predicted and observed solar corona was unexpectedly good. The outer corona appears to be bright at the places where the heliospheric current sheet develops folds. As far as I am aware, this was the first time that the configuration of the outer corona was predicted so realistically.

The method we developed to infer the heliospheric current sheet can be extended to the Earth's distance and beyond for the realistic magnetic equator. One complexity is that the formation of the wavy magnetic equator does not propagate with the speed of light or infinite speed, but is carried by the solar wind particles with a speed of a few hundred kilometers per second. At the same time, the current sheet rotates with the Sun. Taking this into account, we developed a method to construct the current sheet for any given (observed) magnetic equator. Two examples are given in Figures 4.10a and 4.10b for Carrington rotation 1654 and 1664, respectively. Space probe observations indicate that the current sheet extends into the outer heliosphere. John Wilcox was the first to infer the wavy current sheet for an ideal case, but we could extend his work for realistic magnetic equators. In Section 6.2, we shall see that the azimuth angle of the interplanetary magnetic field changes its direction (from toward to away or vice versa) as the Earth crosses the current sheet.

It is understandable that scientists do not renounce easily what they have been taught and what they have based their research on, even if a new finding does not conform with what they believe in. They consider that such a finding is not credible and discredit the new finding. They begin to lose faith only when many more new findings are inconsistent with that they believe in.

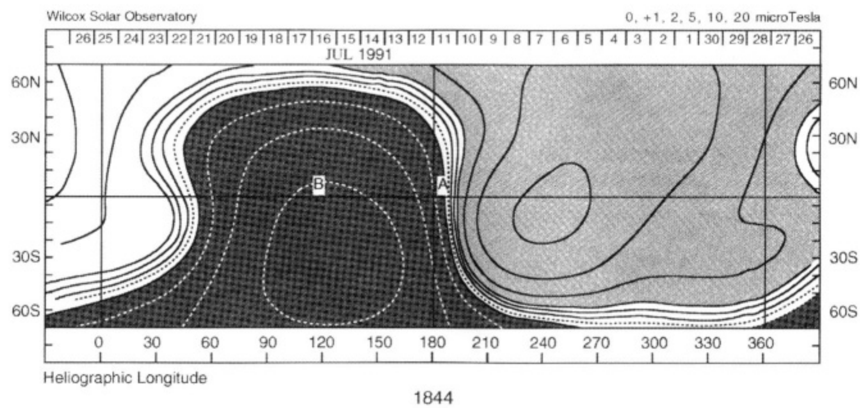


Figure 4.9a Distribution of the magnetic field on the source surface for Carrington rotation 1844: T. Saito, S.-I. Akasofu, Y. Kozuka, T. Takahashi, and S. Numazawa (1993).

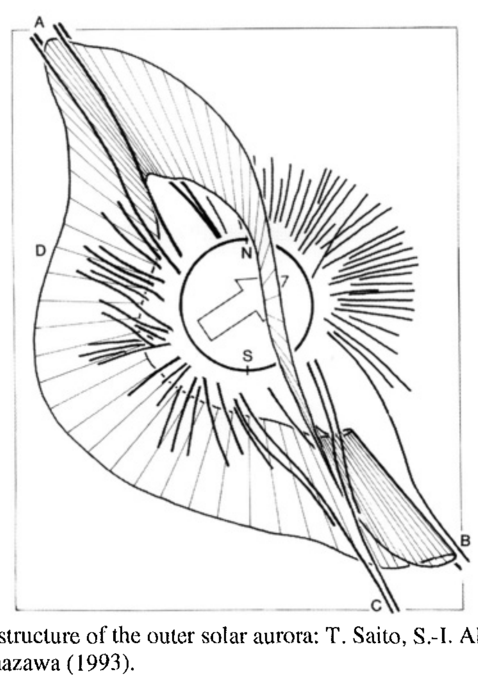


Figure 4.9b Predicted structure of the outer solar aurora: T. Saito, S.-I. Akasofu, Y. Kozuka, T. Takahashi, and S. Numazawa (1993).

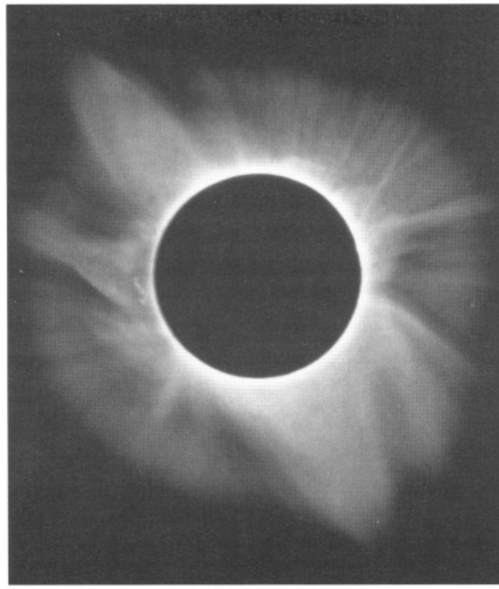


Figure 4.9c Observed outer solar corona during the 1991 solar eclipse: T. Saito, S.-I. Akasofu, Y. Kozuka, T. Takahashi, and S. Numazawa (1993).

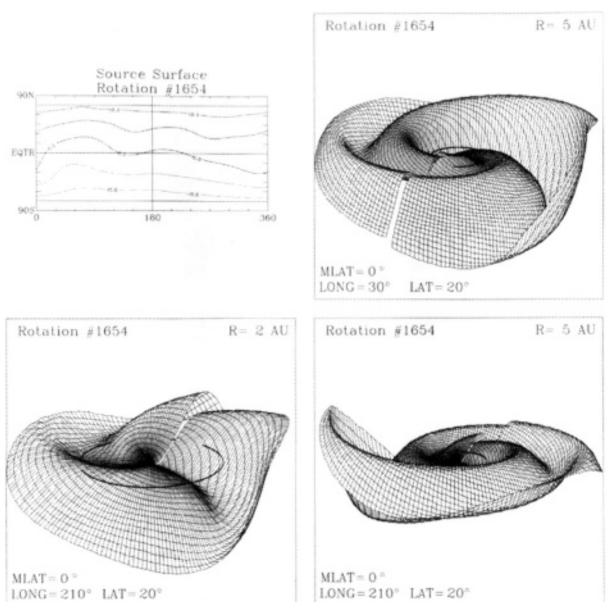


Figure 4.10a The heliosphere current sheet for Carrington rotation 1654. The Earth's orbit is shown: S.-I. Akasofu and C.D. Fry (1986).

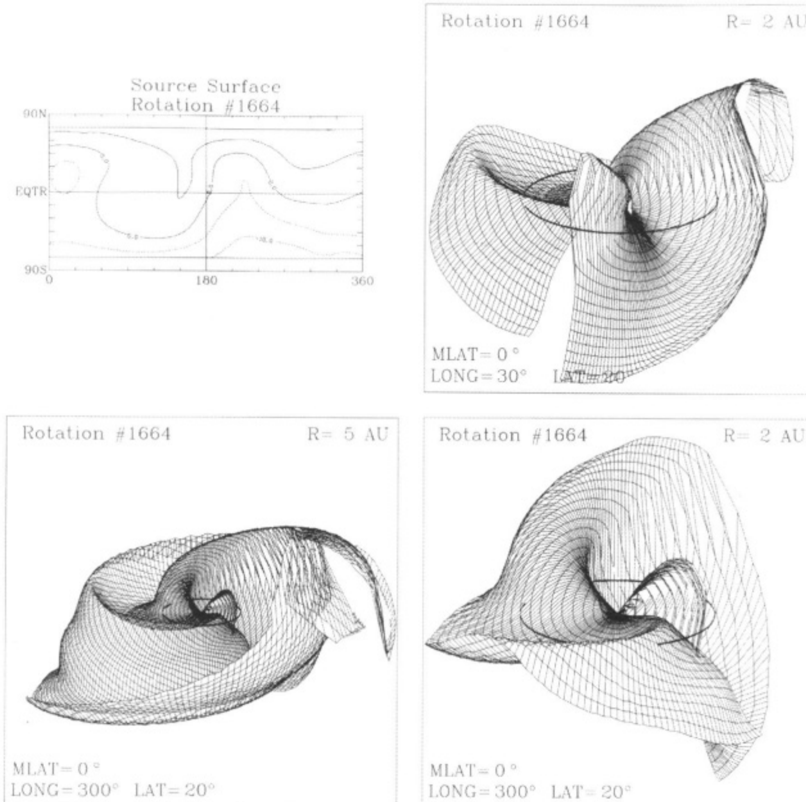


Figure 4.10b The heliospheric current sheet for Carrington rotation 1664. The Earth's orbit is shown: S.-I. Akasofu and C.D. Fry (1986).

Chapter 5

MYTH OF THE EMERGING FLUX TUBES: SUNSPOTS AND SOLAR FLARES

5.1 Introduction

The present guiding concept in searching observationally and theoretically for basic processes of sunspots and solar flares is based on a hypothesis that solar activities are manifestations of interactions of intense magnetic flux tubes that emerge from beneath the photosphere, and the subsequent consequences. The photosphere is considered merely a passive medium through which the magnetic flux penetrates from below. Therefore, the main theoretical efforts have so far been concentrated in examining the emergence of hypothetical magnetic flux tubes for sunspots and instability processes for solar flares, leading to explosive annihilation of the magnetic energy carried up by the flux tubes. For these reasons, all observational/morphological features of solar activities have been discussed in terms of such theoretical implications, e.g., magnetic flux *emergence,* magnetic energy *storage,* flare *buildup, triggering,* instability, and magnetic energy *release,* instead of descriptive terms, as if the hypothesis is confirmed beyond doubt. Thus, most solar physicists working on this particular subject appear to share the paradigm of magnetic flux tubes and of magnetic field annihilation. A very large number of papers have been published with an extremely high degree of agreement on the problems to be solved within the framework provided by this particular paradigm. In these multiple papers, research have articulated and elaborated on the hypothesis. As in the standard paradigm, the solution is assured. Thus, if an anticipated result does not occur, this will be taken as a scientist's failure, not as the theory's failing. Gene Parker (1964) remarked:

> *...it has proved extremely difficult to progress from the general association of flares and magnetic fields to specific processes by which the field actually produces flares. At least on our scratch pads the magnetic field stubbornly refuses to dissipate on command.*

Such a tendency is very unfortunate and even dangerous for the development of solar physics. This is because the identification of observational features corresponding to such hypothetical processes has not necessarily been very definitive, at least not verified. Solar observers must be more independent of theorists for a healthy growth of the field, although this can be said in any scientific field. It is well known that when a particular paradigm becomes dominant, it becomes very difficult to publish a paper that casts doubt on it; the referees could ask almost impossible tests to confirm the claims. The author attempts to overcome such a hindrance usually with great difficulty. On the other hand, observations that appear to conform to the paradigm may be accepted without much scrutiny.

5.2 Emerging Magnetic Flux Tubes

The theoretical difficulties we are facing today in understanding transient solar phenomena may not always be due to our present inability in handling theoretical problems and in sorting out observations. It is likely that the problem is that no doubt has been cast on the guiding concept of hypothetical magnetic flux tubes below the photosphere and of magnetic reconnection. It has been forgotten that the present guiding concept itself consists of a three-step hypothesis:

1. Hypothesize the presence of intense magnetic flux tubes of various sizes and orientations beneath the photosphere.
2. Hypothesize their rise through the upper boundary of the photosphere by magnetic buoyancy.
3. Hypothesize annihilation of their magnetic energy by explosive magnetic reconnection from interacting with other flux tubes.

This three-step hypothesis has been held for several generations. Therefore, it has now become a *doctrine;* the occurrence of a sunspot pair is considered to be proof of the hypothesized flux tube instead of the source of the hypothesis. A typical response from a solar physicist to my question “What is the proof for the presence of a thin magnetic flux tube?" is “A pair of sunspots.” Figure 5.1 schematically illustrates these three steps. An important point to make here is that it has not been observationally and theoretically confirmed that thin flux tubes of various sizes and orientations can be created and exist just below the photosphere. Therefore, it is still only a hypothesis.

We should have several choices: one of them is that intense magnetic flux tubes of various sizes and various orientations can be formed beneath the photosphere and form complex sunspot groups after emerging through the photosphere; in the second, large-scale, weak fields on the photosphere are basic and become *seed* fields for the dynamo process, which concentrates them into sunspot fields, which spread to become the large-scale fields again, becoming the seed field for sunspots.

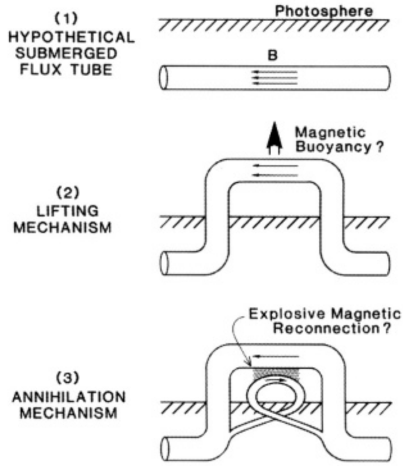


Figure 5.1 Illustrating the three-step hypothesis.

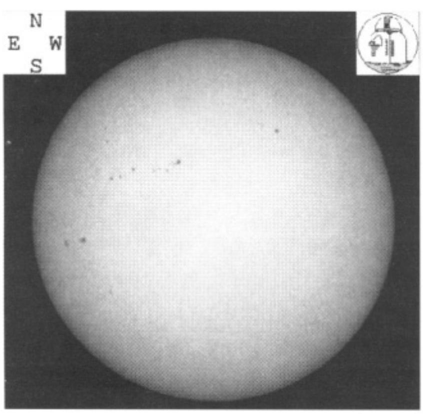


Figure 5.2a Photograph of the solar disk on February 6, 2001. (Big Bear Solar Observatory)

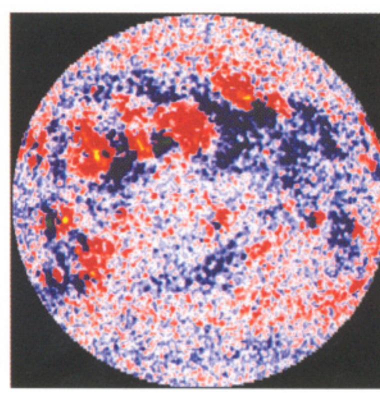


Figure 5.2b Example of solar magnetograph. Red areas show where the magnetic vector points toward the Earth, while blue areas show where the vector points away from the Earth, February 6, 2001. (Kitt Peak Solar Observatory).

There may be others; there is nothing wrong in assuming that dead spots (the large-scale, weak fields) are recycled for new spots. Figures 5.2a, and 5.2b show an example of the observed distribution of solar magnetic fields on the photosphere. It is important to note that a sunspot of one polarity (say, positive) tends to form within a large-scale field of the same polarity (positive) and that a large sunspot pair tends to form in the vicinity of the boundaries of positive and negative large-scale fields (Pat McIntosh, 1981). These will not be such a relationship, if the emergence of magnetic flux tubes can occur randomly.

In this section, we consider that *the photosphere is an active medium,* rather than a passive medium through which the hypothetical flux tube merely penetrates. Specifically, we consider a process associated with vortex motions that can concentrate the *observed* weak field, forming sunspots of a variety of sizes.

One of the fascinating aspects of a vortex flow, such as a cyclone and a hurricane in the Earth's atmosphere, is that it is associated with converging flow near its base. In the photosphere, the converging flow can concentrate weak magnetic fluxes into a relatively small area and the associated dynamo process may amplify the field. However, a concentration process would not work efficiently for a horizontal flux tube beneath the photosphere, because the mounting pressure gradient tends to counteract it, even if the compressibility of the gas is taken into account. In fact, there is hardly a paper that can demonstrate the formation of even a single thin magnetic flux tube below the photosphere.

First of all, solar physicists must theorize the formation of magnetic flux tubes of a variety of sizes that lie horizontally and are oriented in many directions. In the photosphere, there is no difficulty in concentrating *vertical* magnetic fluxes because the photospheric gas can escape from the top of the photosphere. In fact, such an outward flow (the Evershed flow) from the top of sunspots is a well-known feature (Figure 5.3). Figure 5.4 shows a clear indication that a sunspot can be associated with a vortex flow, although solar physicists do not want to pay any attention to such an *observed* feature. In fact, there is no paper on this particular observation except my own (Akasofu, 1985). It is important to note that the cloud structure in a hurricane consists of both large- and small-scale features (Figure 5.4), and all of them are associated with upward flows of air and thus converging flows near the bottom of all the clouds, the smallest one being a cumulonimbus. A similar statement may be made on the structure of sunspots in Figure 5.4. It is basically impossible to figure out the sunspot structure in Figure 5.4 and many others in terms of emerging magnetic flux tubes of different sizes.

In his book titled The Solar Atmosphere, Hal Zirin (1966) noted:

If the theorist, who prefers things in neat packages, were presented with an ideal spherical sun, even with some magnetic field, he would never *predict even sunspots, much less flares. The same is true of tornadoes or hurricanes in the terrestrial atmosphere. But faced with their existence, he must come up with some mechanism that might produce this great energy release.*

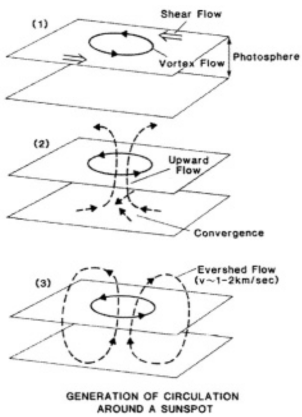


Figure 5.3 Schematic representation of how a sunspot might form as a result of vortex motion in the photosphere.

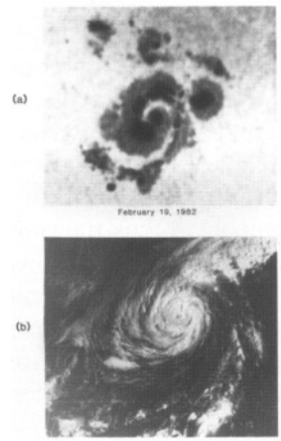


Figure 5.4 Upper: An observed sunspot group in the northern hemisphere (Kitt Peak Solar Observatory). Lower: A photograph of a hurricane near Japan (Japan Meteorological Agency): S.-I. Akasofu (1985).

We must be cautious in interpreting magnetic loops in the corona, which are imaged by x-rays. They are not necessarily isolated magnetic tubes of force and are more likely to be tubes of higher plasma density. If this would be the case, there is less magnetic flux in the tubes because of the diamagnetism of coronal plasma.

5.3 Energy Source

Since many solar transient phenomena are manifestations of electromagnetic processes, we must deal with the dynamo process that can supply the power. Thus, 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 weakly ionized atmosphere, the degree of ionization being 10^{-4} to 10^{-5} in the quiet photosphere and perhaps 10^{-6} to 10^{-7} near 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. Assuming an area of 10^5 km x 10^5 km (a typical radius of an umbra ~10^4 km) and depth of 10^3 km, the density of ~10^{-7} g/cm^3 and the speed of 1000 m/sec, the total bulk kinetic energy of the photospheric flow is ~10^{31} erg. The energy dissipation rate in most intense flares is known to be ~ 10^{29} erg/sec.

Thus, so long as thermal convection, the pressure gradient, and other mechanisms for the neutral component of the photosphere can maintain the flows, it is basically a small portion (1%) of the bulk kinetic (mechanical) energy of the neutral component that is converted into electrical energy. The bulk kinetic 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. The dissipation process of the power thus produced reduces the bulk kinetic energy of the ionized component and therefore its bulk speed $\boldsymbol{V}_p$. However, the differential speed between the ionized component $\boldsymbol{V}_p$ and the neutral component $\boldsymbol{V}_n$ ensures the transfer of the bulk kinetic energy from the neutral component to the ionized component (Figure 5.5).

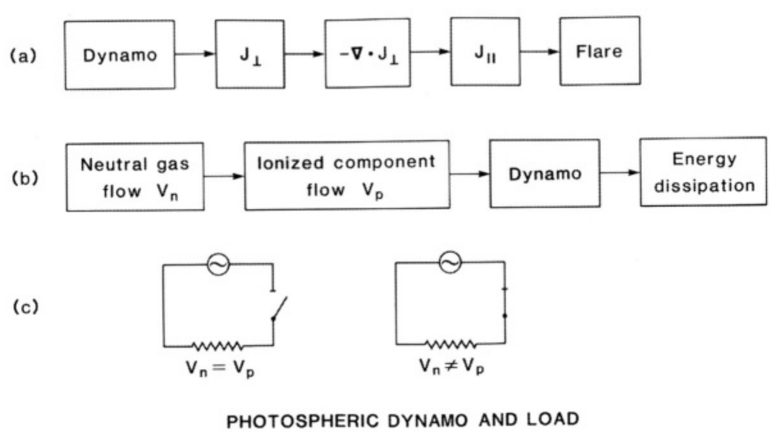


Figure 5.5 Diagram showing how the kinetic energy of the neutral component of the photosphere drives a solar flare.

Note that if there is no dissipation and if both velocities remain the same, the resulting situation corresponds to a dynamo with an open circuit (or without load). Therefore, however small it may be, the velocity differential is most crucial for the photospheric dynamo to generate the power for the *sunspot circuit* and the *flare circuit.* An important point is that one cannot predetermine the velocity differential based on 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 can take place in a region far from the dynamo region, which is connected by the magnetic field lines. Kan et al. (1983) showed that the power of the dynamo is given by:

$$P = -\Sigma_p \boldsymbol{E} \bullet (\boldsymbol{V}_n - \boldsymbol{V}_p) \times \boldsymbol{B}$$
$$\boldsymbol{E} = -\boldsymbol{V}_p \times \boldsymbol{B}$$

Here, $\boldsymbol{\Sigma}_p$ is the Pedersen conductivity of the photosphere which, in the lower photosphere, is similar to that of sea water.

The justification of the common assumption of infinite conductivity for such a low-conductivity medium relies on a large-scale length L

$$VB/L >> B^2/4\pi\Sigma_p L^2.$$

This consideration of justifying infinite conductivity of the photosphere and the assumption of $\boldsymbol{V}_p = \boldsymbol{V}_n$ may eliminate the possibility of understanding solar flares. In magnetospheric physics, we have found that the simplified MHD equations are hardly appropriate in dealing with the auroral potential structure. *In fact, by the initial MHD assumption of infinite conductivity along the magnetic field lines and thus* $\boldsymbol{E}_{||}=0$*, we used to throw away the very solution of the auroral particle acceleration process* ($\boldsymbol{E}_{||}\neq 0$) even before engaging to solve the problem. The MHD method is only one of the tools in understanding magnetospheric and solar phenomena. This is often forgotten. It is a useful tool for understanding certain phenomena, but it is not the universal tool.

Hannes Alfvén was the first to insist that the frozen-in field condition ($\boldsymbol{E}+\boldsymbol{V}\mathrm{x}\boldsymbol{B} = 0$) should be thawed. In his Birkeland Symposium paper titled *The Second Approach to Cosmical Electron Dynamics,* (1968) he stated:

> *...One has good reasons to suspect that there often exist electric fields with components parallel to the magnetic field. The existence of such fields may invalidate the "frozen-in" picture in many cases. We may say that the first new principle is associated with a "thaw" of the frozen-in field lines.*

However, in spite of the fact that he was the founder of MHD physics, some MHD theorists accused him of being a heretic. The MHD formulation became such a powerful paradigm, even its founder had to be accused as a maverick when he warned us of its limitation.

5.4 Sunspots

It so happened many years ago that Gene Parker and I ran into each other at the Logan International Airport in Boston and had a cup of tea. I asked him what was the most difficult problem he had encountered in his life. His response was simply *sunspots.* Although he may not remember the conversation, I was greatly impressed by it.

It is my belief that it is best to go back to observed facts, not a theoretical interpretation of them. As mentioned earlier, it has been proposed that the large-scale fields are basic in considering the formation of sunspots. A positive sunspot appears where there is a positive large-scale field and a negative sunspot appears where there is a negative large-scale field. This large-scale field pattern exists prior to the appearance of large sunspot groups (McIntosh, 1981); see Figure 5.2b. One of the important aspects of sunspot formation is that a pair of sunspots tends to form near a polarity reversal boundary of the large-scale fields, which is often marked by dark chromospheric filaments. Figure 5.6 shows schematically the relationship between the large-scale fields and large sunspot pairs.

It is of great interest that only one side (say, -/+) of the boundary in each hemisphere tends to produce sunspots and that the opposite side (say, +/-) is active in the other hemisphere. L. Svalgaard and John Wilcox (1976) referred to the active boundary as the Hale boundary.

McIntosh (1981) demonstrated that sunspots tend to form near a belt of highly sheared flows, which is far greater than the shear associated with the nonuniform rotation of the Sun. A large-scale shear flow belt forms first in high latitudes at the beginning of a new sunspot cycle and shifts equatorward, just like sunspot groups.

