Shock Acceleration as Source of the Anomalous Component of Cosmic Rays in the Heliosphere

Riaan Steenkamp, M.Sc.

Thesis submitted in the Department of Physics of the Potchefstroom University for Christian Higher Education for the degree Ph.D. in Physics

Supervisor: Prof. H. Moraal
Assistant Supervisor: Prof. M.S. Potgieter

POTCHEFSTROOM
SOUTH AFRICA
January 1995
"Knowledge is but an unending adventure at the edge of uncertainty."

— From *Children of Dune* by Frank Herbert —

To my wife, Anna-Marie, for her love, understanding and companionship.
Abstract

Anomalous cosmic rays are low energy enhancements of the cosmic ray intensities that cannot be explained by standard modulation of galactic cosmic rays entering the heliosphere. Presently it is thought that these anomalous cosmic rays enter the heliosphere as interstellar neutrals that are singly ionized in the inner heliosphere, convected outward to the solar wind termination shock and accelerated there to cosmic ray energies. To study this problem a numerical solution scheme is developed to solve the Parker transport equation as function of time, magnetic rigidity and two spatial dimensions. A requirement of the numerical model is that it must be able to solve the Parker equation across the solar wind termination shock to describe particle acceleration in a self-consistent way. The basic solutions produced by this model are studied to compile a comprehensive set of solutions, including the modulation and re-acceleration of galactic cosmic rays, the acceleration of a low energy source of particles and the effects of curvature and gradient drifts on these solutions. The similarities between the acceleration and modulation of different species of particles in the heliosphere are studied. The quality and characteristics of the solutions produced by the numerical model are studied in detail to demarcate the useful solution ranges of the model. It is shown that the modulation state of singly charged Helium and Oxygen during the solar minima of 1977/78 and 1987 is well explained by this model. Similarly, the model is used to address the problem of anomalous Hydrogen as a combination of the re-acceleration of galactic protons and protons accelerated at the solar wind termination shock. This confirms our present understanding of the origin of these species quantitatively, while it also demonstrates the validity of the newly developed numerical model. Hysteresis or phase lag effects between the modulation of high and low energy particles are well-known. Following several previous calculations, we solve the transport equation to determine to what extent these lags are due to time-dependent effects in the modulation.
Uittreksel

Skokversnelling as bron van die anomale komponent van kosmiese strale in die heliosfeer

Anomale kosmiese strale is verhogings in die intensiteite van lae-energie- kosmiese strale wat nie verklaar kan word deur die standaardmodulasie van galaktiese kosmiese strale wat die heliosfeer binnedring nie. Tans is die mening dat hierdie anomalie kosmiese strale die heliosfeer as interstellêre neutrale atome binnedring wat in die binneheliosfeer enkelgeioniseer, dan uitwaarts na die sonwind-terminasieskok gekonvekteer, en daar na kosmiese-straal-energie versnel word. Ten einde hierdie probleem te bestudeer, word 'n numeriese oplossingsmetode ontwikkel om die Parker transportvergelyking op te los as 'n funksie van tyd, magnetiese styftheid en twee ruimtelike dimensies. 'n Vereiste van die numeriese model is dat dit in staat moet wees om die Parker vergelyking oor die sonwind-terminasieskok op te los sodat die deeltjieversnelling op 'n selfkonsistente manier hanteer kan word. Die basiese oplossings wat deur hierdie model gegenereer word, word bestudeer om 'n omvattende studie van modulasie saam te stel. Dit sluit in die modulasie en herversnelling van galaktiese kosmiese strale, die versnelling van 'n bron van lae-energie-deeltjies en die effekte van krommings- en gradiëntdryf op die oplossings. Die ooreenkomste tussen die versnelling en die modulasie van verschillende spesies deeltjies in die heliosfeer word bestudeer. Die kwaliteit en eienskappe van die oplossings verkry met die numeriese model word in detail ondersoek om die bruikbare toepassingsgebiede van die model af te baken. Daar word aangetoon dat die modulasietoestand van enkelgeioniseerde Helium en Suurstof gedurende die sonminima van 1977/78 en 1987 goed deur hierdie model verklaar kan word. Die model word insgelyks gebruik om die probleem van anomalie Waterstof te hanteer as 'n kombinasie van die herversnelling van galaktiese protone en protone versnel by die sonwind-terminasieskok. Dit bevestig ons huidige begrip van die oorsprong van hierdie spesies kwantitatief; dit illustreer ook die geldigheid van die nuutontwikkelde numeriese model. Histerese of fasevertraginge tussen die modulasie van hoë- en lae-energie-deeltjies is welbekend. In navolging van vorige berekenings los ons die transportvergelyking op om te bepaal in watter mate hierdie vertraginge aan tydsafhanklike effekte in die modulasie toeskryfbaar is.
I wish to use this opportunity to express my gratitude to several important people, powers and institutions without which this thesis, or large parts thereof, would not have been possible.