If a vortex motion occurs in the belt of the shear flow, a large concentration of the magnetic flux can be expected. Assuming that the magnetic field is nearly frozen in the ionized component, it is possible to make a rough estimate of the initial radius $r_o = r_1 (B_o B_1)^{1/2}$, where r_1 and B_1 are the radius and the magnetic field intensity of the umbra, respectively, taking r_1 = 1000 km and B_1 = 1000 G and B_o = 10G, r_o ~ 10,000 km. Assuming this concentration and that the formation of a very small spot can be achieved in one day, the required speed is about 100 m/sec

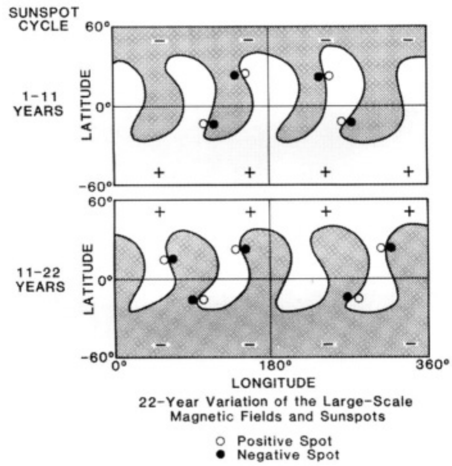


Figure 5.6 Relative location of the large-scale photospheric fields and sunspot pairs during the 22-year cycle variation.

As one vortex grows, both the inward flow and the vortex motion will be transmitted to the conjugate area along the magnetic field lines that loop across the polarity reversal boundary. As a result, the magnetic flux will also be concentrated in the conjugate area, inducing and forming a spot of the opposite polarity. Figure 5.7 shows schematically the formation of a sunspot pair across the polarity reversal boundary. It should be noted that the concept of a magnetic flux tube cannot explain that a single sunspot occurs first, rather than a pair from the beginning.

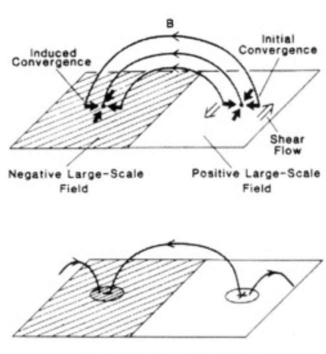


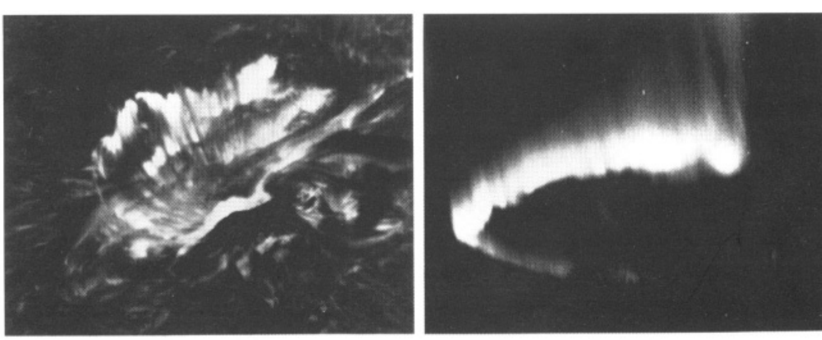
Figure 5.7 Schematic illustrations suggesting how a pair of sunspots might form across the polarity reversal boundary. The converging effect in the positive large-scale field is communicated along the magnetic field lines to the conjugate area, forming a spot of the opposite polarity.

5.5 Force-free Fields and Solar Flares

In his Caltech office, Hal Zirin and I once ran both his solar flare movie and my all-sky aurora movie together side by side. He commented that the aurora is an *Earth flare,* while I insisted that a solar flare is the *solar aurora.* In any case, a flare and the aurora result from optical emissions of the atmosphere of the Sun and Earth, respectively. In fact, there are many similarities between the two phenomena (Figure 5.8). Some of them are:

1. Both are atmospheric emissions caused by impacts of energetic electrons.
2. Both appear in a ribbon-like form (actually curtain-like).
3. Both appear at the feet of magnetic field line loops.
4. Both are associated with a great variety of electromagnetic processes, requiring a dynamo process to power them.
5. Both are associated with the field line currents (see following).

Solar physicists agree in general that the *force-free* field ($\boldsymbol{J} \times \boldsymbol{B} = 0$) is vital in understanding solar activities. In a force-free field, electric currents flow along the magnetic field lines.



SOLAR FLARE AURORA

Figure 5.8 Both solar flares and the aurora result from emissions of the atmosphere particles. Flares have a two-ribbon structure connected by magnetic field lines (Big Bear Solar Observatory). The northern and southern auroras are connected by the geomagnetic field lines.

There have been countless papers on force-free field configurations, that simply solve $(\nabla \times \boldsymbol{B}) \times \boldsymbol{B} = 0$ under various conditions, but there have been hardly any papers on how force-free fields can be generated, and more specifically how field-aligned currents $J_{\|}$ can be generated. The dynamo process associated with vortex flows is an important element in generating field-aligned currents $J_{\|}$ and the resulting force-free and *sheared* fields near active sunspots (Figure 5.5). Solar physicists have become accustomed to considering solar activity in terms of a magnetic flux *tube* and thus bypassing the processes that produce the tubes. They must first consider how long and thin magnetic flux tubes of various sizes and various orientations can be generated below the photosphere. A thin magnetic tube requires solenoidal currents, and it is interesting to see what processes are hypothesized to be responsible for the solenoidal currents. In this context, Alfvén was treated as a maverick again by insisting on the need of considering electric currents. However, solar physicists responded that there is no $\boldsymbol{J}$ term in their MHD equations ($\boldsymbol{J}$ is converted into $\nabla \times \boldsymbol{B}$). It is quite obvious that there is no rigid electrical circuit in the solar atmosphere. However, this does not mean we should forget the physics involved in generating $J_{\parallel}$.

In a model of eruptive prominences by Choe and Lee (1996), they assumed an arcade-like magnetic field configuration (Figure 5.9). In their model, there is an antiparallel flow of photospheric gas across the centerline of the arcade, namely the dynamo process. The field-aligned currents flow from both sides. In this particular model, *the neutral line* is supposed to exist along the centerline of the arch. However, there is a strong horizontal magnetic field there and it tends to align along the central line because of the field-aligned currents. This tendency has been observed.

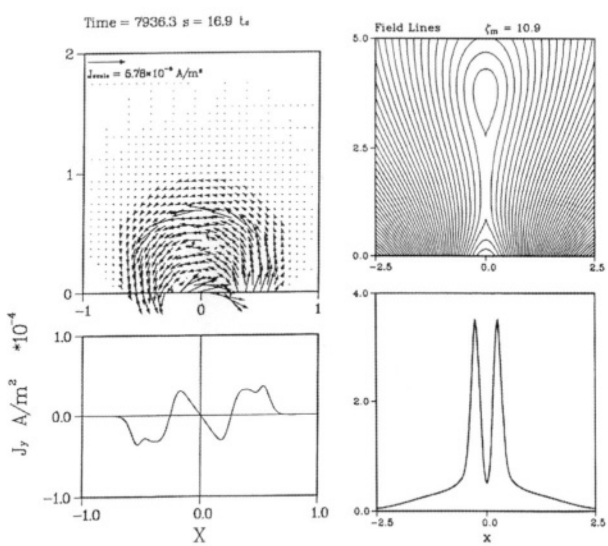


Figure 5.9 A simulation of an erupting prominence and the associated electric current distribution: C.S. Choe and L.C. Lee (1996).

Recall that a force-free field indicates nothing but the presence of field-aligned currents $J_{\parallel}$ in the solar atmosphere (Figure 5.5). In either solar conditions or magnetospheric conditions, the field-aligned current $J_{\|}$ must be related by the dynamo process:

$$J_{\|} = - \nabla \bullet \mathrm{I}_{\perp}$$

where $\mathbf{I}_{\perp}$ must be generated by the dynamo process $\mathbf{V} \times \boldsymbol{B}$, so that $\mathbf{I}_{\perp}$ must be perpendicular to $\boldsymbol{B}$. Thus, if a force-free field is crucial for solar flares, the dynamo process must be involved in generating $J_{\|}$. It is unfortunate that such a dynamo process and thus the ultimate energy source are ignored in solar physics.

It is important to note that the field-aligned currents distort the magnetic field configuration from the potential fields. In a simple case of a pair of sunspots, consider a vertical plane that contains the two spots. There is a loop of magnetic field line contained in this plane. The projection of this particular field line onto the plane horizontal to the vertical plane is a straight line that connects the two spots. The dynamo process around the two spots tends to distort the field in such a way that the projected straight line becomes an S-shaped character. Such a magnetic field is called a sheared magnetic field. Magnetic energy stored in the distorted portion of the magnetic field is expendable, not the potential portion of the field. The degree of the distortion provides a measure of the amount of expendable magnetic energy.

The stored magnetic energy were expended for solar flares by magnetic reconnection, we would not expect an increase of magnetic shear; this is because the magnetic field configuration would have to relax toward a potential field during a flare. Thus, it is important to examine changes of vector magnetic fields, in particular magnetic shear associated with intense flares, on the basis of a high-resolution transverse and longitudinal magnetic field measurement. I proposed to Hal Zirin that it is important to examine how magnetic shear changes at flare onset. Haimin Wang et al. (1993) showed the magnetic shear actually increases after several major flares; Figure 5.9 shows such an example. One can clearly see a large increase of the shear at flare onset.

It is obvious that an enhanced photospheric dynamo process provides energy for both solar flares and the sheared field. This situation is similar to magnetospheric substorms, since the size of the polar cap can increase at substorm onset, indicating that the solar wind-magnetosphere dynamo power goes to both auroral substorms and the magnetotail (Section 3.7).

Haimin Wang, Hal Zirin, and their colleagues reported their finding (Figure 5.10) for five X-class flares in their paper in *Astrophysical Journal* (1994). Toward the end of their paper they noted:

> *It is interesting to note that the direct driving model for solar flares proposed by Akasofu (1984) predicts that the magnetic shear should increase during solar flares. In this model, the energy source for flares is photospheric motions which drive a dynamo process, the energy release is due to dissipation of the field-aligned currents, and the amount of energy released as a function of time is directly related to the shear of the magnetic field. Although the model seems to explain our observations, there is a major inconsistency: the increased shear we observe persists well after the flare emission has ceased, whereas the direct driving model predicts that the shear must decrease again after the flare, otherwise the flare emission would continue.*[3]

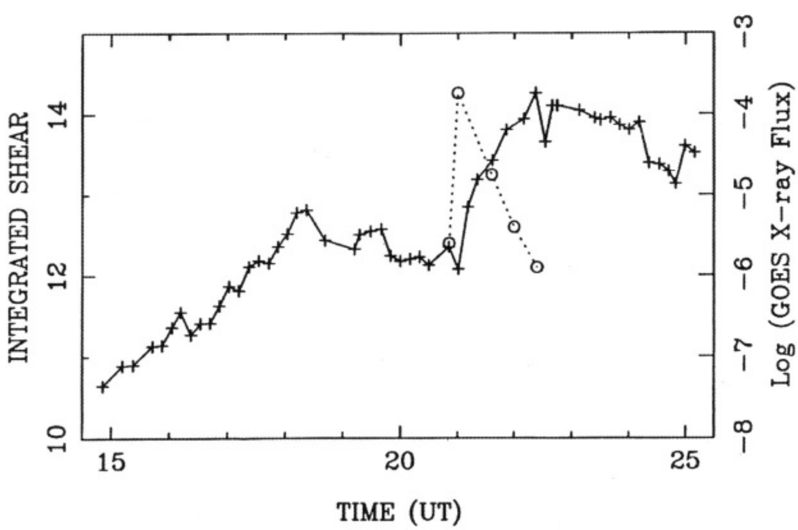


Figure 5.10 Changes of magnetic shear at the time of flare onset. Note a sudden increase of the shear, instead of decrease as expected from magnetic reconnection: H. Wang, M.W. Ewell, Jr., and H. Zirin (1994).

---

[3] Our auroral studies show that there is no simple relationship between the solar wind-magnetosphere dynamo power (ε) and auroral brightness. Auroras tend to be brightest during the first half hour after substorm onset, although the substorm typically lasts three hours. It may also be noted that a strong heating of the solar atmosphere continues well after optical flares.

5.6 Magnetic Reconnection

In Section 1.9, I stated that I intuitively avoided magnetic reconnection from my consideration of understanding various solar-terrestrial phenomena, in spite of the fact that the scientific community has been emphatic that magnetic reconnection was the most important process in this particular discipline, in particular in understanding solar flares and magnetospheric substorms.

The concept of magnetic reconnection originated in an attempt to explain solar flares in the 1950s and 1960s. At that time, it was said that there was no intense flow of the photospheric gases around sunspots, so that the only source of energy that could be tapped for solar flares was magnetic energy. Actually, there are complicated flows around a large sunspot group, but they could not be observed by the Doppler method near the central part of the solar disk.

Magnetic reconnection was considered the process that releases magnetic energy from sunspots. Since solar flares appear to be explosive (in a speeded-up flare movie!), it was concluded that magnetic reconnection must be explosive. Somehow, most researchers in this field began to believe that magnetic reconnection must be an explosive process. Perhaps, most theorists of solar flares have not had an opportunity to observe one on a real-time basis; it is actually a rather slowly (boringly!) developing phenomenon.

Thus, theorists considered an idealized situation in which an antiparallel magnetic field configuration is given as the pre-existing condition to be annihilated. It did not matter for them whether a static, steady antiparallel magnetic field configuration would exist in the turbulent solar atmosphere. It was a great surprise for the theorists to find that it is very difficult for this configuration to annihilate itself explosively. Hundreds of papers were published in claiming that this difficulty could be removed. If flares are not an explosive phenomenon, such an exercise was not needed. As I mentioned in Section 3.6, it is my feeling that if a static, steady antiparallel magnetic configuration could be produced, it would be so stable that an explosive annihilation would not be possible.

The only possible way to produce such a field configuration may be to *drive* two mutually antiparallel fields toward each other. Thus, magnetic reconnection might occur only under a driven condition. The problem is that energy released from such a process may be exactly the same as that needed to drive the two fields together, so that no extra energy can be released from the antiparallel field configuration itself in an *annihilation* process. This may be what happens on the dayside magnetopause, where the southward-oriented interplanetary magnetic field interacts with the northward-oriented magnetic field of the magnetosphere. In this situation, no one has claimed magnetic reconnection on the front of the magnetosphere as the magnetic energy release process.

Furthermore, as expendable magnetic energy in the vicinity of sunspot groups exists only in the nonpotential component, hundreds of papers were written on force-free fields, namely on the solution of $(\nabla \times \boldsymbol{B}) \times \boldsymbol{B} = 0$. However, it appears that many researchers forget that force-free fields arise from field-aligned currents ($J_{\|}$) and that $J_{\|}$ must be produced by a dynamo process, which requires the flow of photospheric plasma across magnetic field lines.

When the magnetotail and its near-antiparallel magnetic field configuration were discovered, many researchers thought that the magnetic energy in the magnetotail was the source of energy for magnetospheric substorms. It was said at that time that there was more than enough magnetic energy for thirty substorms. Those researchers faced the same problem as solar physicists. That is, they found that it is difficult to release energy explosively from the magnetotail. Further, I recall I asked them to consider how to stop explosive magnetic reconnection if they could succeed in initiating it. Otherwise, as mentioned earlier, the whole magnetotail will be in effect burnt up, and we know this does not happen. No one has made any study of this point!

Unfortunately, most solar and magnetospheric physicists, both theorists and experimenters, have believed that magnetic reconnection must occur in spite of such a theoretical difficulty and that the resulting X-line (the neutral line) would explain practically all substorm features, including the sudden brightening of an arc at substorm onset and the acceleration of auroral electrons. These claims have not been substantiated. Unfortunately, this trend has considerably retarded the progress of our discipline. It took more than a decade to convince the scientific community that the magnetosphere must be driven each time for a magnetospheric substorm; magnetospheric substorms do not arise from an explosive release of magnetic energy that is steadily accumulated in the magnetotail. Rather, as we learned in Section 1.9, a substorm occurs after the solar wind-magnetosphere dynamo power is substantially increased for an hour or so.

The TRACE satellite showed recently that the energy source of solar activities lies in the lower corona. It may well be that the source lies below it, certainly not high in the corona, as suggested by many researchers who consider magnetic reconnection. Figure 5.11 shows an example of the TRACE image.

It is my hope that both solar physicists and magnetospheric physicists can learn more from each other, even if solar flares and magnetospheric substorms may be found to have different causes.

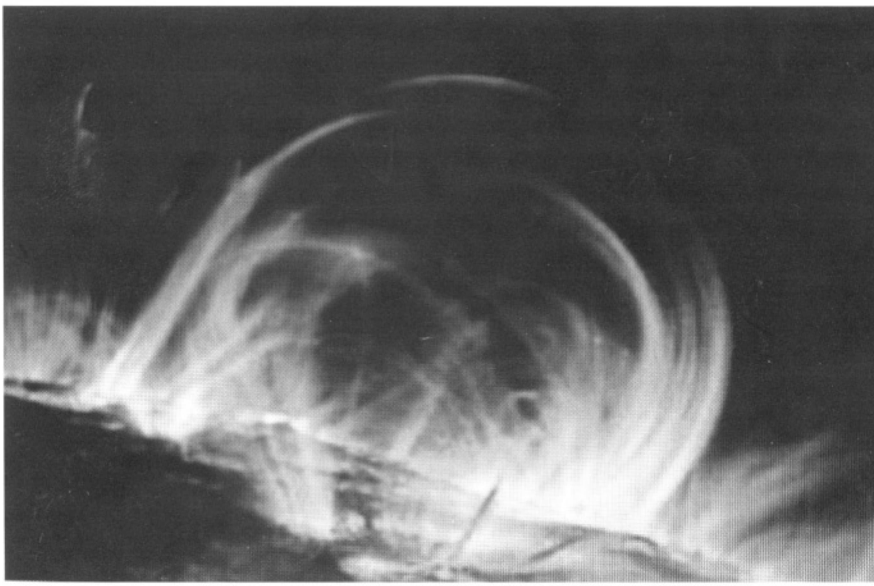


Figure 5.11 Example of the images taken from the TRACE satellite. The Transition Region and Coronal Explorer, TRACE, is a mission of the Stanford-Lockheed Institute for Space Research (a joint program of the Lockheed Martin Advanced Technology Center's Solar and Astrophysics Laboratory and Stanford's Solar Observatories Group), and part of the NASA Small Explorer program.

Chapter 6

SYNTHESIS OF THE FOUR MAJOR DISCIPLINES: PREDICTING GEOMAGNETIC STORMS AS A SPACE WEATHER PROJECT

6.1 Introduction

The study of the solar-terrestrial relationship has developed into four major disciplines: solar physics, interplanetary physics, magnetospheric physics, and ionospheric physics (aeronomy). Researchers have made considerable progress within each field of study during the twentieth century, although many challenging problems have been left unsolved. Meanwhile, there has been much discussion about space weather research during the last few years. The term *weather* in this context implies that the goal of space weather research should be to forecast space weather and, at least, predict geomagnetic storms in terms of the two geomagnetic indices Dst and AE as a function of time.

To be successful in this particular effort, space weather researchers need to establish a new discipline that synthesizes and integrates the four major disciplines and their subdisciplines. This chapter describes a research scheme needed for success in predicting geomagnetic storms by presenting an example of this integration process. The needed efforts are:

1. Modeling of the background solar wind flow
2. Parameterizing solar events on the source surface
3. Modeling the propagation of shock waves (including the simulation of past events)
4. Estimating the velocity, density, and IMF of the solar wind at the Earth
5. Characterizing geomagnetic storms
6. Predicting the size of the auroral oval

Many atmospheric scientists believe that weather forecasting is not their immediate responsibility and there is a similar feeling among many space physicists on space weather forecasting, at least until a few years ago. Many researchers are naturally interested in one or at most two or three subjects in the above list, but very few have been interested in integrating all seven subjects. Indeed, there have been only a relatively small number of papers that dealt with the entire topic in the past. Ghee Fry and I (1986) and Murray

Dryer (1994) attempted to integrate the four disciplines for space weather research. The reason I initiated space weather research in the 1970s, well before this particular research has become so popular among space physicists, was that Colonel Lee Snyder was sent from the U.S. Air Force to study under me for his graduate research. He became responsible for an over-the-horizon (OTH) radar in Maine for Russian bomber surveillance; the problem was that the OTH radar could not detect the bombers when an intense auroral storm was in progress, so that it became necessary to predict the occurrence of auroral storms. For these reasons, Kazuyuki Hakamada, Ghee Fry (also sent by the US Air Force to the University of Alaska) and I developed a space weather prediction scheme that is often referred to as the HAF (Hakamada-Akasofu-Fry) model; in a sense, it was a product of the Cold War. During this project period, I was strongly convinced that it would not be possible to succeed in predicting geomagnetic/auroral storms without integrating the four fields. There are at present many missing links among the four disciplines necessary in order to succeed in space weather prediction.

6.2 Modeling the Background Solar Wind Flow

The solar wind exhibits considerable variation even without any specific solar events, such as solar flares, CMEs, and sudden filament disappearances. This is particularly the case when high-speed streams flowing from long-lasting coronal holes, and the interplanetary sector boundary structures corotating with the Sun, are present. Since any effects of specific solar events propagate into the existing solar wind structures and interact with them, it is important first of all to devise a simple way to model the background flow. Conditions on the source surface, the imaginary spherical surface of 2.5 solar radii, are important in modeling interplanetary conditions. One of the most important aspects of the source surface is the magnetic equator (or the so-called *neutral line).* The axis of the dipolar field on the source surface rotates from 0° to 180° (or from 180° to 0°) during the Sun's 11 -year cycle variations (Chapter 4, Figure 4.3). We found that we can reproduce most of the main features of solar wind variations during the whole sunspot cycle at the Earth or at any point to about a distance of 2 AU by assuming that the solar wind speed is minimum at the sinusoidal magnetic equator and increases toward higher latitudes. Although more realistic models are adopted in an advanced HAF model, this chapter is intended to provide the basic principle involved in our integration effort.

As the Sun and its source surface rotate about every 25 days, a fixed point in space (not on the source surface) at a distance of 2.5 solar radii scans horizontally the velocity field from solar longitude 360° to 0° in one solar rotation along a heliographic latitude line (e.g., 0° at the June and December solstices). Figure 6.1 shows an example of model distribution of the solar wind speed on the source surface and the resulting wind speed at a particular point (fixed in space) on the source surface. Solar wind particles leave radially from this particular point one by one with different velocities as the Sun rotates; the point depicts a sinusoidal variation of the speed of solar wind particles during one rotation. As a result, each solar rotation sends out two waves. Subsequent changes of the radial speed of individual particles can be modeled by adopting a kinematic solution in a method developed by Kazuyuki Hakamada (Hakamada and Akasofu, 1982). A faster flow of particles interacts with a slower flow of particles to form a shock wave and a reverse shock. Thus, by integrating the velocity as a function of time graphically, one can determine the distance traveled by individual particles (see Figure 6.2).

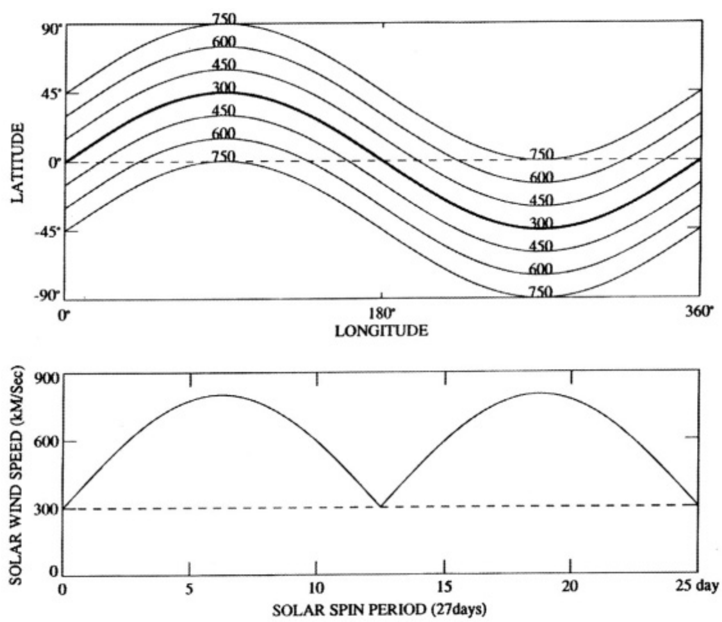


Figure 6.1 Upper: A sinusoidal magnetic equator and the associated solar wind speed distribution on the source surface. Lower: Solar wind speed observation at an equatorial point (fixed in space) at a distance of 2.5 solar radii.

A magnetic field line originating from the source surface can be traced by following particles leaving a particular point on the source surface (not a fixed point in space). The resulting interplanetary magnetic field structure is the familiar Parker spiral, together with the corotating interaction region produced by the formation of the shock wave structure (Figure 6.3). The computed velocity (V), density (n) and IMF magnitude B agrees reasonably well with the observed ones (Figures 6.4a and Figure 6.4b).

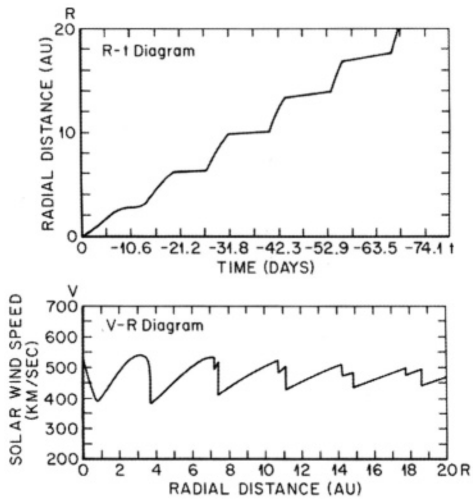


Figure 6.2 Upper: The distance (R) and time (t) relationship for a solar wind particle. Lower: The velocity (V) and distance (R) relationship for a solar wind particle.

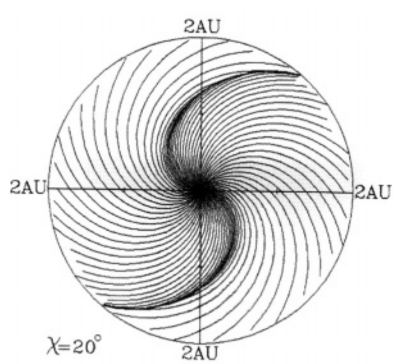


Figure 6.3 The spiral structure of the interplanetary magnetic field lines on the equatorial plane. The magnetic equator on the source surface is assumed to be sinusoidal, the amplitude being 20° in latitude (see top diagram of Figure 6.1).

It is important to realize that such a simple scheme can reproduce reasonably well the observed 27-day variations of the solar wind observed at the Earth. The arrival of the fast wind is associated with a sharp change of the azimuth angle (PHI) of the IMF (from toward to away or away to toward), indicating the crossing of the heliospheric current sheet. It is for this reason that the corotating structure is called the sector boundary.

Some solar wind physicists were upset by our prediction scheme. In general, scientists tend to ignore poor work by others. Since they were upset, we must have done something worthwhile. In fact, to begin with, before Figure 6.3 was published, there had been no quantitative or semiquantitative pattern of the interplanetary corotating structure available except for hand-drawn sketches. Note that one requires the R-t diagram in Figure 6.2 for this particular purpose that requires, in turn, the integration of the equation of motion twice. Solar wind researchers are, in general, not interested in the second integration because satellites and space probes measure only the velocity.

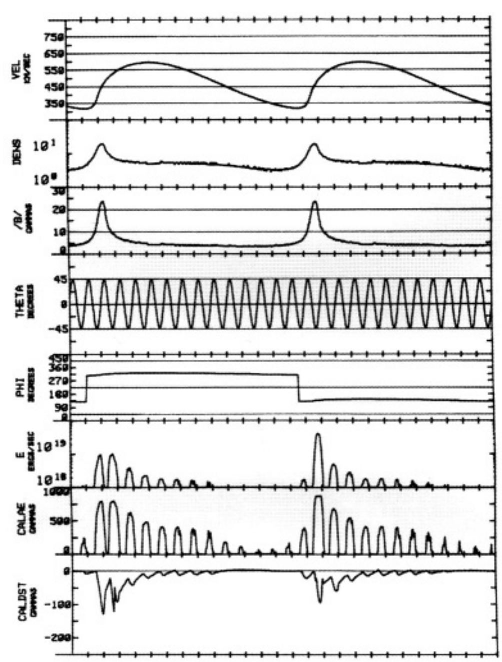


Figure 6.4a The velocity, density, IMF magnitude, IMF latitude angle (THETA), IMF azimuth angle (PHI), ε parameter, calculated AE and Dst indices at the Earth (or any other point at a distance of 1 au) for one rotation period of the Sun (seen from the Earth), namely 27 days for the same conditions as those for Figure 6.1. It is interesting to note that the IMF THETA angle variations assumed to have sinusoidal variations of amplitude 45°, can reproduce reasonably well the observed storm activities associated with the arrival of the sector structures (see Figure 6.4b).

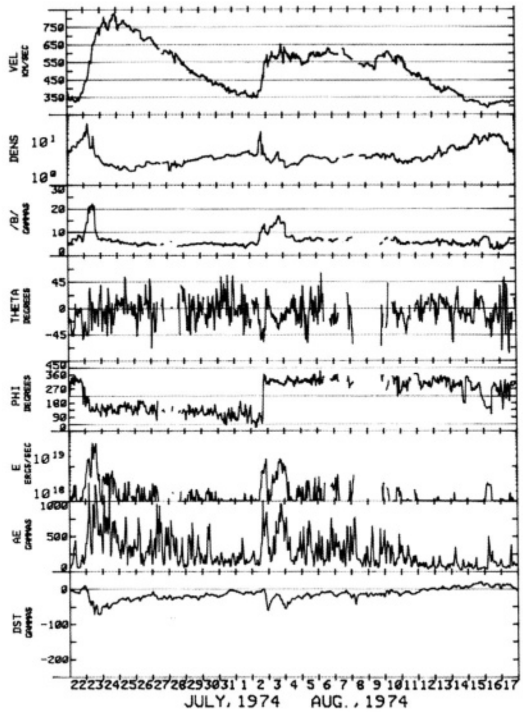


Figure 6.4b A typical 27-day variation of the solar wind speed V, IMF magnitude B, IMF THETA and PHI angles and the AE and Dst indices, when two large coronal holes (separated approximately by 180° in solar longitude), between July 17 and August 12, 1974. These observed variations may be compared with a simulated variation in Figure 6.4a.

6.3 Parameterizing Solar Events

In our scheme, a solar event is represented by a high-speed source area on the source surface, which is superposed on the background structure described in the previous section. The source area is represented by a circular area (or an elliptical area); the speed is highest at the center and has a Gaussian distribution. The speed at the center varies in time in a characteristic way, which is parameterized by

$$V = V_F t e^{-t/\tau_F}$$

Thus, a solar event is parameterized by the maximum speed V_F at the center at the peak of the event, the area size σ_F and the time variations τ_F, together with longitude λ_F and latitude ϕ_F and the start time T_F of the event. Figure 6.5 shows graphically the adopted parameters, $\lambda_F = 180°$, $\phi_F = 0°$, $V_F =$ 500 km/sec, $\sigma_F = 30°$, and $\tau_F = 12$ hrs on the source surface for a hypothetical event. It is assumed that the event takes place on the center of the disk of the Sun at 12 UT, on December 8. Therefore, in this particular case, the center of the Sun, and the location of the solar event and the Earth, are almost on the same solar radial line at the onset of the solar event. If we consider CMEs or sudden filament disappearances, another set of parameters must be needed to describe them. Obviously, the parameterization is crucial for any modeling effort, including a MHD method. This issue has not been discussed much in space weather research; solar physicists are only interested in what happens on the Sun, while magnetospheric physicists consider their problems only after solar disturbances near the magnetosphere.

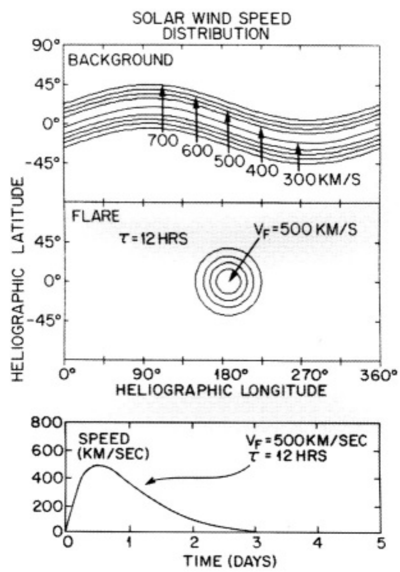


Figure 6.5 From top: The background flow speed of the solar wind on the source surface (the solar longitude-latitude map); the solar wind speed is minimum (300 km/sec) at the magnetic equator and increases toward higher latitudes. The middle diagram represents a solar event; a higher speed area is added to the background flow; the speed reaches V_F= 500 km/sec after flare onset and decreases slowly (τ_F = 12 hrs). The bottom diagram shows how the speed at the center of the circular area in the middle diagram varies in time (expressed by the parameter τ_F = 12 hours in this particular case).

6.4 Modeling the Propagation of Shock Waves

Figure 6.6 shows the propagation of the resulting shock wave in the equatorial plane at 0, 6, 12, and 18 UT, on December 10, at 36, 42, 48, and 54 hours after the hypothetical event on December 8, respectively. The Earth's location is indicated by a star mark. Since the event is assumed to occur on the center of the disk, the center of the shock wave is propagating approximately along the Sun-Earth line. A number of interplanetary observations in the past were modeled to learn about the parameterization. An example is shown in Figure 6.7, in which three successive shock waves are observed by three space probes. Our method can handle a series of flare events. During an active period of the Sun, a number of shock waves generated by flares and the co-rotating structure interact in a very complicated way (Figure 6.7). Our scheme is useful as a guide to interpret solar wind variations during such a period.

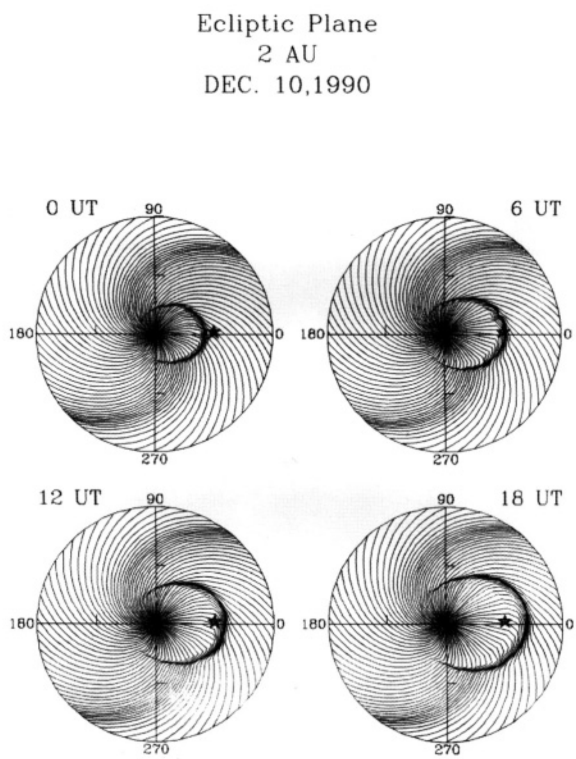


Figure 6.6 The propagating shock wave generated by the growth of the high-speed flow (the middle diagram in Figure 6.5) on the equatorial plane (the circular area of radius 2 au). The Sun is located at the center. The location of the Earth is indicated by a star. The shock wave is represented by the IMF field lines distorted by the propagating shock wave. The resulting solar wind variations at the Earth are illustrated in Figure 6.9.

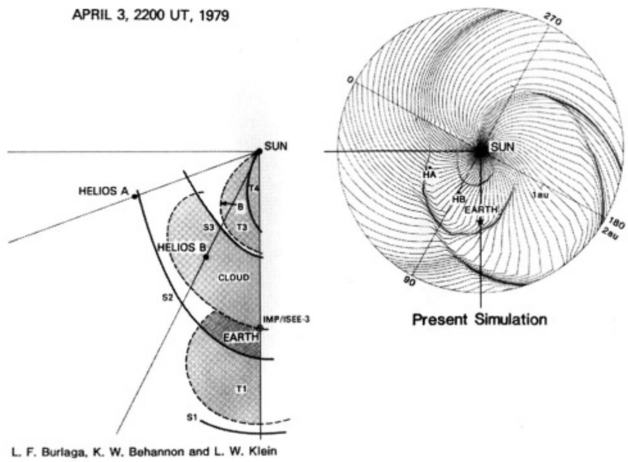


Figure 6.7 Simulation (hind-casting) of the August 1979 events (right), studied by L.F. Burlaga, K.W. Behannon, and L.W. Klein (1981) who inferred the geometry of three shock waves S1, S2, and S3 (left); the geometry of the shock waves in one quarter area of interplanetary space was inferred on the basis of the observations at HELIOS A, B, and the Earth-bound satellites (IMP/ISEE-2): S.-I. Akasofu (1996).

6.5 Detecting Shock Waves by IPS

In predicting a geomagnetic storm after a specific solar event, it is desirable to detect the advancing shocks midway between the Sun and the Earth. A space probe at the midpoint is ideal, but is practically unavailable. For this reason, we have searched for other methods. One of them is to use interplanetary scintillation (IPS). In order to demonstrate this method, a study of an event in September 1978 was undertaken (Akasofu and Lee, 1989, 1990). We constructed successive 3-D surfaces of the shock wave (Figure 6.8a) and projected them onto the *sky map,* a map of the sky centered at the direction of the Sun (Figure 6.8b). The available IPS observation during the event showed an intense IPS area in the upper left of the sky (Figure 6.8c), in agreement with the projection (A. Hewish et al., 1985). As far as I am aware, this was the first comparison of the observed IPS and the model results. Such an observation assures that we chose the necessary parameters of the solar event on the source surface before the arrival of the shock wave reasonably well. Takao Saito and I found that comets between the Sun and the Earth are also useful in calibrating the simulation results, because comet tails tend to show some disturbance when the shock waves hit them. The importance of the integration effort of the four disciplines in space weather research is to lead us to new problems and their solutions. The use of IPS observations and comets is a good example.

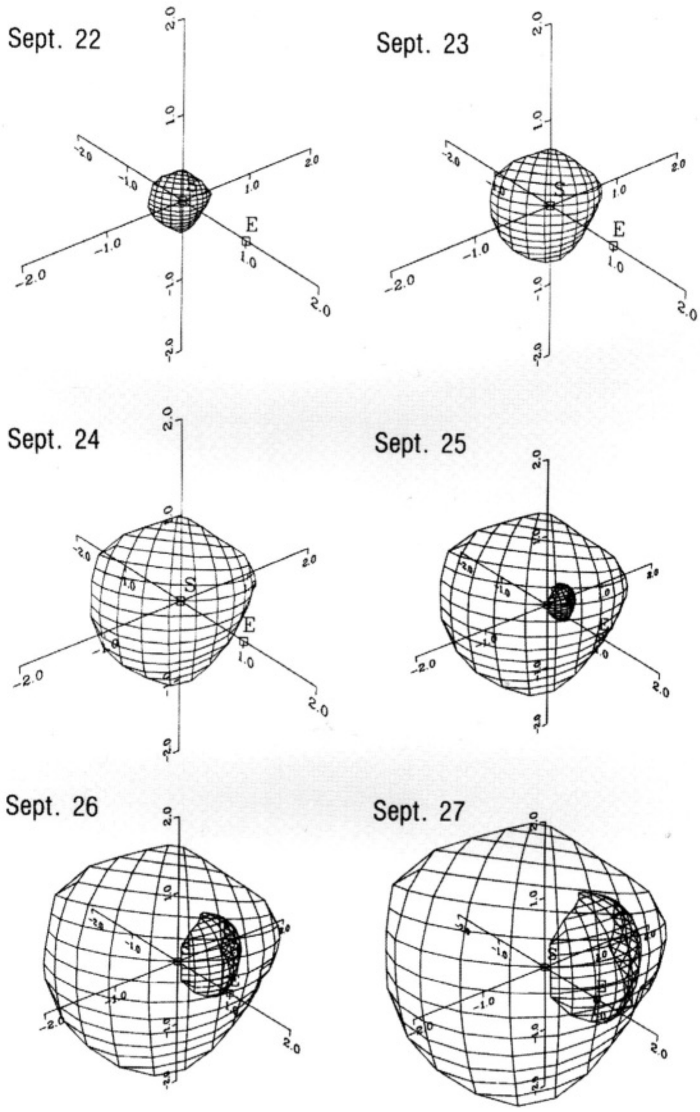


Figure 6.8a Three-dimensional structure of the shock wave for the September 1978 event. The location of the Earth is marked by E.

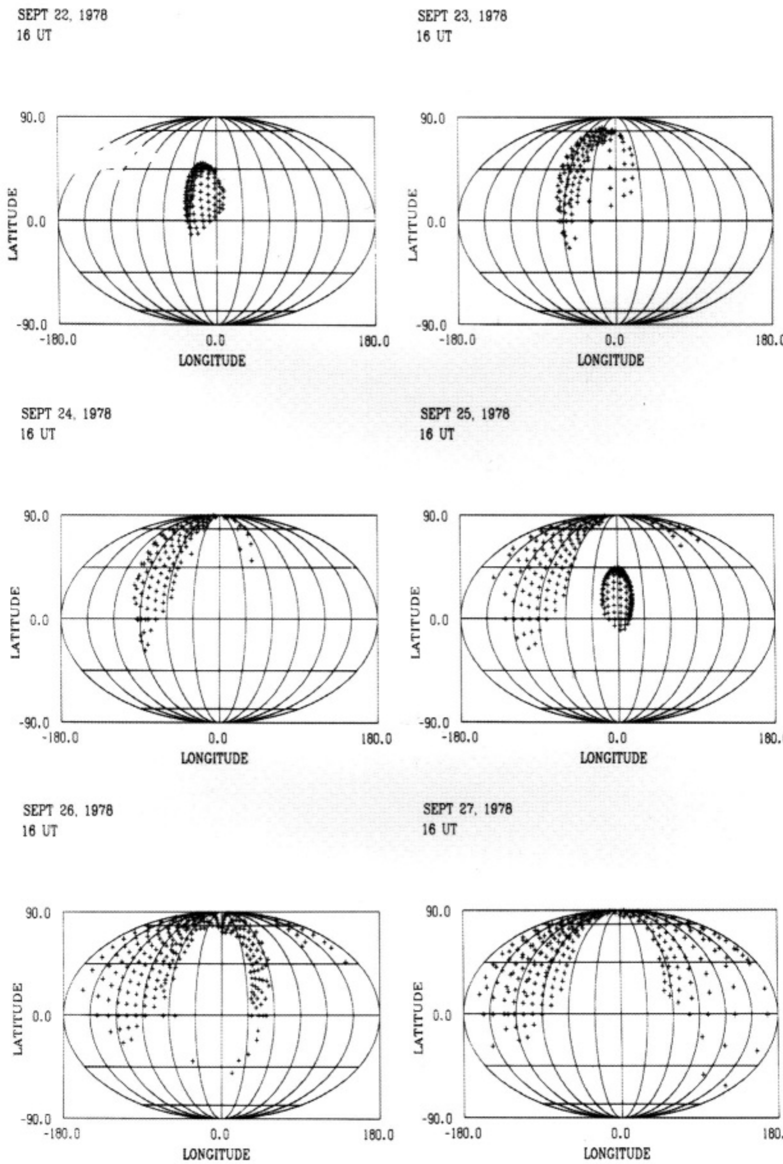


Figure 6.8b The projection of the 3-D shock front on the sky map at 16 UT on September 22-27, 1978. The center of the map is the direction toward the Sun. The shock front is located in the northwestern part of the sky.

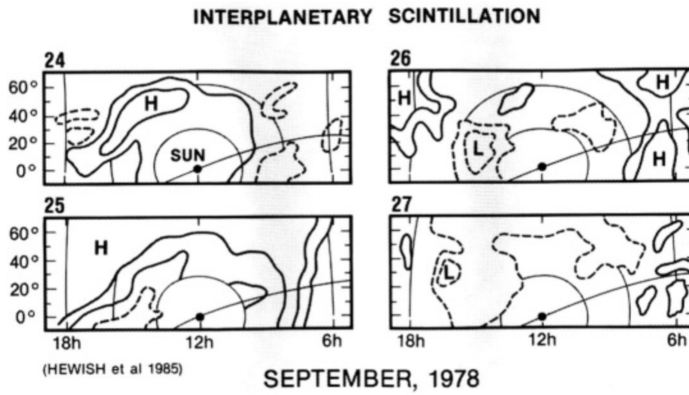


Figure 6.8c The observed IPS sky map for the September 1978 event. A high IPS region is seen in the northwestern part of the sky map on September 24: A. Hewish, S.J. Tappin, and G.K. Faggen (1985).

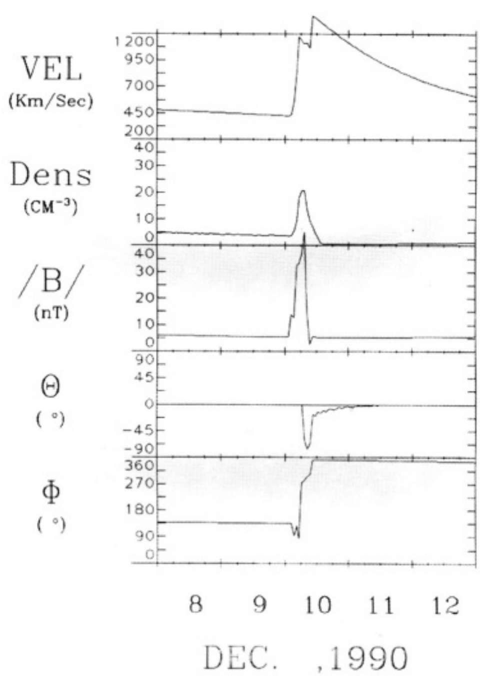


Figure 6.9 The computed solar wind velocity, density, IMF magnitude B, IMF Θ (THETA), and Φ (PHI) angles for the hypothetical event illustrated in Figure 6.5 and 6.6.

6.6 Estimating the Velocity, Density, and IMF at the Earth

Our modeling enables us to predict the solar wind quantities, such as velocity, density, and IMF, at any location within a distance of 2 AU. Figure 6.9 shows these quantities at the Earth for the hypothetical solar event discussed in Section 6.3. A similar estimation can be made at the L5 point or any other points (for example, the L3 and L4 points, which will be occupied by the STEREO mission) in the inner interplanetary space. During the last year or so, a number of comparisons have been made on the predicted arrival time of the shock waves at the Earth by MHD methods and our (HAF) scheme. The results show that the accuracy of both methods is comparable at this time. This result must have surprised some MHD researchers.

6.7 Characterizing Geomagnetic Storms

The prediction of a geomagnetic storm requires the prediction of the Dst and AE indices as a function of time.

The next step is to identify the expression for the electric power that generates the storm components and the resulting storm fields. It is given by (Section 1.9):

$$\varepsilon \text{ (Megawatts)} = 20\, V \text{ (km/sec) } B^2\text{(nT)} \sin^4 (\theta/2) \qquad (1)$$

where θ denotes the IMF polar angle

The first important test of adopting ε (Megawatts = MW) in predicting geomagnetic storms is whether ε can characterize the variety of geomagnetic storms, or more specifically, whether we can infer the two geomagnetic indices AE(t) and Dst(t) as a function of time from ε(t).

Figure 6.10 shows, from the top, ε, calculated Dst, calculated AE and the observed AE for the March, 1973 storm. Knowing that Dst is proportional to the total kinetic energy of the ring current particles (Dessler and Parker, 1959), the calculated Dst can be obtained by:

$$\frac{dDst}{dt} = \alpha\varepsilon - \frac{Dst}{\tau_R} \qquad (2)$$

It is assumed that 70% of the power is dissipated in the ring current, so that α is 0.7; τ_R is the lifetime of ring current particles (7 hrs or less). In terms of ε, the intensity of geomagnetic storms may be roughly classified as follows:

weak storms ε~0.25 x 10^{0} MW (e.g. V = 500 km/sec, B = 5 nT)
moderate storms ε~1.4 x 10^6 MW (e.g. V = 700 km/sec, B = 10 nT)
very intense ε >8.0 x 10^6 MW (e.g. V = 1000 km/sec, storms B = 20 nT)

There is to date no theoretical study that relates ε to the AE index. This is because the magnetosphere responds to an increased ε in two ways, the directly driven component and the unloading component (Section 1.10). The directly driven component correlates fairly well with ε, but the unloading component does not and thus cannot be predicted at this time. The AE index includes both components. Therefore, the empirical relationship between ε and AE has to be established.

$$AE\ (nT) = -300\ (\log \varepsilon)^2 + 11700 \log\varepsilon - 113200 \qquad (3)$$

It can be seen from Figure 6.10 that the Dst variations computed on the basis of ε can reproduce fairly well both the observed characteristics of the storm and its time variations. However, it is obvious that the empirical relationship between ε and AE should be improved. Figure 6.11 shows ε computed for the hypothetical event discussed in Section 6.7. Both the AE (CAE) and Dst (CDST) indices are also computed. The quantity, Φ_{pc}, will be discussed in Section 6.8. The size of the auroral oval can be predicted in a similar empirical way.