Firstly, I am eternally grateful to our Heavenly Father for His grace to have allowed me to complete this work. Secondly, I wish to express my thanks to my long-suffering supervisor, Prof. Harm Moraal, for his advice, support and patience during the course of this study. The long hours he spent poring over my work despite his hectic schedule is much appreciated. Thirdly, my thanks to Prof. Marius Potgieter, who acted as assistant supervisor, for his continuing interest and advice during the development of the numerical model. I also wish to thank the FRD and the Department of Physics at the PU for CHE for financial support. Further, I want to thank the following people at the Department of Physics for invaluable discussions on a wide range of topics relating to my work. They are Profs. Adri Burger, Jakkals Reinecke and Okkie de Jager. A special word of thanks to Conrad Steenberg who was interested enough to take my numerical model and apply it to an independent problem. This application became a full-fledged M.Sc. thesis.

I want to thank Prof. J.R. Jokipii for much insight and useful discussions on the theoretical aspects of particle acceleration, modulation and drift during his visit to Potchefstroom in 1992. I am grateful to Mrs. Petro Sieberhagen for taking many noisome administrative tasks from my hands.

Thank you to the Head of Department at the Physics Department at the University of Namibia, Prof. Detlof von Oertzen, for his patience in allowing me to be absent from duties for several months to complete my thesis.

I wish to thank my parents for their love and my parents-in-law for their love as well as putting up with my wife and I for first one month in January 1994, another one in June/July 1994 and an additional 3 months from November 1994 to January 1995. Without their hospitality everything would have been more complicated.

My thanks also to my friends, Willem, Harry, Frans, Fanie, Laurette, Corné, Louis, Stef, Estie and Okkie, and others who befriended me during the eight years of my stay in Potchefstroom. It would have been a dreary time without them.

Thanks to Jan van Rooy and Anton Opperman, both system programmers at the University’s Department of Information Technology and Administration, whose expert advice and help was much appreciated during my term in caring for the Department of Physics’ IBM RS/6000 workstations.

Last, but not least, Anna-Marie, my wife, thank you for checking up on my English, and thank you for your love and understanding for the hours that I sometimes kept.

This thesis was typeset with the \TeX\ document formatting system, operating with Eberhard Mattes’ excellent \TeX\ implementation. This was done on a PC running IBM’s award winning OS/2 operating system. My utmost gratitude and respect to the Free Software Foundation and all those people on the Internet who selflessly design excellent software and publish it under the Free Software Foundation’s GNU Public License to provide a multitude of people with free
software of excellent quality. Long live the Free Software Foundation!

*Soli Deo gloria!*

Riaan Steenkamp
Potchefstroom 1995
Contents

1 Introduction 1

2 Structure of the Heliosphere 5

\hspace{1em} 2.1 Introduction .................................................. 5

\hspace{2em} 2.2 The Solar Wind ................................................. 6

\hspace{3em} 2.2.1 The Parker Solution ........................................... 6

\hspace{3em} 2.2.2 Solar Wind Observations ..................................... 9

\hspace{3em} 2.2.3 The Solar Wind Termination Shock ......................... 10

\hspace{3em} 2.2.4 Structure Outside the Shock .................................. 14

\hspace{2em} 2.3 The Interplanetary Magnetic Field .......................... 14

\hspace{3em} 2.3.1 The Parker Spiral Field .................................... 15

\hspace{3em} 2.3.2 The Jokipii-Kóta Modification ............................... 17

\hspace{3em} 2.3.3 The Moraal Modification .................................... 17

\hspace{3em} 2.3.4 The Smith and Bieber Modification ......................... 18