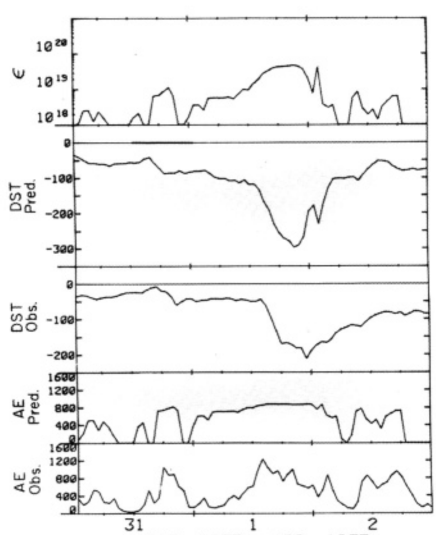


Figure 6.10 From the top, ε computed on the basis of the solar wind observations, the calculated Dst (CAL Dst), observed Dst, calculated AE (CAL AE), and observed AE and Dst for the March 1973 storm.

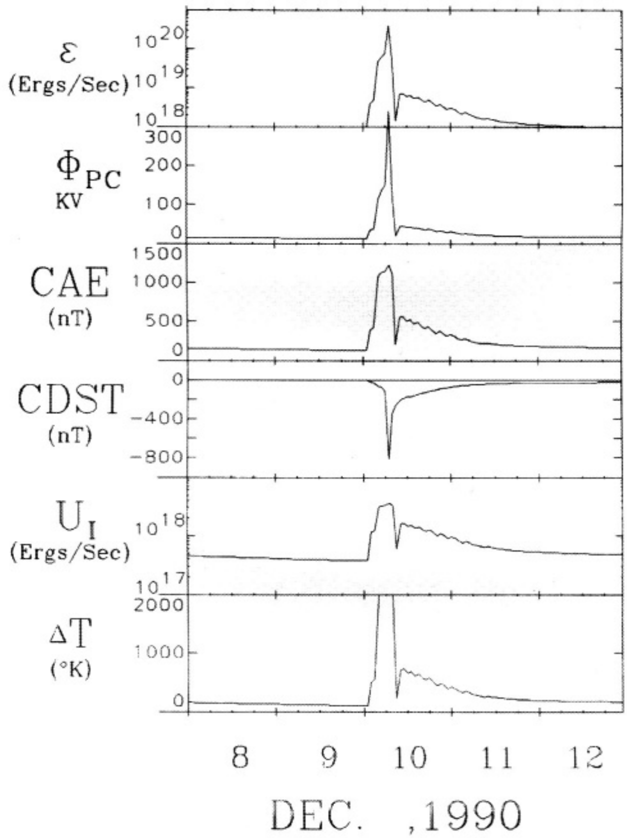


Figure 6.11 The computed ε, polar cap potential (Φ_{pc}), calculated AE (CAE), CDst indices, the ionospheric energy dissipation rate (U_I) and the temperature variation (ΔT).

6.8 Predicting Ionospheric Effects

Pat Reiff and her colleagues found that the ε parameter has a good correlation with the polar cap potential: $\Phi_{pc} = (0.93\ \varepsilon - 3.19)^{1/2}$. This potential drives a flow of ionospheric plasma from the dayside hemisphere into the polar cap region. Sergei Maurits and Brenton Watkins (1996) demonstrated that it is possible to forecast the electron density distribution of the F region in the polar cap on the basis of observed solar wind and the IMF (Figure 6.12).

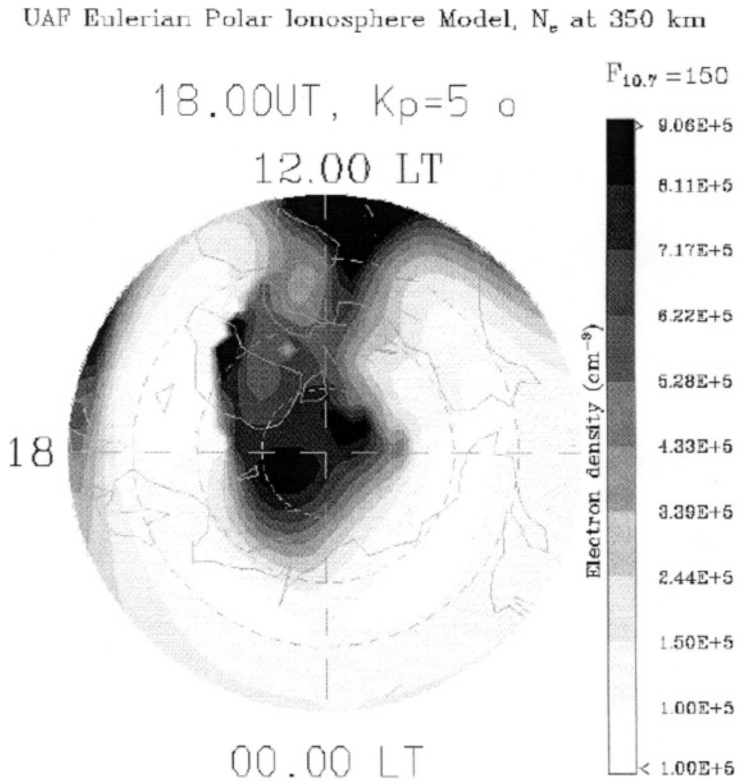


Figure 6.12 An example of the simulation of the (***ExB***) flow of the ionospheric plasma from the dayside hemisphere into the polar cap at an altitude of 350 km (18 UT, Kp = 5_0: S. Maurits and B. Watkins (1996).

6.9 Effects on Power Transmission Lines and Oil/Gas Pipelines

Although it has been emphasized by the space science community that solar events could cause serious problems on power transmission lines, there have so far been only a few studies to examine how geomagnetic storms can actually affect them. Bob Merritt, John Aspnes, and I (1979, 1982) demonstrated that changing magnetic fields produced by the auroral electrojet induce electric currents in the neutral line of a three-phase transmission line and that such extra currents are converted into pulse signals in the circuit breaker system (Figure 6.13). It is generally very difficult to obtain such data from a power company. It was fortunate that many of the engineers at the local power company (Golden Valley Electric Association) were University of Alaska graduates. They were Bob Merritt's former students and were very helpful for this project. It can be seen that large impulsive changes of the

Earth's current tend to produce pulses in the relay system. However, a simple mathematical relationship between them could not be found.

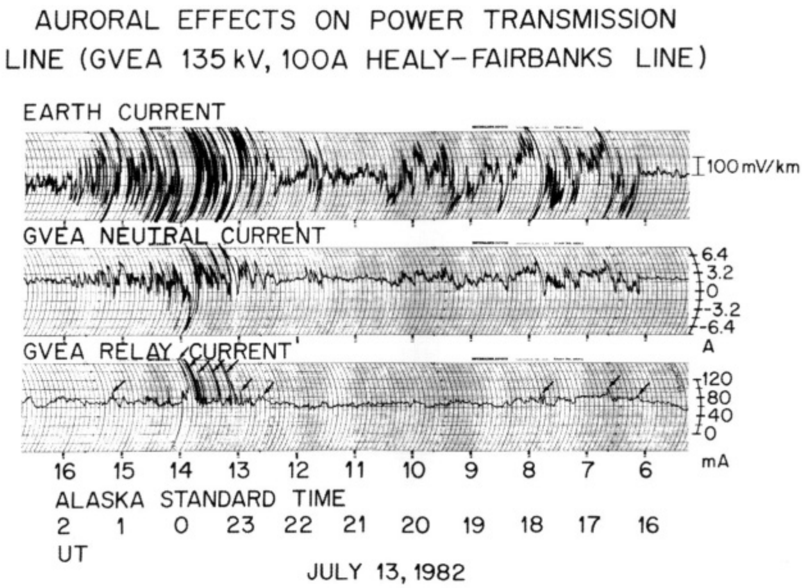


Figure 6.13 From the top, the Earth current recorded at the Geophysical Institute, University of Alaska, electric currents induced in the Golden Valley Electric Association (GVEA) power transmission line and the relay current in the GVEA substation on July 19, 1982: S.-I. Akasofu and J. Aspnes (1982).

During the construction of the Trans-Alaska Oil Pipeline, I wrote to the president of the pipeline company, saying that the aurora can induce strong currents in the pipe. He responded by saying that his corrosion engineers knew what to do about the problem. However, after the pipeline company asked us to monitor the induced current during the construction, Gene Wescott, Bill Sackinger, Bob Merritt, and I found a rather simple formula; the current I in the pipe is given by $I = V/\Omega$, where V is the induced voltage (~ 1 volt/m) for moderate auroral activity (the total length of the pipe ~1000 km) and Ω the total resistance of the pipe. Thus, I ~100 amperes for V = 1000 volts and Ω = 10 ohms. Currents leaking from the pipe to the ground cause corrosion of the pipe. Our finding is now used to monitor the corrosion of the Trans-Alaska Oil Pipeline (Figure 6.14).

6.10 Missing Links

In examining the variability of the solar wind speed *V*, the IMF magnitude *B* and the IMF polar angle θ, the most variable quantity is θ. In order for ε to be greater than 1 million MW, for V = 500 - 1000 km/sec and B ~ 10 nT, it is necessary for θ to be greater than ~90°. This situation is generally called the southward turning of the IMF. It is likely that intense solar events produce a high value of V ~ 500 - 1000 km/sec and B >10 nT. However, if θ happens to be 0° or very small, ε cannot reach 10 million MW.

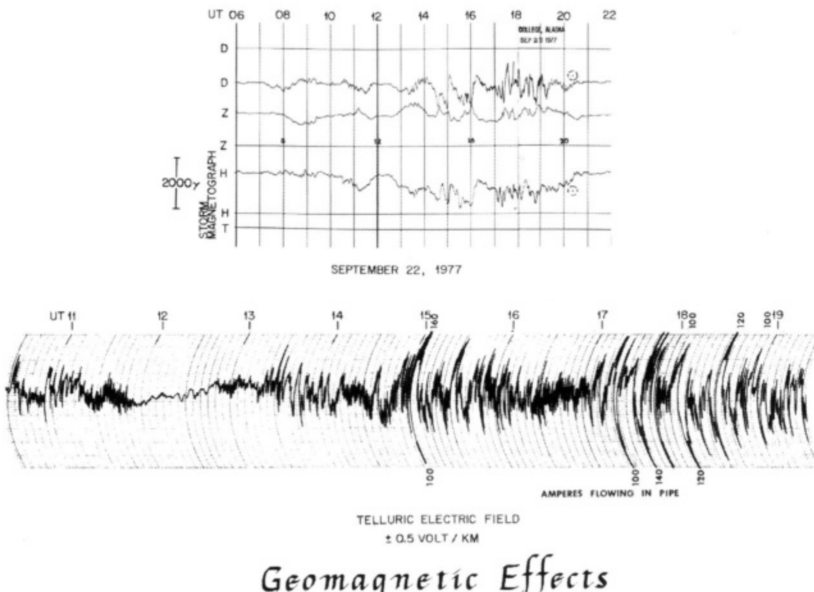


Figure 6.14 Upper: College magnetic record on September 22, 1977. Lower: Electric current in the Trans-Alaska Pipeline on the same day: E.M. Wescott, W. Sackinger, and S.-I. Akasofu.

It is thus obvious that we cannot succeed in predicting geomagnetic storms until we can find a way to predict $\theta(t)$. Thus, the prediction of $\theta(t)$ after a specific solar event is most crucial in predicting geomagnetic storms.

Recently, IMF changes associated with solar events were discussed in terms of magnetic *clouds, magnetic flux ropes, loops, tubes,* etc. Perhaps some efforts are needed to standardize such terms. For example, if the structures are detached magnetically from the Sun, they may be called *clouds,* while if the structures are magnetically anchored to the Sun, they may be called *loops.*

Three important issues in this regard are how these structures are related to:

1. Coronal Mass Ejections (CMEs),
2. whether the expanding CMEs constitute the so-called *driver gas,* which is responsible for generating the shock wave, and
3. if it would be at all possible to predict changes of the IMF polar angle θ as a function of time on the basis of solar observations.

Bruce Tsurutani, Francis Tang, and I examined several possibilities for the causes of θ changes (Tsurutani et al., 1988). Most sunspot pairs align in the east-west direction. However, in some exceptional cases, the pairs tend to align in the north-south direction. It was our finding that the orientation of the sunspot field is unrelated to that of the IMF at the Earth's location. Figure 6.15 shows such an example. In this case, a sunspot pair aligned in the north-south direction. Thus, if the sunspot field would simply expand toward the Earth, we would expect a southward-oriented field at the location of the Earth. However, the observed field was pointing northward. Therefore, the magnetic structure within an enhanced solar wind is not a simple expansion of magnetic loops anchored in the vicinity of active spots.

Further, there occur often several very sharp changes (in time) of the azimuth angle of the IMF during a major geomagnetic storm, suggesting sometimes a large-scale movement of the heliospheric current sheet. Figure 6.16 shows how a shock wave associated with a solar event in the southern hemisphere can push up the current sheet, so that the Earth's position with respect to the current sheet (above, before the passage of the shock) can change (below, after the passage). Some of the changes of θ may be related to flapping of the current sheet, because the field lines associated with the current sheet are supposed to be parallel to it. Further, the passage of the shock wave can increase the magnitude of the IMF, so that ε can be increased if $\theta > 90°$.

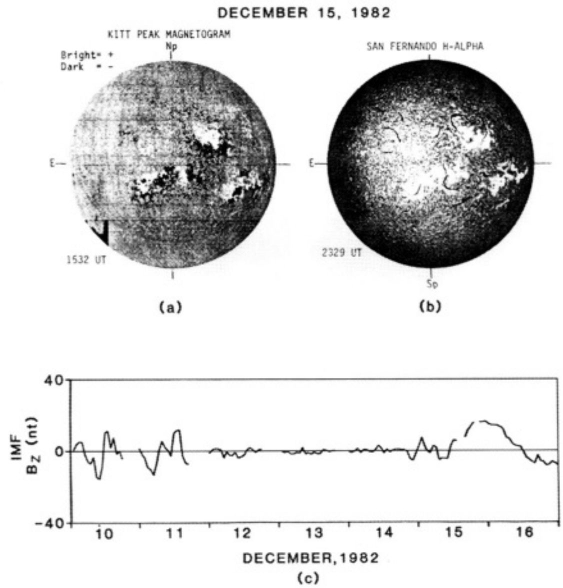


Figure 6.15 A north-south oriented sunspot pair on December 15, 1982 was associated with a major flare and a major change in the IMF. If the sunspot field expanded, it was supposed to cause a southward-oriented IMF. However, at the location of the Earth, the observed IMF was oriented in the northward direction: F. Tang, S.-I. Akasofu, E. Smith, and B. Tsurutani (1985).

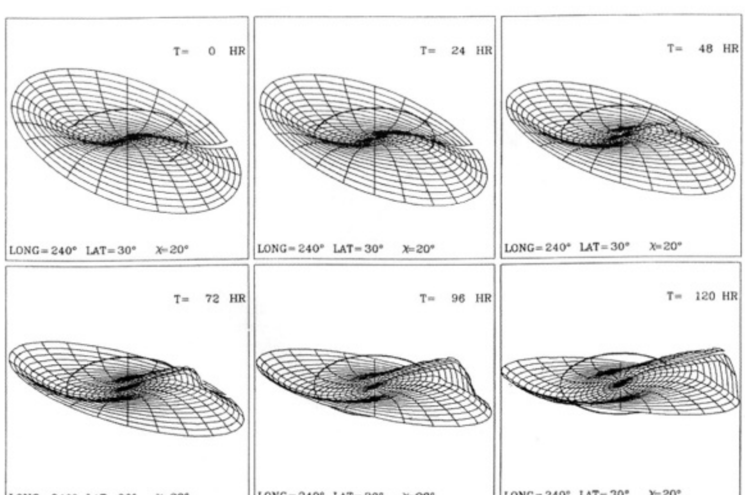


Figure 6.16 Deformation of the heliospheric current sheet caused by a hypothetical flare in the southern hemisphere: C.D. Fry and S.-I. Akasofu (1985).

6.11 High-speed Streams from Coronal Holes

It was mentioned in the Prologue that E.W. Maunder found a high-speed stream from a spot-free area (his ninth statement). Large coronal holes tend to be an extension of the polar coronal holes (Figure 6.17). We have simulated high-speed streams from equatorial coronal holes in Figure 6.3 and 6.4a.

The two streams in one solar rotation occur because the sinusoidal magnetic equator is a representation of the extension of two coronal holes, one from the northern and the other from the southern polar region (Figure 6.17). Indeed, in a typical situation, two coronal holes occur that are separated by 180° in longitude. As a result, two high-speed streams appear, causing two peaks of speed of the solar wind during one solar rotation (Figures 6.4a and 6.4b).

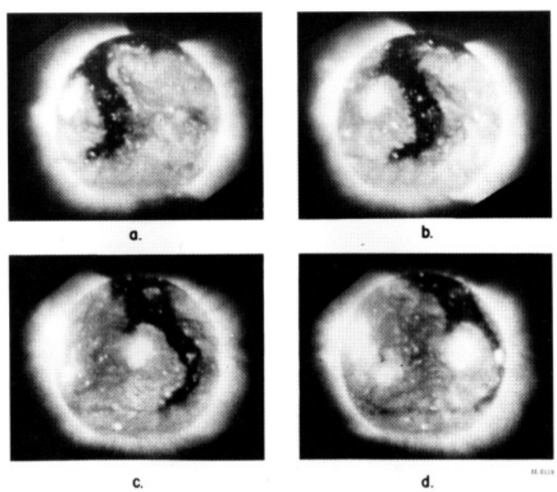


Figure 6.17 A typical coronal hole. As the Sun rotates, it rotates with the Sun, so that it will be seen every 27 days: NASA Sky Lab photograph.

Coronal holes tend to appear a few years after the maximum phase of a sunspot cycle. The resulting high-speed streams rotate with the Sun and sweep by the Earth; it takes about one week to pass by the Earth. As a result, the Earth tends to be immersed in the high-speed stream twice, each time for one week, in each solar rotation. The resulting auroral activity occurs twice (each lasting for 10 days) during each solar rotation, a period of 27 days (see Figure 6.4b). Thus, the auroral activity tends to peak well after the period of the sunspot maximum (Figure 6.18). Although this fact has been well known for a long time, it appears that it has recently been forgotten. Thus, as a result, many solar physicists and auroral physicists claim that the sunspot maximum years coincide with the auroral peak year. This is not the case in high latitudes. It is unfortunate that auroral activity caused by the solar wind from coronal holes has to be rediscovered every eleven years.

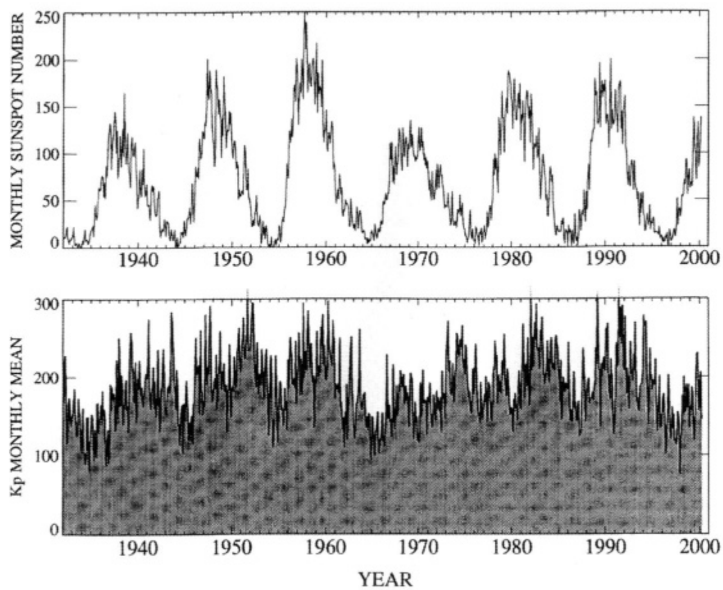


Figure 6.18 Upper: The sunspot cycles. Lower: The Kp index that represents auroral activity. Note that the auroral activity represented by the Kp index tends to lag behind changes of the sunspot number: S.-I. Akasofu.

6.12 A New Example

A large sunspot group appeared near the eastern limb of the Sun during the last week of March 2001 and reached the central meridian on about March 28. Almost 20 solar transient activities occurred from March 28 to April 18. Some of them caused intense interplanetary shock waves. An advanced version of the HAF model was used to simulate the propagation of the shock waves. The results are compared with the observations at the libration point. Figures 6.19a, b, c, show the solar conditions on March 29, 2001. Details of the sunspot group are shown in Figure 6.20.

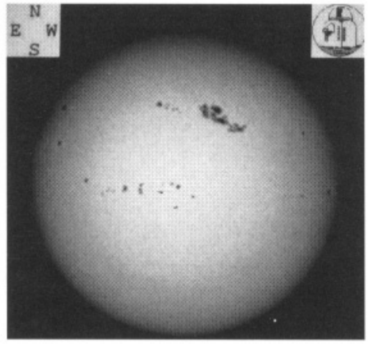


Figure 6.19a Visible image on March 29, 2001: Big Bear Solar Observatory.

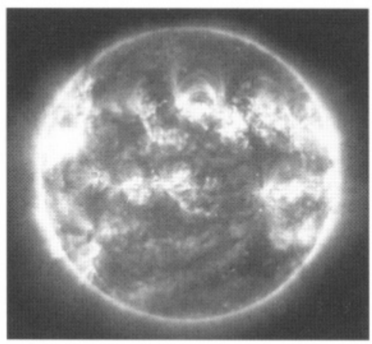


Figure 6.19b SOHO Image on March 29, 2001: SOHO Project.

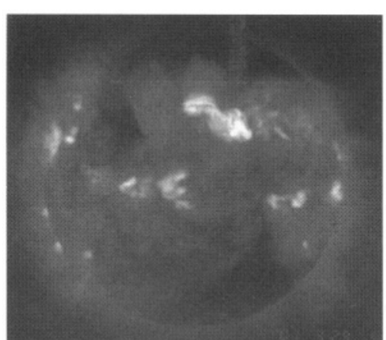


Figure 6.19c YOHKOH Image on March 29, 2001: YOHKOH Project.

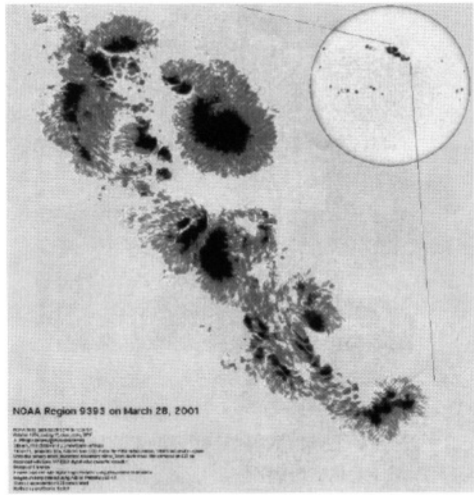


Figure 6.20 One of the largest sunspot groups that appeared Sunspot Cycle 23 (NOAA Region 9393): D. Rust (2001).

One of the largest transient events occurred at 1004 UT on March 29. It was an X1.7 flare at longitude 12° West and latitude of 16° North on the solar disk. In Figure 6.21, the simulated patterns of the interplanetary condition to a distance of 2 AU are shown for the series of events between 29 March and 22 April 2001, using the interplanetary magnetic field lines. In each circular pattern, the familiar spiral pattern and its deformation by the propagation of the shock waves are shown. The red lines show the field vector pointing away from the Sun, while the blue lines show those pointing toward the Sun.

The shock wave caused by it overtook the slower shock wave produced by an early event and they merged together before reaching the Earth. The combined shock arrived at the Earth at 0021 UT, on 31 March 2001, and caused one of the most intense geomagnetic storms since the last decade (minimum Dst was -358 nT). The following patterns show the interplanetary conditions at the arrival time of the other shock waves.

Figure 6.22 shows the comparison between the observation by the ACE satellite and the simulation results at L1. The first two diagrams show the observed speed and density as a function of time and the simulation results by using the initial speed given by the Type II radio emissions. There is a considerable disagreement between them.

By assuming that the main disagreement arises from the initial speed of the driver gas/shock wave, we adjusted the initial speed was adjusted until the predicted *ex post facto* shock arrival times are the same as the observed ones. This is because the initial speed appeared to be the most uncertain factor in the parameterization. The third and fourth diagrams show the simulation results obtained by adjusting initial speed, together with the observations. Figure 6.23 is constructed on the basis of this adjustment. This is a learning process in space weather prediction. Figure 6.23 shows the Dst index for the same period.

It is obvious that there is an inevitable uncertainty in identifying the correspondence between a particular solar event and a particular magnetospheric event, until a direct observation of the shock waves between the Sun and the Earth becomes possible. Until then, we must make our best effort for their identification.

The initial parameterization is crucial in studying the propagation of transient solar events, regardless of the simulation models adopted. Although the disagreement between the observation and simulation results may depend on many factors, it appears that adjusting the initial speed alone can make a significant improvement.

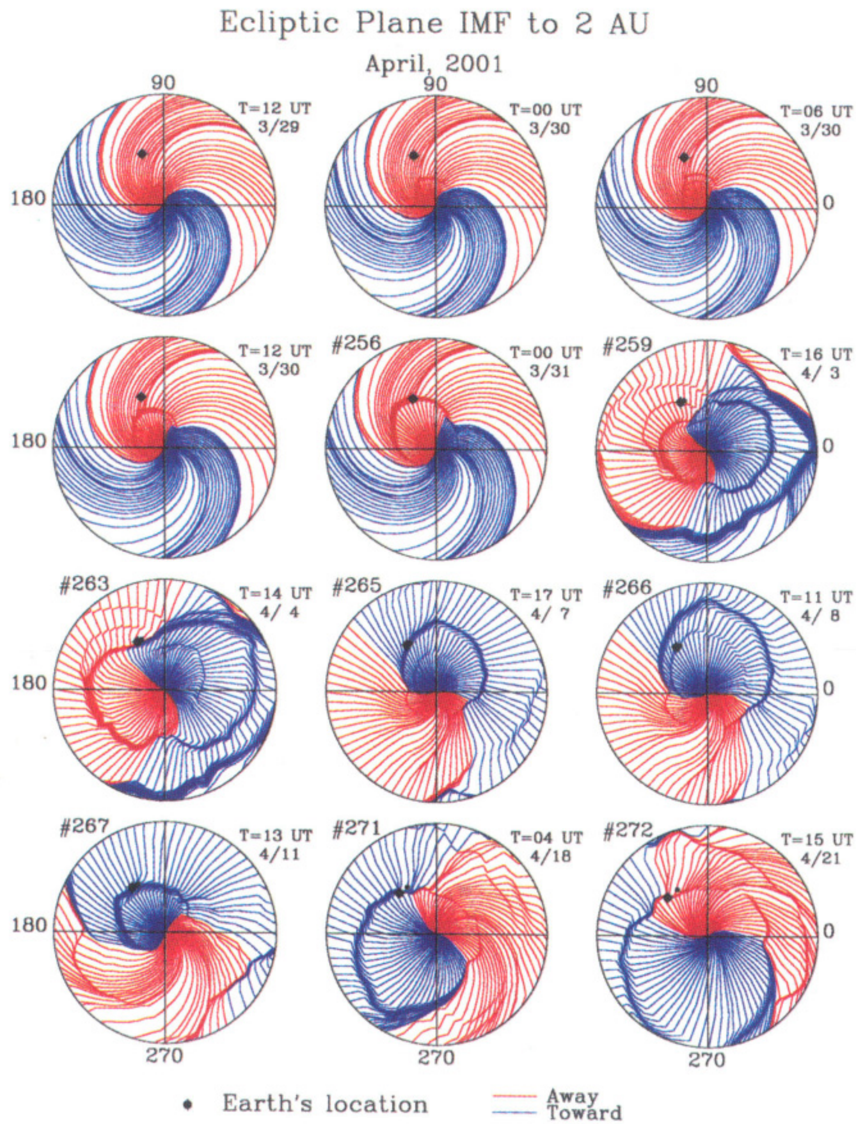


Figure 6.21 A series of shock waves that propagated in interplanetary space during the period of March 29 and April 21, 2001. The location of the Earth is indicated by a star mark. The first five patterns show the propagation of the shock wave generated by the March 29 flare: W. Sun, M. Dryer, C. D. Fry, C.S. Deehr, Z. Smith, S.-I. Akasofu, M.D. Kartalev, and K.G. Grigorov (2002).

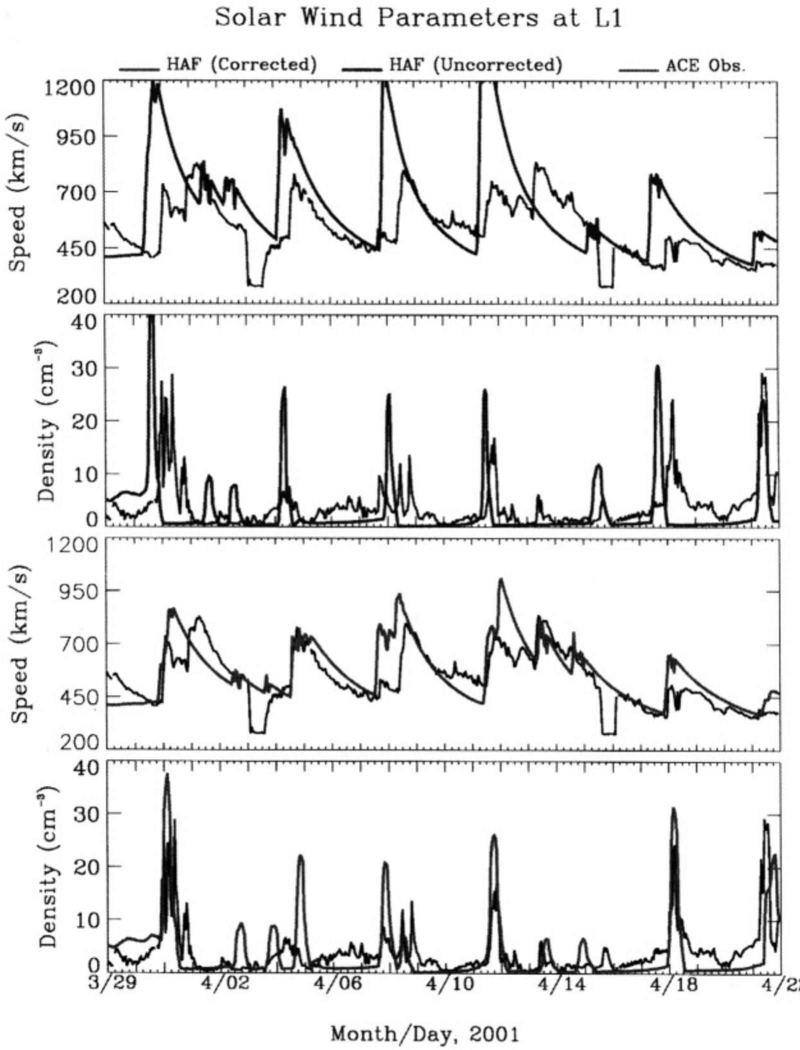


Figure 6.22 Comparisons between the observed shock waves and the simulation results. The upper two rows show the simulation without adjusting the initial speed of the shock, and the lower two after the adjustment: W. Sun, M. Dryer, C. D. Fry, C.S. Deehr, Z. Smith, S.-I. Akasofu, M.D. Kartalev, and K.G. Grigorov (2002).

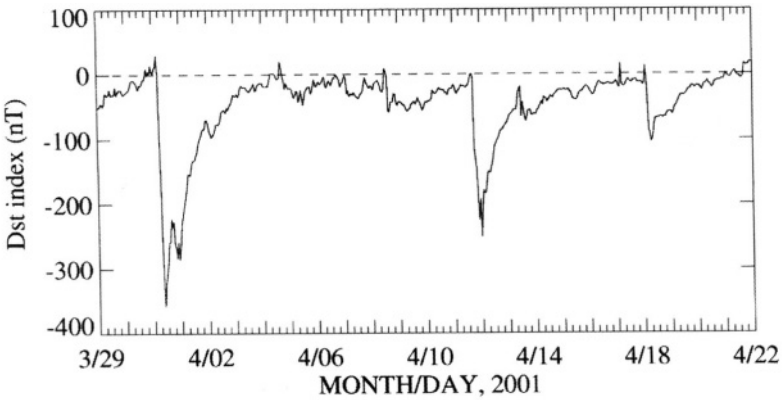


Figure 6.23 A series of geomagnetic storms associated with April 2001 events: Geomagnctism Data Center, Kyoto University.

6.13 Summary

As a summary, Figure 6.19 shows the geomagnetic storm prediction scheme presented in this chapter in a block diagram form. It is satisfying to see that many solar physicists and magnetospheric physicists have started to work together to study space weather jointly by learning the others' discipline. However, many more concerted efforts are needed for the success in the prediction of geomagnetic storms. In particular, we should focus our effort in predicting IMF $\theta(t)$; this requires the knowledge of the interplanetary magnetic structure associated with the clouds and loops.

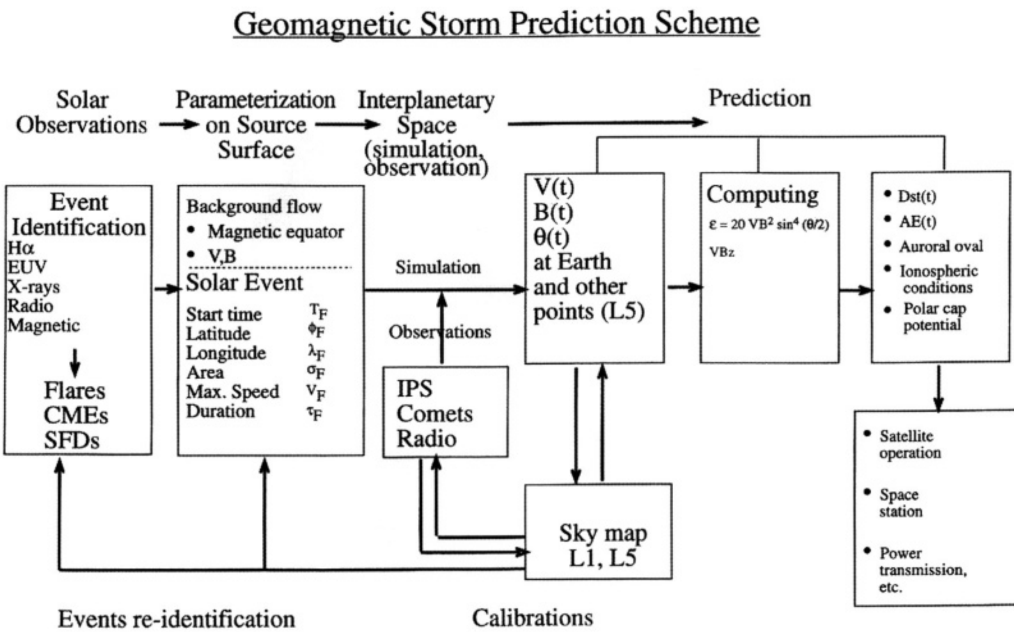
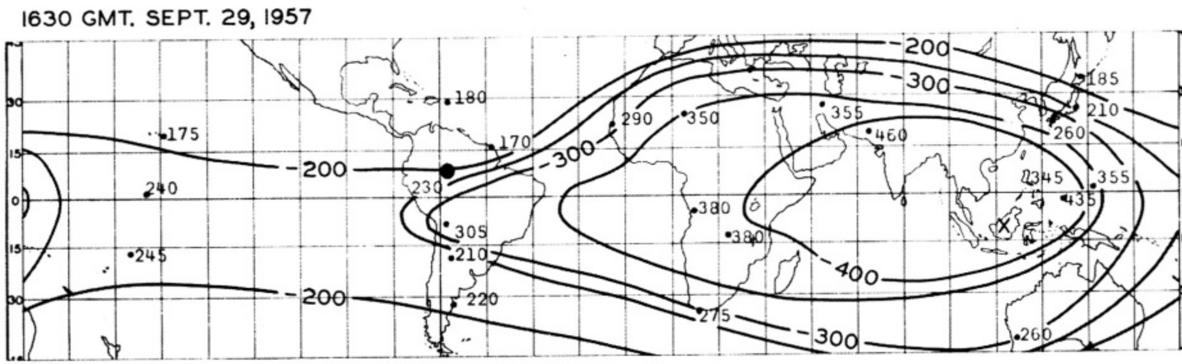


Figure 6.24 The geomagnetic storm prediction scheme in a block diagram form.

There is one subject that has not been mentioned in this chapter. The main phase of geomagnetic storms tends to develop asymmetrically. Figure 6.25a and 6.35b show three examples of the asymmetric development during two intense storms. Actually, this phenomenon was described in terms of the DS component by Chapman in 1918. It was taken up later by Akasofu and Chapman (1964). As mentioned in Section 3.1.3, Alfvén attempted to explain the asymmetry in terms of the field-aligned currents associated with the electrojets. Although the substorm current system can explain a small portion of the asymmetry, it cannot explain the fact that the asymmetry is larger in lower latitudes, as well as the magnitude of the asymmetry (Figure 3.8). It appears that O^+ ions are preferentially injected into the ring current belt in the evening sector where the field-aligned currents are directed upward. In Figure 6.25a, it can be seen that the magnitude of the asymmetry is as large as 150-200 nT.



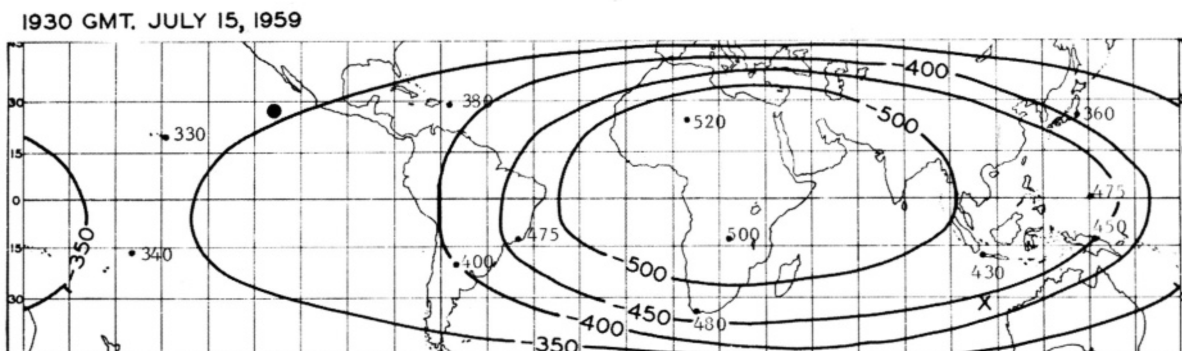


Figure 6.25a The distribution of the main phase field (the horizontal component) for the two intense storms of September 29, 1957 and July 15, 1959. The subsolar point is indicated by a black circle and the anti-subsolar point by a cross: S.-I. Akasofu and S. Chapman (1964).

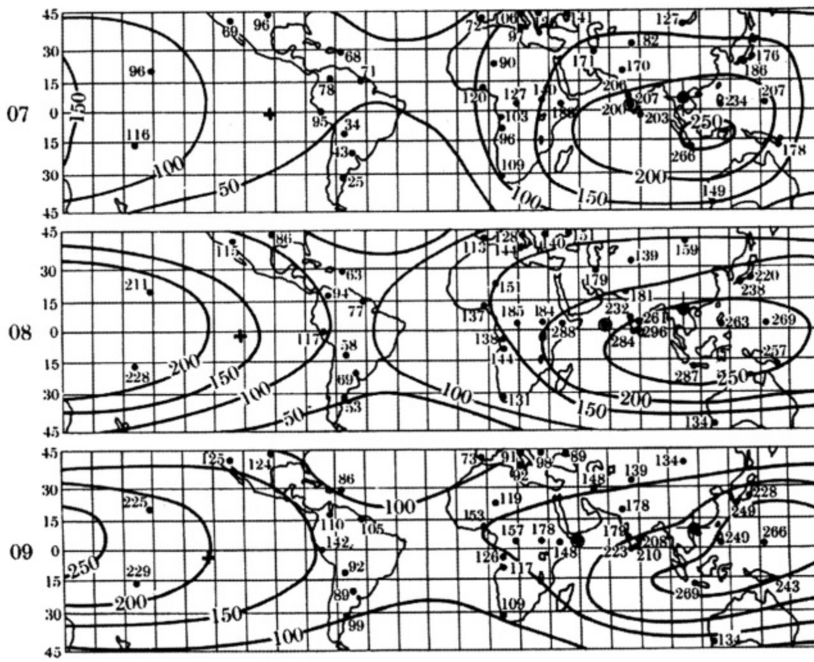


Figure 6.25b The distribution of the main phase field (the horizontal component) for the storm of April 18, 1965. A dot with a circle and a cross denotes the subsolar point and anti-subsolar point, respectively: S.-I. Akasofu and S. Chapman (1972)

This page intentionally left blank

Chapter 7

BEYOND THE INNER HELIOSPHERE: THE MAGNETIC FIELD STRUCTURE OF THE OUTER HELIOSPHERE: A Three-Dimensional Model

Establishing the three-dimensional structure of the heliosphere is an important problem in space physics. Some postulate that the heliosphere has a magnetosphere-like structure as it moves through interstellar gas. On the other hand, we are aware that the magnetic field structure in the heliosphere deviates considerably from a dipolar field. One of the purposes of this chapter is to suggest that a crude model of the magnetic field structure can be made by considering the distribution of electric currents in the heliosphere, as suggested by Hannes Alfvén (1977). Figure 7.1 shows the distribution of the heliospheric current.

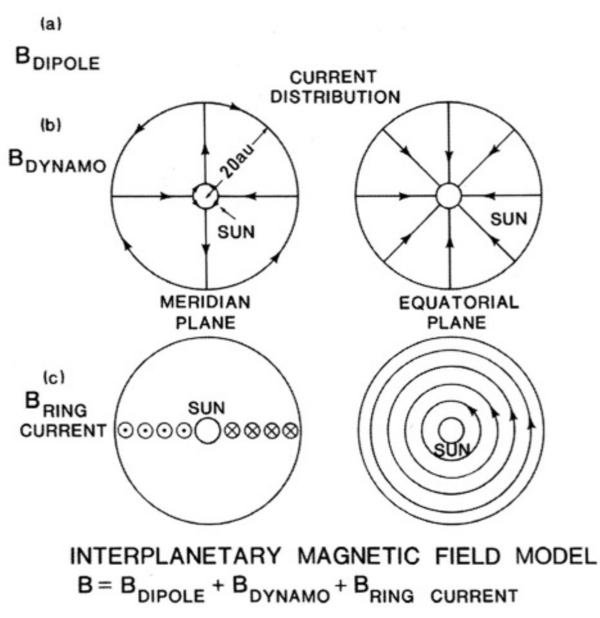


Figure 7.1 Heliospheric current system: S.-I. Akasofu and D.N. Covey (1981).

One can, indeed, consider that the familiar Parker spiral arises from two currents, both the radical (J_r) and azimuthal (J_ϕ) components, flowing perpendicular to the spiral magnetic field lines (Akasofu et al., 1980). We first consider the case when the solar main dipole is directed southward (the north polar region being the south pole). The unipolar induction generates the radially outward-flowing current from both polar regions (Alfvén, 1950, 1977). After reaching the pole of the heliosphere, the current flows along the outer surface (assumed to be spherical) of the heliosphere to the equator and then flows radially (J_r) toward the Sun.

The sector boundary can be taken as a thin current sheet where the current flows azimuthally on the equatorial plane eastward when the solar main dipole is directed southward); the IMF has the sunward component above and the antisunward component below the current sheet. In fact, the boundary is often called the heliospheric current sheet. Thus, the basic heliospheric magnetic field is composed of the magnetic fields produced by these currents, together with the intrinsic solar dipolar field. When the solar dipole reverses its polarity (because of the eleven-year cycle variations), the direction of the currents reverses as well. Note that the current continuity and $\nabla \bullet \boldsymbol{B} = 0$ conditions are automatically satisfied in this method.

The solar system is also embedded in the interstellar magnetic field of the Orion arm. Therefore, the heliospheric magnetic field interacts with it. The magnitude of the field is inferred to be about 0.22 nT. The orientation of the field with respect to the heliosphere is not known. In modeling the heliospheric magnetic field, it is assumed that the orientation is vertical to the solar equatorial plane and directed southward, but only a small fraction of it (~ 5%) is allowed to penetrate into the heliosphere. Note that only a small fraction of the interplanetary magnetic field can penetrate into the magnetosphere.

Figures 7.2a and 7.2b show an example of modeling of such a situation. As far as I am aware, Figures 7.2a and 7.2b are the first of their kind to show the three-dimensional structure of the Parker spiral, although it is crude. So far, all I have seen are hand drawings. Ed Smith and his colleagues (1995, 1997) reported, on the basis of their Ulysses observations, that the spiral structure is present at high latitude in the heliosphere. The field tends to have a spiral structure, although it tends to be more radial than what Parker's model predicts. This may be because the solar wind speed is known to be higher in higher latitudes, producing an *underwound* condition. However, this particular model cannot handle such a situation. Depending on the orientation of the solar dipolar field, some of the high-latitude hemispheric magnetic field lines may be connected to the interstellar magnetic field lines (Figure 7.2b).

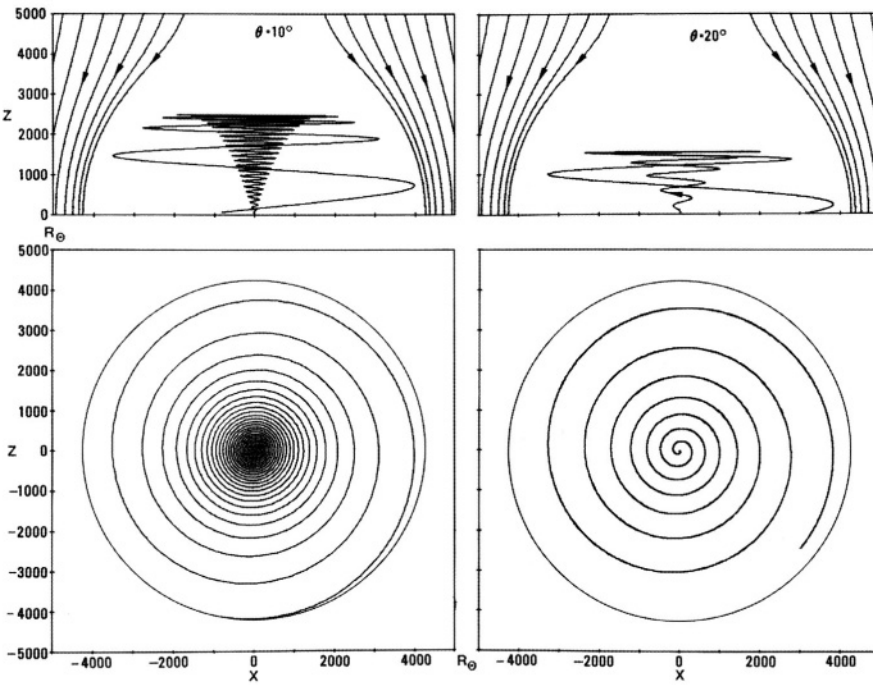


Figure 7.2a Heliospheric current structure: S.-I. Akasofu and D.N. Covey (1981).

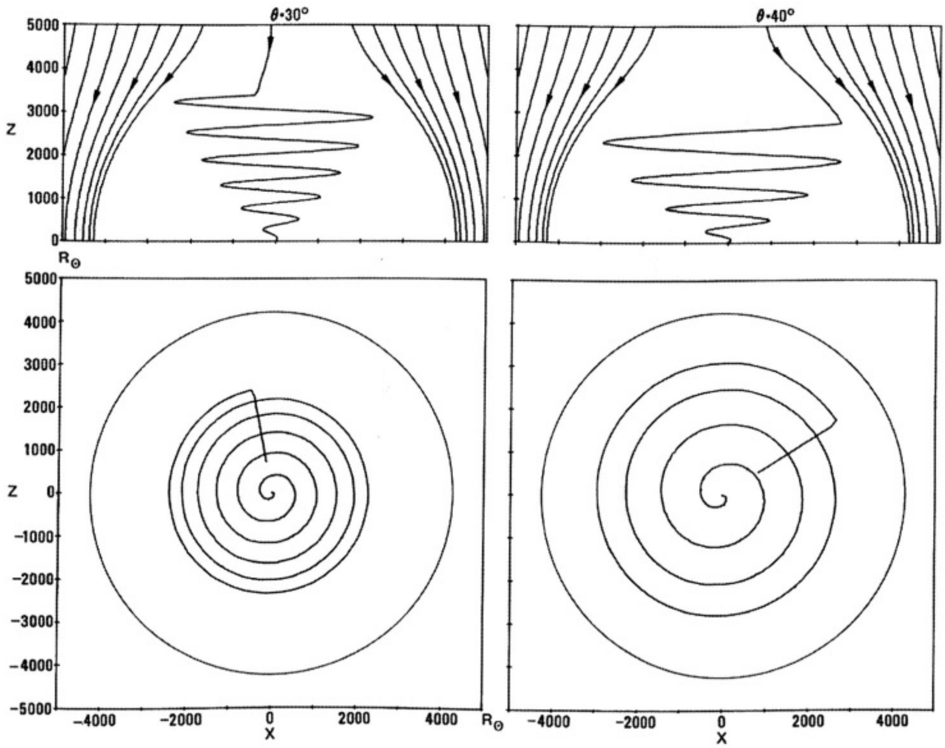


Figure 7.2b Heliospheric current structure: S.-I. Akasofu and D.N. Covey (1981).

As described in Section 6.1, a magnetic field line originating from the source surface can be traced by following particles leaving a particular point on the rotating source surface. The resulting interplanetary magnetic field structure, together with the corotating interaction region produced by the formation of the shock wave structure, is the familiar Parker spiral. Figure 7.3 shows the magnetic field configuration on the equatorial plane within a distance of 20 AU by assuming the magnetic equator has a sinusoidal structure of amplitude of 20° in latitude.

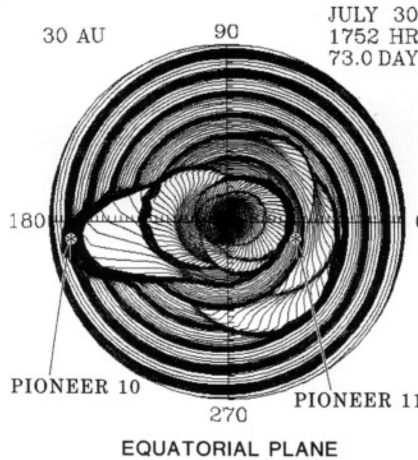


Figure 7.3 Heliospheric magnetic field 73 days after a series of major solar activities in June-August 1982: S.-I. Akasofu, W. Fillius, W. Sun, C.D. Fry, and M. Dryer (1985).

Solar disturbances considerably distort the background structure illustrated in the figure in the Prologue. As an example, simulation results of a series of April-June 1978 and June-August 1982 events in Figure 7.3 are shown. There were seven successive major flares in the period.

It can be seen that the seven shock waves and the background sector structure form a very complicated magnetic structure in the outer heliosphere. Based on Pioneer 10 and 11 missions, Smith (1990) noted that the observations are consistent with such models. Both Pioneer 10 and 11 were in the outer heliosphere during these periods, so we could make some comparison between the simulation and the observations. It is somewhat surprising that our crude method reproduced well the observed shock structure for up to ten days after the shock wave was generated near the Sun. As far as I am aware, this is the first semiquantitative simulation of the disturbed condition of interplanetary space caused by a successive solar activity.

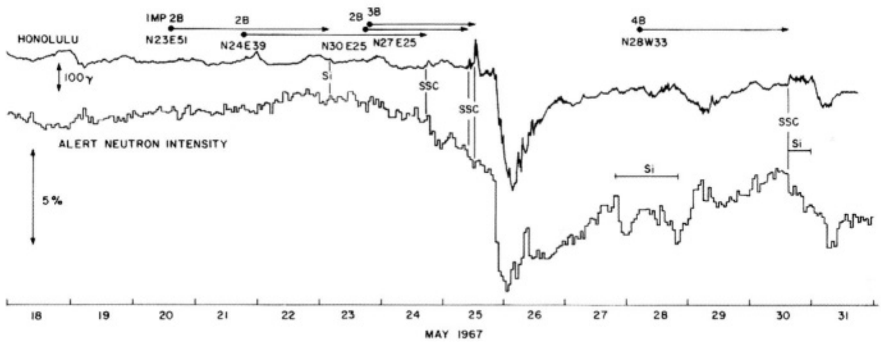


Figure 7.4 Honolulu magnetic record and Alert neutron intensity, together with solar activity during the period of May 18-31, 1967, showing a Forbush decrease: S. Yoshida, N. Ogita, and S.-I. Akasofu (1971).

That the shock structures are expected to have considerable effects on the propagation of galactic cosmic rays from interstellar space to the center of the heliosphere (Van Allen, 1996). Indeed, a significant decrease of cosmic ray flux was observed at the location of both space probes, as well as on the Earth (Akasofu et al., 1985). This phenomenon is called the *Forbush decrease*, honoring Scott Forbush, who discovered it. Figure 7.4 shows an example of a large Forbush decrease caused by solar activities on May 23-24, 1967.

It is also well known that the cosmic ray intensity has a clear anticorrelation with the eleven-year cycle of solar activity. This phenomenon was considered in the past in terms of the degree of diffusion of cosmic rays from interstellar space. Another possibility is that the eleven-year cycle variations of cosmic ray intensity result from such successive sweeping effects of cosmic ray particles by the shock waves. It is interesting to note that the Forbush decrease is not uniform around the Earth. Figure 7.5 shows this nonuniformity. Sekiko Yoshida and I worked on this particular feature, although the space physics community has not shown much interest in it. Since each shock wave tends to sweep away cosmic rays from a limited direction, such an anisotropy may be a good indication of effects of the shock waves.

The magnetic field structure of the outer heliosphere needs much more modeling effort than in the past, because space probe observations are very limited. It is our hope that our effort is of use in providing some idea about the geometry of the heliospheric magnetic field structure during quiet and disturbed periods.

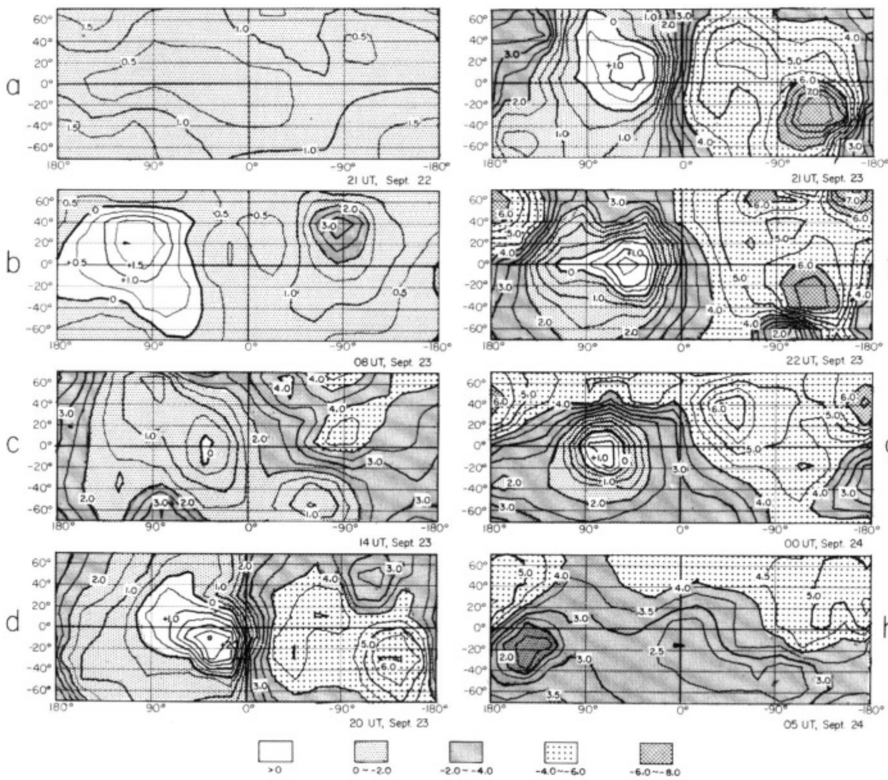


Figure 7.5 The distribution of cosmic ray intensity variations: S. Yoshida, N. Ogita, S.-I. Akasofu, and L. Gleeson (1973).

Epilogue

If history teaches us merely a simple chronological description of past events, it is of little interest to us. Thus, the history of science is of little interest if it tells us only the standard stories, such as who discovered what, when and so forth. It was T.S. Kuhn who told us what to learn from the history of science, by introducing the concept of *paradigm* in his book *The Structure of Scientific Revolutions.*

Kuhn pointed out that in the history of science, there are periods during which there is a high degree of agreement, both on theoretical assumptions and on the problems to be solved within the framework provided by those assumptions. The resulting coherent tradition of scientific research is called a paradigm. Scientists whose research is based on shared paradigms are committed to the same rules (including established viewpoints) for scientific practice. That commitment and the apparent consensus it produces are prerequisites. The members of the community are tasked to solve puzzles defined by the paradigm. Like exercises and examples in our scientific textbooks, which are a product of paradigms, the solution is assured. In fact, the term paradigm originally is said to be related to an example in a textbook. Our discipline of solar-terrestrial physics is no exception in establishing many paradigms; we have witnessed a number of examples of such situations in this book.

Paradigm	Pioneer	New Paradigm/(initiated by)
Solitary Particles	Birkeland	Diamagnetic Plasma Flow (Chapman)
Closed Magnetosphere	Chapman	Open Magnetosphere (Dungey)
2-D Current System	Chapman	3-D Current System (Alfvén)
Auroral Zone	Loomis	Auroral Oval (Feldstein)
Fixed Auroral Pattern	Fuller/Heppner	Auroral Substorm (Akasofu)
MHD ($E_{\Pi} = 0$)	Alfvén	"Thawing" (Alfvén) ($E_{\Pi} \neq 0$)
Magnetospheric Convection	Axford/Hines	Surviving
Magnetic Reconnection	Hoyle/Parker	Surviving
Magnetic Tube (Sunspot)	Babcock	Not Available
Solar Wind (Thermal)	Parker	Not available

Most all scientists inevitably spend most of their time in this puzzle-solving work. Scientists articulate, verify, elaborate, and consolidate those theories that the paradigm supplies, resolve some of the minor details (residual ambiguities), and attempt to reconcile anomalies. Scientists are generally *practitioners* engaged in such *mop-up* work. Even the brilliant mathematicians Euler, Lagrange, Laplace, and Gauss spent their lives elaborating Newton's paradigm, so that there is nothing to be ashamed of or embarrassed about. We do it all the time. However, there is a danger that bright young scientists tend to be attracted to the puzzle-solving problems and often become almost fanatic supporters of a particular paradigm. Furthermore, because a representation of reality is easier to grasp than reality itself, researchers tend to confuse the two and take concepts for reality. They forget that as their model becomes mathematically rigorous, it becomes increasingly detached from the real world (truth). Is a sunspot really a static, cylindrical magnetic tube of force emerging from beneath the photosphere?

There is often one popular model (paradigm) in each field that is the creation of one scientist. When his model becomes very popular, there is a tendency for a large number of scientists to swarm around it and attempt to *improve* it. The accepted truth about reality is nothing but a consensus of contemporary experts. Meanwhile, all other models are often forgotten.

Eventually, however, there may be a growing number of unsolved puzzles and anomalies for a particular paradigm. As a result, the scientific community's confidence in the paradigm is eroded. This crisis of confidence means that the agreement, which constitutes the sharing of the paradigm, begins to dissolve.

Kuhn observes, however, that even when confronted by severe and prolonged anomalies, scientists do not, in general, respond to the resulting crisis. They think that anomalies are just a few rotten apples and their accidental presence in a barrel does not discredit the other apples. Scientists will push the rules of their paradigm harder to find where and how far they can be made to work. Although they may begin to lose faith, they do not renounce the paradigm that led them into the crisis. They will devise numerous articulations and *ad hoc* modifications of their theory in order to eliminate any apparent conflict. Because of different assumptions used by different researchers in explaining the unsolved anomalies, the initial agreements are lost, causing controversies within a particular paradigm.

The transition from a paradigm in crisis to a new one, a *scientific revolution,* is inaugurated by this growing sense that the existing paradigm has ceased to function adequately. This occurs when scientists carefully examine the barrel and find more rotten apples. Kuhn further observes that scientists do not reject paradigms simply because they are confronted with anomalies or counter instances. Once it has achieved the status of paradigm, a scientific theory is declared invalid only if an alternative candidate is available to take its place. New facts alone do not destroy an outlived theory. As a result, the longer a powerful paradigm survives, the more damage it inflicts. When a powerful paradigm dies, there may be a long period of vacuum because no one can think of anything else. Since a paradigm is fated to die (otherwise, there is no progress) eventually, a long-lasting paradigm actually retards the progress in its field. Max Planck was quoted as saying:

> *A scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die and a new generation grows up that is familiar with it.*

The emergence of a new paradigm candidate is most often far from a cumulative process, one achieved by an articulation or extension of the old paradigm. It is more often the creation of an imaginative mind spurred by an epiphany by a single scientist rather than a result of logical thinking; he suggests that what was thought to be rotten apples are oranges. The act of creation is intuitive, irrational, illogical, and, above all, unscientific (as A. Koestler puts it in his book *The Act of Creation*). Suffice it to say, there will be no inspiration and no breakthrough if one logically elaborates his thinking on the basis of an old, unworkable paradigm.

However, one must be careful about the sudden flash of inspiration. The sudden flash would not arise unless a researcher has had many months or years of struggle and agony in solving his/her problem. A hint of the flash should be available to everyone on an equal opportunity basis, but only one particular person can grasp the hint, because of his/her struggle in the past. A falling apple gave Newton a hint for universal gravity, but even a child is aware that an apple can fall from an apple tree. It is unfortunate that many stories of discovery are distorted, emphasizing only the sudden-flash aspect.

The unscientific nature of creating a new paradigm will most often bring a battle over its acceptance. I suggest that readers learn how the concept of plate tectonics became accepted after A. Wegener proposed the concept of continental drift; those who held the paradigm of geosyncline did not accept him until after World War II, when strong support began to emerge. Sir W.L. Bragg learned about Wegener's idea from Sydney Chapman. After learning more details from Wegener, he gave Wegener's paper to the Philosophical Society. Bragg (1967) mentioned:

> *The local geologists were furious; words cannot describe their utter scorn of anything so ridiculous as this theory, which has now proved so abundantly to be right.*

Since an old paradigm must be fully developed near the end stage, mathematical rigor can be used as a powerful arm against such a creative act (which is often full of errors by its nature) to give an impression that imprecision is a defect of the new idea. Actually, such imprecision is almost a prerequisite for a pioneering paper.

As new paradigm candidates begin to emerge, scientists tend to respond in a way similar to members of any other community. Here, Koestler observed:

> *Like other establishments, they are consciously or unconsciously bent on preserving the status quo, partly because unorthodox innovations are a threat to their authority of the paradigm, but also because of the deeper fear that their laboriously erected intellectual edifice might collapse under the impact.*

A scientific establishment is highly conservative and will attempt to preserve the power of its ruling group against any rebels. Thus, a pioneer often must standalone and be independent-minded on the fringe of the scientific establishment, and perhaps be a rebel (W.B.I. Beveridge in his *The Art of Scientific Investigation*).

During this turbulent period, called the preparadigm period, scientists get involved in passionate controversies. Eventually, one paradigm candidate gains the status of a new paradigm because it is more successful than its competitors. The chosen paradigm will be said to be beautiful, artistically creative, imaginative, inspirational, novel, and elegant after having been treated as a crackpot idea by those clinging to the old paradigm.

The emergence of a new paradigm does not necessarily mean, however, new progress in that particular field. As Koestler put it:

> *Progress by definition never goes wrong. Evolution constantly does, and so does the evolution of ideas, including those of exact science.*

In his address as retiring president of the American Association for the Advancement of Science, S.P. Langley (1889) noted that the progress of science is not like the march of an army towards truth, but…

> *...not wholly unlike a pack of hounds, which in the long-run perhaps catches its game, but where, nevertheless, when at fault, each individual goes his own way, by scent not by sight, some running back and some forward: where the louder-voiced bring many to follow them nearly as often in the wrong path as in a right one: where the entire pack even have been known to move off bodily on a false scent...*

All these confusions are left out of the textbooks by their authors, who are mostly compilers. Thus, students learn only that science progresses monotonically, asymptotically approaching truth. One of the mentor's tasks is to tell his/her students how science actually progresses. The mentors need to teach them to swim not only in a pool, but also in a swift river or ocean.

Suppose that a group of scientists agree that they are going to solve a Garfield jigsaw puzzle – we may call this the "Garfield paradigm." A jigsaw puzzle has many rules. First of all, you have to think of Garfield only, not other jigsaw puzzles, and you cannot use scissors. An important point here is that, like examples in a textbook, the jigsaw puzzle is supposed to be solvable if you do it right. All scientists working on the Garfield puzzle believe so, at least at the beginning.

Unfortunately, the frontier of science is not like a simple jigsaw puzzle for many reasons such as, for example, the limit of accuracy of observations; the pieces do not necessarily match together well; many pieces are still missing. Most scientists have not learned how to deal with such a jigsaw puzzle.

Suppose one piece does not seem to fit at all in the Garfield puzzle. In this situation, many scientists throw the piece away perhaps saying that it came from another puzzle. Some scientists in the Garfield paradigm think that they are not capable of solving the puzzle or they are not working hard enough. They reject, often violently, someone who suggests that the puzzle is not a Garfield puzzle, but another one.

One scientist will finally find that all the pieces, including the odd piece, match better together by supposing that the puzzle is actually the "Snoopy puzzle," not the Garfield puzzle. After some confusion, everyone begins to believe that he is solving the Snoopy puzzle.

It is important to note that scientific research consists of three steps. In the first step, both observations and analyses of a particular phenomenon should be conducted. In the second step, researchers are supposed to synthesize new and earlier observations and then formulate a new interpretation and scheme. The new scheme of a particular phenomenon must be proven quantitatively in the third step.

First Step	Second Step	Third Step
Observation Analysis Interpretation of a particular phenomenon	Synthesis by choosing a set of observations (new and earlier) and formulating a new theory or scheme	Quantitative examination of the new theory conceived in the second step
Observed Facts X, A, K, ?, α, Ω, β, P, τ, M,D, W, Y, C, π, O, ?, B	Prevailing Paradigm K→Ω→C→Y→M→Phenomenon New Paradigms A→β→P→α→O→Phenomenon M→Y→C→Ω→K→Phenomenon α→X→?→π→X→Phenomenon	Quantitative confirmation of the scheme proposed in the second step

The chart above illustrates the three steps. In the second step, individual researchers choose a set of observed facts from a large number of observations and propose their own sequence in explaining the cause-effect relationship for a particular observed phenomenon. The chosen set may differ considerably depending on researchers, and may also differ from the set that corresponds to the prevailing paradigm. In some cases, a successful new paradigm can often include one or more old paradigms with a new and higher order of interpretation. In another case, the chosen set may be identical to the set of the prevailing paradigm, but the sequencing may be reversed; see case (2) in the Second Step. A situation similar to this happened during the development of theories of stellar evolution. A red giant was a newly born star (younger than the Sun) in the old paradigm, but is an old star (older than the Sun) in the latest paradigm.

It is important to know the definition of creation in science in this context, because the second step is an act of creation in science. Creation in science consists of perceiving a new thought pattern on the basis of *already available data and theories.* That is to say, creation in science is synthesis by combining, relating, and integrating data that are often seemingly unrelated to each other. Thus, the second step is indeed this act of creation.

Perhaps Charles Darwin, a 19th century British scientist, is the most creative scientist in history. He had an amazing synthesizing power. He found that there are three kinds of coral reefs – the first one is, the reef surrounding the shore of an island. The second one, is an atoll surrounding an island in the center, and the third one is simply a circular atoll. This observation was good enough. However, he went further. He inferred that these different types are related. By hypothesizing the increase of sea level, he suggested that the different types of reefs represent different states of the atoll formation. This example shows how great his synthesizing power was. In fact, his hypothesis has recently been confirmed by drilling at the center of an atoll. In another example, he observed:

> *What can be more curious than the fact that the hand of a man, formed for grasping, that of a mole for digging, the leg of a horse, the paddle of the porpoise, and the wing of the bat should all be constructed on the same pattern?*

Charles Darwin was one of the first scientists to conceive the concept of the evolution of life and developed the hypothesis. We are surrounded by many living creatures: single-cell organisms, starfish, squids, horses, dogs, cats, crabs, scorpions, mushrooms, flower-bearing plants, ferns, algae, etc. They do not seem to be related at all. With his great synthesizing power, however, Darwin inferred that all these creatures have evolved from a single-cell organism and have branched out into different species.

Now, returning back to the three steps, there are very few researchers who can engage in all three steps satisfactorily. Most of them can engage in one of the three steps. There is nothing wrong with that; most researchers are supposed to be good at one of the steps. The problem is that they tend to spend all their effort on only one of the steps and do not even realize that they are doing so. An excellent researcher who is good at the first step may be intentionally or unintentionally considering the second and third steps and thus can design a crucial observation to advance his/her discipline. An excellent researcher in the first or third step may suggest a new scheme in the second step or a new observation to discover an undiscovered element. Indeed, there are many set elements that have not yet been discovered; see the ? sign in the table. An excellent observation specialist may be able to identify one of them by carefully designing the observation on the basis of an intentional or unintentional consideration in the second step; see case (3) in the Second Step above. Such a case may be called a *discovery.*

As described in detail earlier, the second step is often misunderstood to be *unscientific.* In order to comprehend this statement, let us take the case of the extinction of dinosaurs. There are a large number of observed facts on this particular phenomenon. It is impossible to propose a theory that can explain all the observed facts. One has to choose a set of observed facts that are considered to be essential. Some researchers are confident that the cause of the extinction is internal (e.g., climate change), while some others believe it is external (e.g., an impact by asteroids) and set up a scheme to explain the cause of the extinction in either case. Actually, there are many observations to support both. Thus, in the history of this particular discipline, the paradigm shifts between the two (namely, internal or external causes) from time to time. The set depends greatly on individual researchers, so that *this choice is a very subjective process* and becomes also controversial.

It is for this reason that many researchers are uncomfortable with the second step and tend to avoid it by saying that it is not a scientific process. Thus, considerable courage is needed to pursue the second step. This is also the step to confront an established paradigm, if the new interpretation is radically different. After all, it is important to realize that in science the first and the third steps are designated to serve the second step. The second step is the one that leads to a breakthrough.

Unfortunately, some of those who are mainly interested in the first step, or are not comfortable with pursuing the second step, keep making only their observations by saying that observations are most crucial. Yet, some others keep a new (important) observation under the rug, because it does not conform to the prevailing paradigm. Some of those who are mainly interested in the third step keep improving a particular theory that is a synthesis by someone else, by criticizing that the second step is not science. It is in such a way that some waste considerable amounts of supercomputer time, but believe they are doing good science.

There is no question that the third step distinguishes science from science fiction, so that it is a vital scientific process. Unfortunately, however, many of those who are mainly interested in the third step tend to insist that the third step is the only important aspect of science. If the second step is not appropriate, the third step will not be very useful (although the effort may not be totally wasted). There is another serious problem in the third step. For example, with proper initial and boundary conditions, one can solve, say, a set of MHD equations and show that one can reproduce an observed phenomenon. However, such a process does not provide any physical insight into the observed phenomenon. One has simply demonstrated that a particular phenomenon can be described by a set of MHD equations, so that the physical processes associated with the phenomenon have not been further elucidated.

Unfortunately, human nature dictates that the revolution will soon turn into a new orthodoxy, with its unavoidable symptoms of one-sidedness, specialization, loss of contact with other provinces of knowledge and ultimate estrangement from reality (Koestler). Since the new paradigms can never reach truth, a new crisis will eventually arise, leading to a new revolution, namely a new synthesis, and the cycle starts all over again. History repeats itself. Scientific communities are no different from other establishments (except that scientific knowledge is cumulative, unlike the political one). After all, human beings called scientists created science, and so the history of science is nothing but a human drama.

After all this, if I have to conclude this book by choosing one word, I will choose the word *open-mindedness.* Each of us scientists wishes to make some small progress in our discipline for a better understanding of Nature. This is all we can hope for. Open-mindedness is the only way to make one small step eventually in the right direction. This is what I have learned in my research life, although it is so difficult to be open-minded.

Further Readings: Mostly Historical

Akasofu, S.-I., B. Fogle and B. Haurwitz (Ed.), *Sydney Chapman, Eighty,* University Corporation for Atmospheric Research, Boulder, Colorado, 1968.

Akasofu, S.-I., Evolution of Ideas in Solar-terrestrial Physics, Geophys. J.R. Astro. Soc., **74**, 257, 1983.

Akasofu, S.-I., *Physics of Magnetospheric Substorms,* D. Reidel Pub. Co., Dordrecht-Holland, 1977.

Akasofu, S.-I., *Polar and Magnetospheric Substorms,* D. Reidel Pub. Co., Dordrecht-Holland, 1968.

Akasofu, S.-I., S. Chapman, and A.B. Meinel, The Aurora, *Handbuch der Physik,* Vol. XLIX/1, Springer-Verlag, Berlin, 1966.

Akasofu, S.-I. and S. Chapman, *Solar-Terrestrial Physics,* Oxford Univ. Press., 1972.

Alfvén H., *Cosmical Electrodynamics,* Oxford Univ. Press, 1950.

Alfvén H., *Cosmic Plasma,* D. Reidel Pub. Co., Dordrecht-Holland, 1981.

American Geophysical Union, Pioneers of Space Physics 2, Reprinted from Journal of Geophysical Research, **101**, Number 5, 1996.

Armstrong, E.B. and A. Dalgarno, *The Airglow and the Aurorae,* Pergmon Press, London, 1956.

Banks, P.M. and G. Kockarts, *Aeronomy,* Academic Press, New York, 1973.

Beveridge, W.I.B., *Seed of Discovery,* Norton, New York, 1980.

Beveridge, W.I.B., *The Art of Scientific Investigation,* Vintage Press, Random House, New York, 1957.

Biermann, L., Kometenschureife und Solare Korpuskularstrahlung, Zeitschriffefur Astrophysik, **29**, 274, 1951.

Birkeland, K., *On the Cause of Magnetic Storms and the Origin of Terrestrial Magnetism,* Norwegian Aurora Polaris Expedition, 1902-1903.

Bryant, D., *Electron Acceleration in the Aurora and Beyond,* Institute of Physics Publishing, Bristol, 1999.

Burgers, D.J.M. and L.H.C. van de Halst (Ed.), *Gas Dynamics of Cosmic Clouds,* North Hallard Pub. Co., Amsterdam, 1955.

Chamberlain, J.W., *Physics of the Aurora and Airglow,* Academic Press, New York, 1961.

Chapman, S. and J. Bartels, *Geomagnetism,* I, II, Oxford Univ. Press, 1940.

Cook, A.H. and T.F. Gaskell (Ed.), *The Earth in Space,* Royal Astrom. Soc., London, 1968.

Cowling, T.G., *Magneto-hydrodynamics,* Interscience Pub., Inc., London, 1957.

Dungey, J.W., *Cosmic Electrodynamics,* Cambridge Univ. Press, 1958.

Dyer, E.R. (Ed.), *Solar-Terrestrial Physics/1970,* D. Reidel Pub. Co., Dordrecht-Holland, 1972.

Eather, R.H., *Majestic Lights,* AGU, Washington, 1980.

Egeland, A. and J. Holtet (Ed.), *The Birkeland Symposium on Aurora and Magnetic Storms,* September 18-22, 1967, Centre National de la Recherche Scientifique.

Fuller, V.R. and E.R. Bramhall, Auroral Research at the University of Alaska, 1930-1934, Univ Alaska Pub., 1937.

Gillmor, C.S. and J.R. Spriter (ed.), *Discovery of the Magnetosphere,* America Geophysical Union, Washington, D.C., 1997.

Gubbins, D., Geomagnetic polarity reversals; A connection with secular variation and core-mantle interaction?, Rev. Geophys., **32**, 61, 1994.

Hoyle, F., *Some Recent Researches in Solar Physics,* Cambridge Univ. Press, 1949.

Humboldt, Alexander von, *Cosmos: Physical Description of the Universe,* Vol. 1, London, 1871.

Keiper, G.P. (Ed.), *The Sun,* Univ Chicago Press, 1953.

Kellner, L., *Alexander von Humboldt,* Oxford Univ. Press, Oxford, 1963.

Kelvin, President Address, Proc. Roy. Soc., **52**, 299, 1892/93.

Kennel, C.F., *Convection and Substorms,* Oxford Univ. Press, 1995.

Koestler, A., *The Act of Creation,* Picadov, 1964.

Kuhn, T.S., *The Structure of Scientific Revolutions,* (2nd ed.), Univ. Chicago Press, 1962.

Langley, S.P., Amer. J. Sci., 3, **37**, 1, 1809, see Barr, E.S., Amer. J. Phys. **28**, 42, 1960.

LeGalley, D.P. and A. Rosen (Ed.), *Space Physics,* John Wiley and Sons, Inc., New York, 1964.

Lindermann, F.A., Note on the Theory of Magnetic Storms, Phil. Mag., **38**, 1919.

Mackin, K.J., Jr. and M. Neugebauer (Ed.), *The Solar Wind,* Permagon Press, 1966.

Massey, H.S.W. and R.L.F. Boyd, *The Upper Atmosphere,* Hutchinson, London, 1958.

Maunder, E.W., Mon. Not. R. Astr. Soc., **65**, 2, 1905.

McCormac, B.M., *Aurora and Airglow,* Reinhold Pub. Co., New York, 1967.

Merrill, R.T. and P.L. McFadden, Geomagnetic Polarity Transitions, Rev. Geophys., **37**, 201, 1999.

Newton, H.W., *The Face of the Sun,* Penguin Books, 1958.

Odishaw, H. (Ed.), *Research in Geophysics,* Vol. 1 and 2, The M.I.T. Press, Cambridge, Mass, 1964.

Parker, E.N., *Interplanetary Dynamical Process,* Interscience Pub., John Wiley and Sons, New York, 1963.

Parker, E.N., C.F. Kennel, and L.J. Lanzerotti, *Solar System Plasma Physics,* Nutt-Holland Pub. Co., Amsterdam, 1979.

Piddington, J.H., *Cosmic Electrodynamics,* Wiley-Interscience Pub., New York, 1969.

Rees, M.H., *Physics and Chemistry of the Upper Atmosphere,* Cambridge Univ. Press, Cambridge, 1989.

Spitzer, L., Jr., *Physics of Fully Ionized Gases,* Interscience Pub. Inc., London, 1956.

Stern, D.P., A brief history of magnetospheric physics during the space age, Rev. Geophysics, **34**, 1, 1996.

Stewart, C. and J.R. Spreiter (Ed.), *Discovery of the Magnetosphere,* American Geophys. Union, Washington D.C., 1997.

Störmer, C., *The Polar Aurora,* Oxford Univ. Press, 1955.

Svestka, Z., *Solar Flares,* D. Reidel Pub. Co., Dordrecht-Holland, 1976.

Thomson, J.J. and G.P. Thompson, *Conduction of Electricity Through Gases,* General Publications, Tronto, 1903.

Zirin, H., *Astrophysics of the Sun,* Cambridge Univ. Press, 1988.

Name Index

Ahn, B.-H.---99, 100
Akasofu-, S.-I. V, 14, 17, 18, 19, 20, 21, 22, 23, 25, 26, 29, 30, 31, 33, 39, 40, 43, 47, 48, 49, 50, 51, 52, 56, 57, 58, 59, 62, 63, 65, 67, 68, 69, 71, 73, 74, 78, 79, 80, 95, 99, 100, 101, 103, 104, 105, 110, 111, 112, 116, 117, 119, 120, 123, 124, 125, 128, 131, 134, 135, 136, 137, 138, 141, 142, 143, 146, 147, 148, 152, 153, 161, 166, 167, 173, 181, 182, 184, 186, 190, 191, 193, 194, 195, 196, 198, 199, 200, 201, 202, 203, 204, 213
Alfvén, H. ---- XXVIII, XXIX, XXXII, 39, 40, 89, 92, 94, 95, 96, 155, 159, 193, 196, 198, 203, 212
Anderson, B.J.---108
Anger, C.D.---53, 62, 67
Armstrong, T.P.---128
Arnoldy, R.L.---31, 64
Asbridge, J.R.---23
Aspnes, J.D.---180
Axford, W.I.---92, 93, 204
Babcock, H.W.---204
Baker, D.N.---64
Bame, S.J.---XXIX, 24
Banks, P.M.---64
Bartels, J.---XXII, XXVII, 6, 9, 34, 121, 133, 140
Bates, D.R.---59, 143
Baumjohann, W.---64, 100, 115
Behannon, K.W.---173
Berkey, F.T.---64
Beveridge, W.I.B.---207 214
Bhardwaj, A.---77
Biermann, L.---XXVII, 11
Birkeland, K.---XXII, XXVII, 2, 3, 7, 34, 89, 90, 91, 94, 95, 155, 204 214
Birn, J.---64
Boström, R.---100, 101, 102, 103
Bragg, W.L.---204
Bramhall, E.H.---46, 55, 214
Brekke, A.---64
Bryant, D.A.---73
Buchau, J.---50, 62
Bullard, E.C.---122
Burch, J.---64, 121
Burlaga, L.F.---173
Cahill, L.J., Jr.---10, 17
Cain, J.---10, 17
Carlheim-Gyllensköld, N.---44
Carlson, C.W.---73
Carrington, R.---XIX, XX, XXXII, 133, 134, 144, 145, 147, 149

Chapman, S. --XVI, XXII, XXVI, XXVII, 3, 5, 6, 8, 9, 10, 11, 13, 14, 15, 17, 18, 19, 20, 21, 24, 30, 34, 39, 40, 45, 54, 58, 60, 61, 68, 69, 70, 71, 89, 90, 91, 92, 93, 95, 96, 109, 110, 112, 120, 121, 122, 123, 126, 133, 140, 193, 194, 195, 204, 206, 213, 214
Choe, C.S. ---------- 159
Clarke, J. ---------- 75
Coleman, P.J., Jr. ---------- 17
Colgate, S.A. ---------- 28
Coroniti, F.V. ---------- 64
Covey, D.N. ---------- 197, 199
Cowley, S.W.H. ---------- 64
Craven, J.D. ---------- 51, 64
Daglis, I. ---------- 119, 120
Darwin, C. ---------- 209, 210
Davis, T.N. ---------- 54, 120
Dawson, J. ---------- 70
Deehr, C.S. ---------- 45, 190, 191
Delory, G. ---------- 72
Dessler, A.J. ---------- 18, 177
Devik, O. ---------- 3
Dryer, M. ---------- 166, 190, 191, 201
Dungey, J.W. ---------- XXIX, XXX, 24, 204, 214
Eather, R.H. ---------- 74, 214
Eddington, A.S. ---------- 15
Egeland, A. ---------- 67, 128, 214
Elphinstone, R.D. ---------- 64
Elvey, C.T. ---------- 16, 45, 46, 54, 58, 60
Euler, L. ---------- 204
Ewell, M.W. ---------- 160
Fadden, J. ---------- 72
Faggen, G.K. ---------- 176
Fairfield, D.H. ---------- 24, 26, 27, 31, 64
Fälthammer, C.-G. ---------- 64, 67
Feldstein, Y.-I. ---------- 45, 47, 48, 50, 59, 64, 110, 112, 204
Ferraro, V.C.A. XXII, XXVII, XXIX, 5, 6, 7, 8, 9, 10, 13, 15, 18, 21, 38, 42, 89, 95, 122
Fillius, W. ---------- 201
Forbush, S. ---------- XXXIV, 202
Frank, L.A. ---------- 17, 51, 64, 65, 76, 77, 113, 117, 118
Friss-Christensen, E. ---------- 100
Fritz, T.A. ---------- 42, 45, 47, 128
Fry, C.D. ---------- 145, 147, 148, 165, 166, 184, 191, 201
Fukunishi, H. ---------- 78
Fuller, V.R. ---------- 46, 54, 55, 204, 214
Fuselier, S.A. ---------- 121
Galeev, A.A. ---------- 131
Galperin, Yu.-I. ---------- 64

Gartlein, C. ---- 54, 68
Gassmann, G. ---- 50
Gauss, C.F. ---- 140, 204
Gladstone, G.R. ---- 77
Gleeson, L. ---- 203
Gold, T. ---- XXV, 12, 13
Green, J.L. ---- 121
Greenwald, R.A. ---- 64, 93, 94
Grigorov, K.G. ---- 190, 191
Gurnett, D.A. ---- 64, 73
Haerendel, G. ---- 64
Hakamada, K. ---- 145, 166, 167
Hallinan, T.J. ---- 70, 71, 72
Hamilton, D.C. ---- XXIX
Hasegawa, A. ---- 114
Heikkila, W.J. ---- 50, 64, 115
Helfferich, M. ---- 74
Heppner, J.P. ---- 11, 46, 54, 204
Herman, H. ---- 64
Hewish, A. ---- 173, 176
Hines, C.O. ---- 92, 93, 204
Hoeksema, T. ---- 135
Honda, K. ---- 41
Hones, E.W., Jr. ---- 23, 24, 64
Hoyle, F. ---- 24, 204, 214
Hultqvist, B. ---- 45, 67
Humboldt, A. ---- XXVI, 1
Iijima, T. ---- 52, 53, 101, 102
Kamide, Y. ---- 52, 64, 65, 97, 99, 100, 112, 132
Kamiyama, H. ---- 41
Kan, J.R. ---- 64, 67, 126, 155
Kartalev, M.D. ---- 190, 191
Kato, Y. ---- 41
Kawasaki, K. ---- 64, 95
Kellner, L. ---- 1
Kelvin, W.T. ---- XX, XXII, 1, 5
Kennel, C.F. ---- 64, 78, 215, 216
Khorosheva, O.V. ---- 48
Kimball, D.S. ---- 62, 73
Kirkpatrick, C.B. ---- 95
Klein, L.W. ---- 173
Koestler, A. ---- 206, 207, 215
Kozuka, Y. ---- 138, 146
Kroehl, H.W. ---- 100
Kuhn, T.S. ---- 204, 205, 215
Lagrange, J.L. ---- 204
Langley, S.P. ---- 207

Langmuir, I. ---- 5
Laplace, P.S. ---- 204
Lassen, K. ---- 47, 67
Lebedinsky, A. ---- 54
Lee, L.C. ---- 64, 71, 141, 142, 159, 173
Liepmann, H.W. ---- 12
Lin, W.C. ---- 39
Lindemann, F.A. ---- 5, 7
Liou, K, ---- 108
Loomis, E. ---- 42 43, 45, 47, 204
Ludwig, G. ---- XXIX
Lui, A.T.Y. ---- 23, 53, 62, 64, 114, 132
Lundin, R. ---- 64
Lyatsky, W.B. ---- 64
Lyons, L.R. ---- 64
Makita, K. ---- 117
Matsushita, S. ---- 97
Maunder, E.W. ---- XXI, XXII, XXVII, 1, 5, 6, 185
Maurer, H. ---- 100
Maurits, S. ---- 179
Maxwell, J.C. ---- XX, 7
McCracken, K.G. ---- 4
McIlwain, C.E. ---- XXIX, 34, 64, 79, 107, 128
McIntosh, P. ---- 152, 156
McPherron, R.L. ---- 64, 78
Meng, C.-I. ---- 31, 62, 64, 78, 79, 94, 95, 110, 111, 112, 117
Merritt, R. ---- 180, 181
Mishin, V.M. ---- 64
Ness, N.T. ---- XXVIII, 11, 139, 140
Neugebauer, M. ---- 11, 215
Newton, H.W. ---- 6
Newton, I. ---- 131, 205, 206, 215
Nikolsky, A.V.U. ---- 109, 110
Nishida, A. ---- 64, 78
Numazawa, S. ---- 146, 147
Ogita, N. ---- 202, 203
Oki, T. ---- 136, 137, 138
Olmsted, C. ---- 136, 137
Olson, J. ---- 78, 80
Opgenoorth, J. ---- 64
Osborne, D.L. ---- 72
Parker, E.N. ---- XXVII, 11, 149, 156, 167, 177, 198, 200, 204, 216
Parks, C.G. ---- 64, 65
Paterson, W.R. ---- 118
Pellat, R. ---- 64
Pellinen, R. ---- 64
Perreault, P. ---- 30, 64

Piddington, J.H. ---- XXVIII, 216
Pike, C.P. ---- 50
Pizzella, G. ---- 128
Planck, M.K. ----206
Potemra, T.A. ----52, 53
Pudovkin, M.I. ----32, 64
Randall, B. ---- 128
Rao, U.R. ----4
Reiff, P.H. ---- 33, 64, 179
Richmond, A.D. ---- 97, 100
Roberts, W.O. ---- 70, 129
Roederer, M. ---- 14
Romick, G.J. ---- 60
Rosenfeld, A.H. ----5
Rostoker, G, ---- 35, 64, 78, 99, 100
Russell, C.T. ---- 64
Rust, D. ---- 188
Sackinger, W. ---- 181
Sagdeev, V.A. ---- 131
Saito, T. ---- 135, 136, 137, 138, 141, 142, 143, 145, 146, 147, 173
Sato, T. ---- 114
Scherrer, P.H. ---- 135
Schultz, M. ---- 80
Schuster, A. ----XXII, 1, 5
Sergeev, V.A. ---- 64
Shea, M.A. ----4
Sigwarth, J. ---- 113, 118
Silsbee, H.C. ---- 92
Singer, S.F. ---- 16
Siscoe, G.L. ---- 64
Smith, E.J. ---- 184, 190, 191, 198, 201
Snyder, A.L. ---- 11, 49, 50, 62, 64, 166
Spiro, R.W. ---- 100
Stenbaek-Nielsen, H.C. ---- 72
Stoffregen, W. ---- 54
Storey, L.R.O. ---- XXVIII
Störmer, C. ----XXII, 3, 4, 7, 15, 16, 37, 38, 39, 44, 55, 93, 96, 216
Sugiura, M. ---- 65, 120
Sun, W. ---- 36, 37, 101, 103, 104, 105, 190, 191, 201
Svalgaard, L. ----156
Swift, D.W. ---- 64
Sydora, R.D. ---- 71
Tajima, T. ---- 71
Takahashi, T. ---- 146, 147
Tang, F. ---- 183, 184
Tappin, S.J. ---- 176
Taylor, H. ---- 128

Temerin, M. ---- 72
Thomson, J.J. ---- 2
Townshend, J. ---- 97, 132
Troshichev, O.A. ---- 64
Tsurutani, B. ---- 183, 184
Tsyganenko, N.A. ---- 64
Vampola, A.L. ---- 37, 38
Van Allen, J.A. ---- XXVIII, XXIX, 16, 39, 51, 113, 128, 202
Vasyliunas, V.M. ---- 27, 64, 78
Venkatesan, D. ---- 128
Vernov, S.N. ---- 48
Vestine, E.H. ---- 42, 45, 92
Voots, G.R. ---- 73
Wagner, J.S. ---- 70, 71
Wagner, R. ---- 50
Walker, J. ---- 100
Wang, H. ---- 160, 161
Waters, C.L. ---- 108
Watkins, B. ---- 179
Wegener, A.L. ---- 206
Wescott, E.M. ---- 45, 181
Whalen, B.A. ---- 50, 51, 62
Wilcox, J.M. ---- 137, 145, 156
Winckler, J.R. ---- 64, 79
Winningham, J.D. ---- 50, 64
Xu, W.-Y. ---- 37
Yamamoto, G. ---- 41
Yasahara, F. ---- 64
Yoshida, S. ---- 202, 203
Zaitzev, A.W. ---- 100
Zirin, H. ---- 152, 158, 160, 161, 216
Zmuda, A.J. ---- 51

INDEX

A

ACE satellite.. 188
AE...120, 165
Alaska meridian chain................47, 97, 99
Alfvén waves XXVIII
Alfvén's theory...96
all-sky camera..................................45, 97
atmospherics XXVIII
auroral electrojet.............................94, 177
auroral kilometric radiation 73
auroral oval ..XXX, 37, 47, 48, 50, 51, 52, 109, 116, 178
auroral potential structure 71, 155
auroral rays 70
auroral storm......................................68, 69
auroral substormXXXI, 21, 54, 58, 62, 69, 124, 160
auroral visibility.......................................42
auroral zone 42, 45, 48, 108, 109

B

Birkeland current......................................95
Boström's current loops......................... 100

C

Chapman-Ferraro current.............9, 10, 89
Chapman-Ferraro theory 6, 13, 18, 21, 24, 38, 96
closed magnetosphere XXIX
CME 6, 166, 171, 183
Cold War.. 166
comet tails 11, 77
comets 173
continental drift 206
convective (**E**x**B**) motions 92
corona 163
coronal hole .... XXI, 6, 12
coronal holes... 185
coronal mass ejection (CME) .... XXXII, 6, 183
corpuscular school....................................2
cosmic ray ....XXXIV, 201
Cosmos ....XXVI, 1
creation in science.................................209
cyclone 152

D

Debye length..8
diffuse aurora..113
dipolarization 105, 113
directly driven component.............35, 178
discovery 210
DMSP satellites................................62, 73
driver gas 183
DS component.. 91
Dst index...............................91, 165, 120
Dynamic Explorer satellite..............59, 64

E

ε Parameter 24, 177
(**ExB**) plasma flow 35, 114
Earth flare 158
Earth’s current 180
Earth's main dipole............................... 133
eastward electrojet..........................91, 109
emerging magnetic flux tubes 150
Emperor Showa 76
energetic neutral atoms (ENA)............ 120
equivalent current system 90, 93, 94
eruptive prominences 159
Evershed flow.. 152
evolution of life 210
Explorer 12 satellite 10

F

field-aligned currents 52, 94, 100, 104, 158, 159, 161, 163
First Polar Year 45
flare buildup 149
Forbush decrease...................XXXIV, 202
force-free field......................158, 160, 163
frozen-in field condition 155

G

geomagnetic indices............................. 120
geomagnetic pole 42
geomagnetic storm ..XXVI, XXXI, 13, 14, 18, 34, 68, 69, 119, 125

geomagnetic storm prediction 192
geomagnetic storms............ 10, 16, 34, 182
geosyncline 206
gradually commencing storms................19
green line...75
guiding center XXVIII

H

HAF..................................... 166, 177, 186
heliosphere ....XXXIV, 197
heliospheric current sheet ...... XXXI, 145, 183, 197
High Energy Neutral Atom (HENA) ... 120
high-speed streams 185
Hubble Space Telescope.........................77
hurricane 152

I

IGY 45, 47
IMP-1 satellite XXVIII
MP-6 73
IMS Meridian Chain 99
initial phase...............................XXVII, 13
International Conference on Substorms (ICS) 67
International Geophysical Year (IGY) ....XXV, 45
International Magnetosphere Study (IMS) 99
International Polar Years 42
interplanetary corotating structure........ 169
interplanetary magnetic field (IMF) 21, 24, 134
interplanetary physics............................ 165
interplanetary scintillation (IPS) 173
interstellar gas...........................XXXI, 197
interstellar magnetic field 198
ionospheric physics (aeronomy)........... 165
ionospheric substorm 27
Iridium satellite..................................... 108
ISIS-2 satellite 53, 62
isochasms 42

J

Jovian aurora...74
Jupiter..XXX, 77

K

KRM computer code.................97, 99, 102

L

libration point satellite, S3...................... 34
loops 182

M

M Region 6
magnetic clouds.................................... 182
magnetic energy storage 149
magnetic equator 166
magnetic flux emergence...................... 149
magnetic flux ropes.............................. 182
magnetic flux tubes 149
magnetic loops...................................... 153
magnetic reconnection .....XXVIII, 27, 28, 35, 36, 114, 160, 162
magnetic shear...................................... 160
magnetized planets 139
magnetohydrodynamics (MHD) XXVIII
magnetometers 41,97
magnetosphere............XXV, XXIX, 13, 21
magnetosphere-ionosphere (MI) coupling 35, 95
magnetospheric physicsXXV, XXVII, 165
magnetospheric storm ... XXXI, XXXII, 69
magnetospheric substorm ... XXX, XXXI, 21, 27, 33, 34, 69, 119, 124, 160, 163
magnetotail XXVII, 27, 35, 106, 115, 118, 163
main phase ....XXVI, 10, 13, 17, 68, 124
mantle... 24
Mariner 2 spacecraft XXV11, 11
Mars .... XXX, 77
Method of Natural Orthogonal Components (MNOC)....................... 36
MHD 155, 159, 171, 177, 210
midday aurora... 74

N

National Science Foundation.................48
near-Earth neutral line 113
negative sudden impulse......................... 19
Neptune................. XXX, XXXI, 133, 139
neutral hydrogen atoms.....................21, 24
neutral line.............28, 135, 136, 137, 166
neutron stars 141

Nikolsky's Spiral 109

O

off-centered dipole 142
offset tilted dipole (OTD) 139
omega band 56
open field lines 52
open magnetosphere 36
open model 25
open region 115, 118
open-mindedness 212
Orion arm 197
outer radiation belt 51
over-the-horizon (OTH) radar 166
oxygen ions (O^+) .... XXIX, XXXI, 35

P

paradigm 149, 150, 204
Parker spiral 167, 198
photosphere 149, 152, 153
photospheric dynamo 154
photospheric magnetic field. 133, 135, 138
Pioneer 10 and 11 201
pitcher model 28, 118
planetary magnetism 133
plasma .... XXII, 5, 8, 96
plasma mantle XXIX
plasma sheet .... XXVIII, 120
plasmasphere XXVIII
polar cap 37, 52, 116
polar cap potential 179
polar magnetic substorm ...... XXXI, 21, 27
polar rain 115
POLAR satellite 70, 87
power transmission lines 180
Poynting flux 32

R

recovery phase XXVII
reversal of the polarity 136, 143
ring current 13, 119
ring current belt .... XXVII, 124
ring current ions 119
rockoons XXVIII

S

Saturn XXX
Scandinavian school 96
Science Citation Index 58
science fiction 211
scientific revolution 205
SD current 109
sector boundary 198
shock wave . XXXII, XXXIV, 9, 167, 171, 173, 183, 186, 188, 200, 201, 202
sky map 173
Solar and Heliospheric Observatory (SOHO) 81, 85
solar aurora 158
solar corona 11, 76, 145
solar flares ..XIX, XXXI, 6, 27, 149, 155, 160, 166
solar magnetic fields 151
solar physics 165
solar wind 11, 145, 166
solar wind-magnetosphere interaction XXX, 89
solar wind-magnetosphere dynamo 32, 160, 163
solar-terrestrial physics .... XX, XXII
source surface .... XXXI, 133, 135, 166
South Pole 74
southward turning 24
space weather 165, 192
SSC 12, 13, 18, 26
stellar evolution 209
STEREO mission 177
storm sudden commencement (SSC) .... XXVII, XXXII, 9
storm time radiation belts 16
storm-substorm relationship 34, 119
storm-time ring current belt 119
substorm onset 112
sudden filament disappearances 166
sudden flash of inspiration 206
sunspot cycle 138, 156, 166
Sunspot Cycle 21 135
sunspots .... XXI, 149, 150, 156, 186
SuperDARN 93

T

27-day recurrence tendency XXI
27-day variations 168
terrella .... XXII, 3
The Ionospheric Research Committee ... 15
three-dimensional current system 94
tippy bucket model 28
TRACE satellite 163

Trans-Alaska Oil Pipeline.................... 181
TRIAD satellite 51

U

Ulysses 197
unipolar region...................................... 135
unloading component......................35, 178
Upsilon Andromedae.............................75
Uranus ....XXX, XXXI, 133, 139

V

Van Allen belt....................XXVII, 16, 113
Venus ....XXX, 77

W

westward auroral electrojet............. 91, 109
westward traveling surge..........56, 62, 111

X

X-line 28, 163

Z

zodiacal light 11

A Scenario on Substorm Onset (dedicated to the International Conference on Substorms (ICS-6), March 25-29, 2002, Seattle, Washington).

(Note added on March 1, 2002)

1. Source of Auroral Electrons

No two substorms are alike. This is understandable by considering that the solar wind-magnetosphere generator power ε(t) is quite variable (Section 1.9). The substorm pattern presented as Figure 2.17 in page 59 contains only the minimum number of what we might call "the common denominators" of all such storms, although there is always a great danger of over-simplification in this selection process. Indeed, there are a large number of other observational facts. Thus, the causes of the magnetospheric substorm can provide a good example for our practice of the scientific process described in Epilogue (page 204), in particular following the three steps given in the table on page 209.

Figure A shows the latest observation of the field-aligned currents (Walters et al., 2001; see also Figure 3.12). Since it is well established that the auroral electrons are the field-aligned current carrier, Region II current in the dark sector (in red) must be carried by auroral electrons (See also page 52 and Figures 2.10 and 2.11). Together with Region I current (in blue), Region II current constitutes Boström's Type II current system (Figure B); for details see Figure 3.9. This was confirmed by projecting the ionospheric Pedersen currents on the equatorial plane and comparing it with the distribution of the radial currents on the equatorial plane (Figure C; for details, see Figure 3.10). *Thus, Region I and II currents are connected by the Pedersen current in the ionosphere and the radial current by the equatorial plane.*

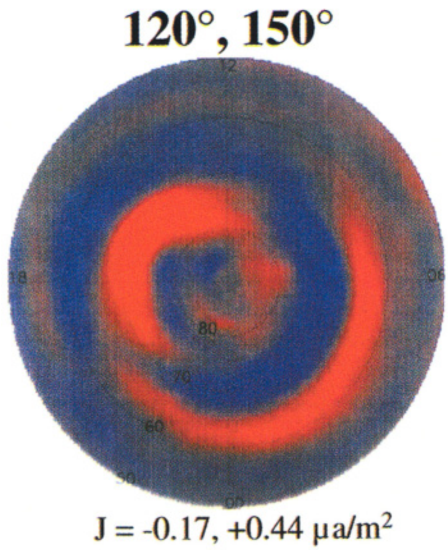


Figure A: Distribution of the field-aligned currents (red areas indicate where upward field-aligned currents are present and blue areas indicate where downward field-aligned currents are present). For details, see Figure 3.12 on page 108.

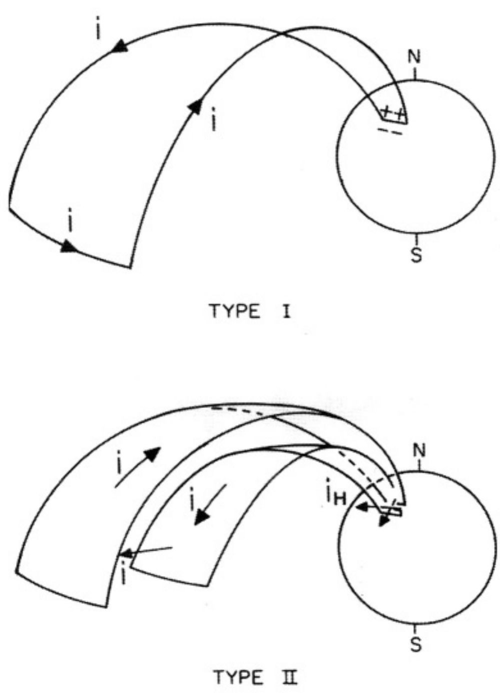


Figure B: Boström's Type I and Type II current systems. For details, see page 102.

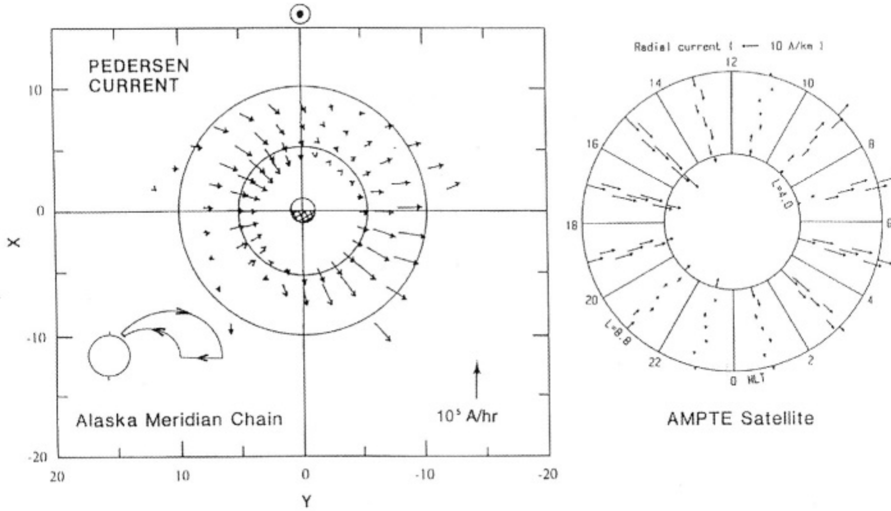


Figure C: Comparison of the equatorial projection of the ionospheric Pedersen current and the radial current on the equator. For details, see Figure 3.10 on page 103.

Note that the distribution of the field-aligned current (Figure A) is not symmetric with respect to the magnetic noon-midnight meridian. This clockwise rotation of the axis of symmetry and the extension of the morning cell across the midnight meridian must be related to the rotation of the axis of symmetry of the evening and morning convection cells (See Figure 3.6d on page 99 and Figure G of this Note), which arises from the fact that the ionosphere has an anisotropic conductivity.

As shown in Figure 2.23 (page 67), the aurora consists of two groups, the dayside group and the nightside group, although both groups form the auroral oval together. Elphinstone et al. (Rev. Geophys. **34**, 169, 1996) showed that the dayside part of the oval is connected to the magnetopause and that the nightside group is connected to a belt located deep in the magnetosphere at a distance of less than 10 earth radii in the magnetotail. Thus, the nightside group requires a different consideration from the dayside group.

Now, since the aurora electrons are the upward field-aligned current carrier, *the cause of the nightside group of the aurora must be directly related to the generation process of Boström's Type II current* *system* and also to the divergence of the equatorial directed $\boldsymbol{I}_{p'}=\Sigma_p\boldsymbol{E}$ (where Σ_p denotes the Pedersen conductivity) and thus to $\nabla \cdot (\Sigma_p\boldsymbol{E})$.

2. Substorm Onset

The westward electrojet grows suddenly at substorm onset. It is conclusively shown by incoherent scatter radar observations (Figure D) that the westward electrojet current is generated by an equatorward-directed electric field $\boldsymbol{E}$ (Brekke et al., 1974); the westward electrojet is the Hall current which is a concentrated eastward ($\boldsymbol{E}$ x $\boldsymbol{B}$) flow of electrons in the E region of the ionosphere. Thus, substorm onset is associated with a sudden growth of the equatorward-directed electric field $\boldsymbol{E}$ that can explain both the growth of the westward electrojet, the Boström's Type II current system and thus Region II current. The latter will enhance the upward field-aligned current and brighten auroral arcs and increases the ionization in the ionosphere. The resulting increase of the ionization increases both the electrojet and the Pedersen current and thus will brighten further auroral arcs, constituting a positive feedback system.

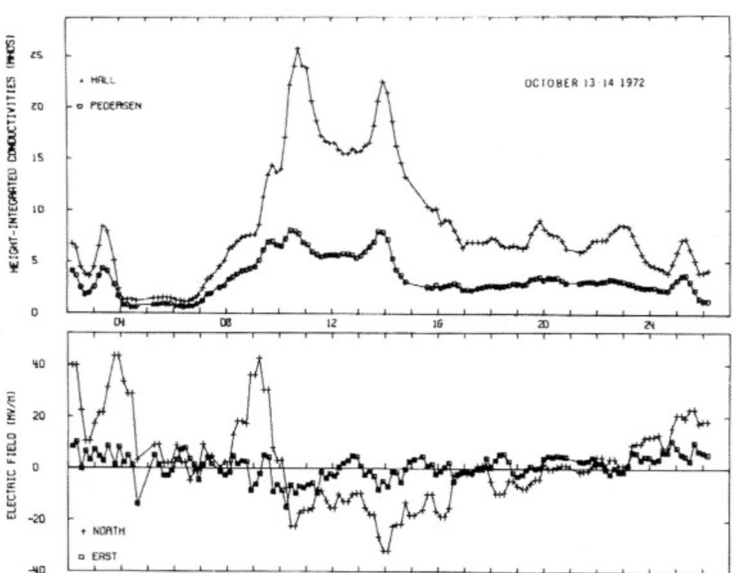


Figure D: Estimated Hall and Petersen conductivities and the observed north-south and east-west electric fields: A. Brekke, J.R. Darpnik, and P.M. Banks, J. Geophys. Res., **79**, 3773, 1974.

An enhanced westward electrojet must be associated with Boström's Type I current in Figure B. Figure E shows the projection of the electrojet on the equatorial plane; the westward electrojet in the ionosphere is connected to the eastward current on the equatorial plane. The eastward current on the equatorial plane tends to reduce the combined ring current and the magnetotail current; both are directed westward. This reduction can cause the "dipolarization" discussed in Section 3.4 and pages 105-108. However, the intense westward current can actually reverse the current direction at a distance of 5 and 8 earth radii at an early epoch of the expansive phase. This must cause the "over-dipolarization" which was discussed in pages 105-106.

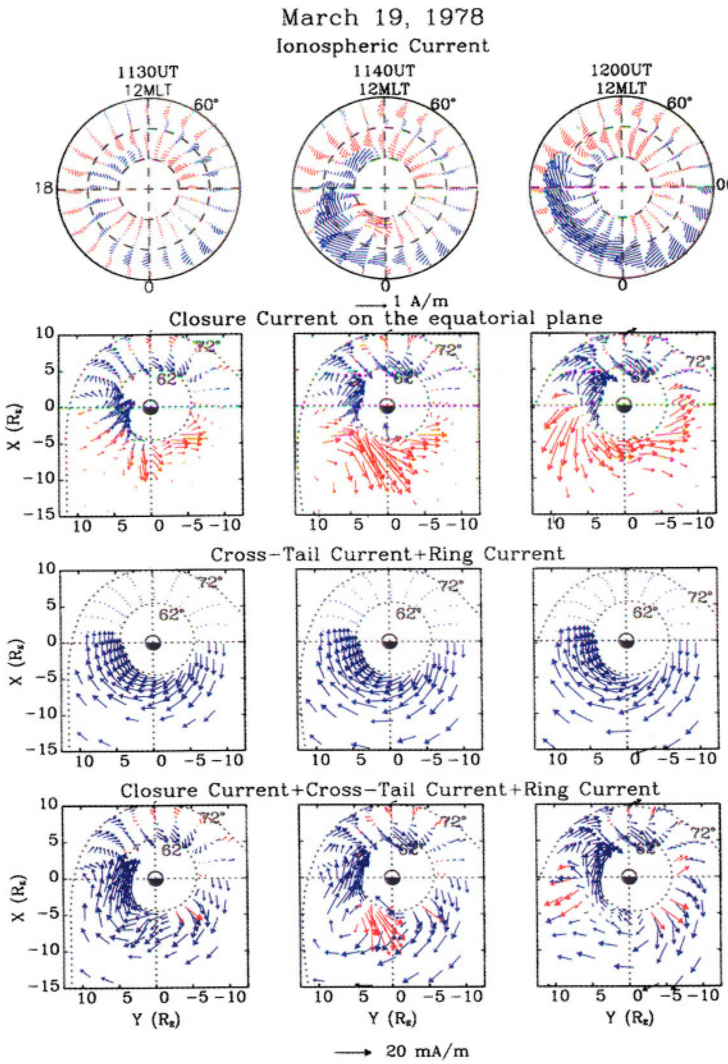


Figure E: From the top, the ionospheric currents, their projection on the equatorial plane, the cross-tail current, the ring current, and the total current, at 11:30, 11:40,and 12:00 UT on March 19, 1978. Blue arrows indicate westward currents and red arrows indicate eastward currents. For details, see Figure 3.11a on page 104.

In the past, we have long thought that the diversion of the magnetotail (westward) current is the cause for the westward electrojet (Figure F). The present consideration does not require such a diversion. The "over-dipolarization" is an important evidence for this conclusion. *The ionosphere is directly responsible for the "over-dipolarization."*

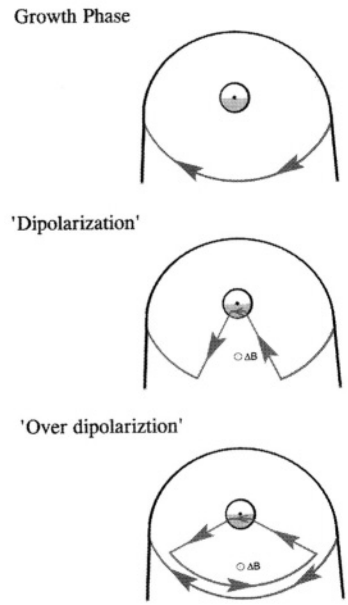


Figure F: In the past, it has been believed that the enhanced cross-tail current during the growth phase is disrupted and diverted to cause the westward electrojet. Another possibility is that the growth of the westward electrojet develops Boström's Type I current, causing the "over-dipolarization": S.-I. Akasofu.

Figure G shows an average pattern of both the directly driven component and the unloading component of the equivalent current lines (which are similar to the flow lines of plasma; see Section 1.10 and Figures 1.15 on page 37). The driven component grows during the growth phase and remains intense so long as the magnetosphere is driven. *On the other hand, the unloading component grows suddenly at substorm onset and is impulsive (Figure H; see Figure 1.15 on page 37).*

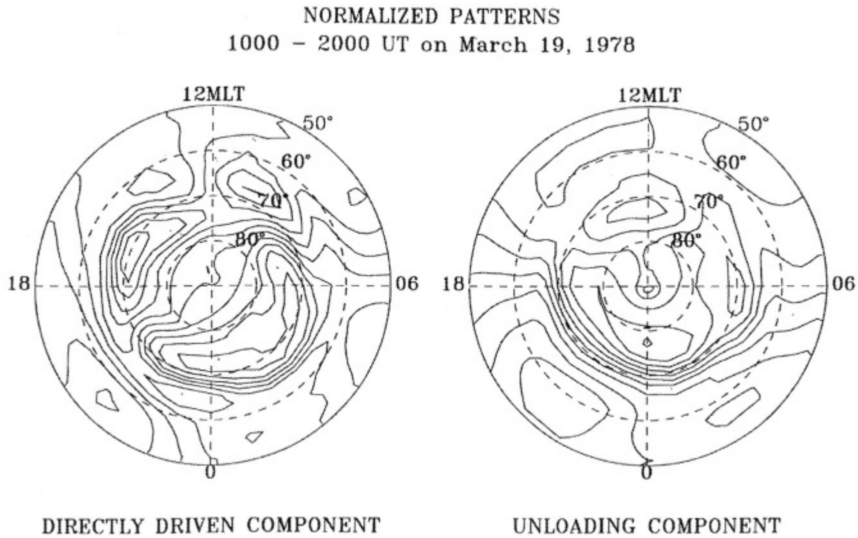


Figure G: Directly driven and unloading components of the equivalent current. The current lines are similar to the flow lines of ionospheric plasmas: For details, see Figure 1.15 on page 37.

Thus, the growth of the equatorward-directed electric field can explain the three main facts associated with the substorm onset. The next question is why the Boström's Type II current system grows suddenly and impulsively in the belt of Region I and II currents at substorm onset.

A certain percentage of substorms are caused by external 'triggering processes', such as the impact of the interplanetary shock waves and of the so-called 'northward turning' of the interplanetary magnetic field. However, a large percentage of substorms occur when $\varepsilon(t)$ is above 10^{18} erg/sec (namely, when the magnetosphere is strongly driven) without any obvious external 'triggering processes'. Thus, it is necessary to look for internal causes for most substorms.

Two important hints for the unloading process are that it is impulsive and short-lived (Figure H) and that the equatorward Pedersen current must be fed from outside the ionosphere. *This means that the radial current $\boldsymbol{J}_r$ on the equatorial plane of Boström's Type II current system must be generated by a dynamo process* ($\boldsymbol{V}$ x $\boldsymbol{B}$) on the equatorial plane.

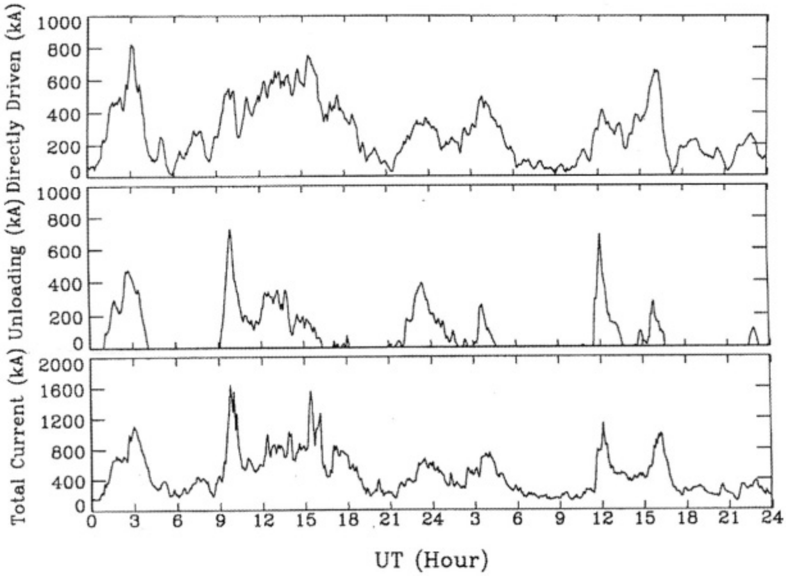


Figure H: Separation of the directly driven component and the unloading component for the period of March 18-19, 1978. Note the impulsive nature of the unloading component. For details, see Figure 1.15 on page 37.

The "over-dipolarization" can trigger the release of energy stored in the stretched field lines during the growth phase (Figure 3.11b) in the form of kinetic energy of the earthward/eastward plasma flow along a narrow belt just outside the outer radiation belt as the field lines contract. The dynamo process on the equatorial plane will be such that $\boldsymbol{E} \bullet \boldsymbol{J}_r < 0$ in the narrow belt and thus that $\boldsymbol{E}$ becomes the equatorward-directed electric field in the ionosphere. It should also be remembered that substorm onset occurs in the late evening sector; it is likely the place where the extending morning convection cell contacts the afternoon cell (see page 113) and where the first over-dipolarization occurs.

So far, we have assembled a number of observations (Step 1) and formulated a scenario for substorm onset (Step 2) on page 203. The next step is to examine quantitatively the chain of processes we considered here in the light of physics. Until then, the scenario presented in this Note will remain as a science fiction.

This page intentionally left blank

Previously published in Astrophysics and Space Science Library book series:

- **Volume 279: Plasma Astrophysics, Second Edition**
  Author: Arnold O. Benz
  Hardbound, ISBN 1-4020-0695-0, July 2002
- **Volume 277: The Sun and Space Weather**
  Author: Arnold Hanslmeier
  Hardbound, ISBN 1-4020-0684-5, July 2002
- **Volume 273: Lunar Gravimetry**
  Author: Rune Floberghagen
  Hardbound, ISBN 1-4020-0544-X, May 2002
- **Volume 272: Merging Processes in Galaxy Clusters**
  Editors: L. Feretti, I.M. Gioia, G. Giovannini
  Hardbound, ISBN 1-4020-0531-8, May 2002
- **Volume 271: Astronomy-inspired Atomic and Molecular Physics**
  Author: A.R.P. Rau
  Hardbound, ISBN 1-4020-0467-2, March 2002
- **Volume 270: Dayside and Polar Cap Aurora**
  Authors: Per Even Sandholt, Herbert C. Carlson, Alv Egeland
  Hardbound, ISBN 1-4020-0447-8, July 2002
- **Volume 269: Mechanics of Turbulence of Multicomponent Gases**
  Authors: Mikhail Ya. Marov, Aleksander V. Kolesnichenko
  Hardbound, ISBN 1-4020-0103-7, December 2001
- **Volume 268: Multielement System Design in Astronomy and Radio Science**
  Authors: Lazarus E. Kopilovich, Leonid G. Sodin
  Hardbound, ISBN 1-4020-0069-3, November 2001
- **Volume 267: The Nature of Unidentified Galactic High-Energy Gamma-Ray Sources**
  Editors: Alberto Carramiñana, Olaf Reimer, David J. Thompson
  Hardbound, ISBN 1-4020-0010-3, October 2001
- **Volume 266: Organizations and Strategies in Astronomy II**
  Editor: Andre Heck
  Hardbound, ISBN 0-7923-7172-0, October 2001
- **Volume 265: Post-AGB Objects as a Phase of Stellar Evolution**
  Editors: R. Szczerba, S.K. Górny
  Hardbound, ISBN 0-7923-7145-3, July 2001
- **Volume 264: The Influence of Binaries on Stellar Population Studies**
  Editor: Dany Vanbeveren
  Hardbound, ISBN 0-7923-7104-6, July 2001
- **Volume 262: Whistler Phenomena**
  Short Impulse Propagation
  Authors: Csaba Ferencz, Orsolya E. Ferencz, Dániel Hamar, János Lichtenberger
  Hardbound, ISBN 0-7923-6995-5, June 2001
- **Volume 261: Collisional Processes in the Solar System**
  Editors: Mikhail Ya. Marov, Hans Rickman
  Hardbound, ISBN 0-7923-6946-7, May 2001
- **Volume 260: Solar Cosmic Rays**
  Author: Leonty I. Miroshnichenko
  Hardbound, ISBN 0-7923-6928-9, May 2001

- **Volume 259: The Dynamic Sun**
  Editors: Arnold Hanslmeier, Mauro Messerotti, Astrid Veronig
  Hardbound, ISBN 0-7923-6915-7, May 2001
- Volume **258: Electrohydrodynamics in Dusty and Dirty Plasmas Gravito-Electrodynamics and EHD**
  Author: Hiroshi Kikuchi
  Hardbound, ISBN 0-7923-6822-3, June 2001
- **Volume 257: Stellar Pulsation-Nonlinear Studies**
  Editors: Mine Takeuti, Dimitar D. Sasselov
  Hardbound, ISBN 0-7923-6818-5, March 2001
- Volume **256: Organizations and Strategies in Astronomy**
  Editor: Andre Heck
  Hardbound, ISBN 0-7923-6671-9, November 2000
- Volume **255: The Evolution of the Milky Way Stars versus Clusters**
  Editors: Francesca Matteucci, Franco Giovannelli
  Hardbound, ISBN 0-7923-6679-4, January 2001
- **Volume 254: Stellar Astrophysics**
  Editors: K.S. Cheng, Hoi Fung Chau, Kwing Lam Chan, Kam Ching 1
  Hardbound, ISBN 0-7923-6659-X, November 2000
- Volume **253: The Chemical Evolution of the Galaxy**
  Author: Francesca Matteucci
  Hardbound, ISBN 0-7923-6552-6, May 2001
- **Volume 252: Optical Detectors|for Astronomy II State-of-the-art at the Turn of the Millennium**
  Editors: Paola Amico, James W. Beletic
  Hardbound, ISBN 0-7923-6536-4, December 2000
- Volume **251: Cosmic Plasma Physics**
  Author: Boris V. Somov
  Hardbound, ISBN 0-7923-6512-7, September 2000
- Volume **250: Information Handling in Astronomy**
  Editor: Andre Heck
  Hardbound, ISBN 0-7923-6494-5, October 2000
- Volume **249: The Neutral Upper Atmosphere**
  Author: S.N. Ghosh
  Hardbound, ISBN 0-7923-6434-1, (in production)
- Volume **247: Large Scale Structure Formation**
  Editors: Reza Mansouri, Robert Brandenberger
  Hardbound, ISBN 0-7923-6411-2, August 2000
- **Volume 246: The Legacy of J.C._Kapteyn Studies on Kapteyn and the Development of Modern Astronomy**
  Editors: Piet C. van der Kruit, Klaas van Berkel
  Hardbound, ISBN 0-7923-6393-0, August 2000
- Volume **245: Waves in Dusty Space Plasmas**
  Author: Frank Verheest
  Hardbound, ISBN 0-7923-6232-2, April 2000