\hspace{3em} 2.3.5 The Wavy Neutral Sheet .................................... 18

\hspace{3em} 2.3.6 Solar Cycle Variations ........................................ 19

\hspace{3em} 2.3.7 The IMF at and beyond the Termination Shock ............ 19

3 Cosmic Rays in the Heliosphere 21

\hspace{1em} 3.1 Introduction .................................................. 21

\hspace{2em} 3.2 The Parker Transport Equation .................................. 21

\hspace{2em} 3.3 First-Order Fermi (Shock) Acceleration ...................... 24

\hspace{2em} 3.4 Pick-Up Ions in the Solar Wind and the Acceleration of ACR 28

\hspace{2em} 3.5 The TPE in a Parker Spiral Field ............................. 28

\hspace{2em} 3.6 Drift Velocities in a Parker Spiral Field .................... 30

\hspace{3em} 3.6.1 The Drift Velocities between the Sun and the SWTS ....... 30

\hspace{3em} 3.6.2 The Drift Velocities Beyond the SWTS ..................... 31

\hspace{2em} 3.7 Neutral Sheet and Shock Drift .................................. 32

\hspace{2em} 3.8 The Neutral Sheet Drift Model .................................. 33

\hspace{3em} 3.8.1 Azimuthal Averages of Gradient and Curvature Drift ........ 33


3.8.2 Azimuthal Average of Neutral Sheet Drift ........................................ 35
3.9 The Diffusion Coefficients ................................................................. 36
  3.9.1 Diffusion Parallel and Perpendicular to the Magnetic Field .............. 36
3.10 Summary ......................................................................................... 38

4 Numerical Solution of Partial Differential Equations with Finite Difference Methods ........................................................................................................ 39
  4.1 Introduction ..................................................................................... 39
  4.2 Classification of PDEs ..................................................................... 40
    4.2.1 Classification of Second-Order PDEs ........................................ 40
    4.2.2 Classification of First-Order Equations ...................................... 40
  4.3 Basic Numerical Techniques ............................................................. 41
    4.3.1 Introduction ............................................................................. 41
    4.3.2 Finite Difference Formulae ...................................................... 42
    4.3.3 Implementation ...................................................................... 44
  4.4 Numerical Solution of Parabolic PDEs .............................................. 45
    4.4.1 PDEs in One Spatial Dimension ............................................. 46
    4.4.2 The Thomas Algorithm ......................................................... 49
    4.4.3 Two-Dimensional Equations ................................................... 51
    4.4.4 Stability Considerations ......................................................... 55
  4.5 Methods for Solving First-Order Hyperbolic PDEs .......................... 56
  4.6 Existing Solutions of the Transport Equation .................................... 58
    4.6.1 Steady-State Non-Acceleration Solutions ............................... 59
    4.6.2 Steady-State Acceleration Solutions ....................................... 61
    4.6.3 The TPE Dilemma with Time-Dependent Solutions ................... 61
    4.6.4 Time-Dependent Non-Acceleration Solutions .......................... 62
    4.6.5 Le Roux's Modified 2D ADI Solution ..................................... 63
    4.6.6 Time-Dependent Acceleration Solutions ................................. 63
  4.7 LOD Solution of the Time-Dependent TPE ...................................... 64
    4.7.1 Solving the Radial Equation .................................................... 64
    4.7.2 Solving the Polar Equation ..................................................... 67
    4.7.3 Solving the Energy Equation .................................................. 70
  4.8 LOD Solution of the Time-Dependent TPE with a Discontinuity ........ 72
    4.8.1 The Finite Difference Form of the Matching Condition ............. 73
    4.8.2 Solving the Radial Equation .................................................... 75
    4.8.3 Solving the Polar Equation ..................................................... 76
    4.8.4 Solving the Energy Equation .................................................. 76
  4.9 Summary of Algorithm .................................................................... 77
## 5 Basic Solutions

5.1 Introduction ........................................ 80
5.2 The Model Heliosphere and its Transport Parameters .......... 81
5.3 A Steady-State Modulation Solution for Protons ................. 85
5.4 A Time-Asymptotic Modulation Solution for Protons .............. 92
5.5 The Acceleration of Cosmic-Ray Protons by the SWTS ......... 96
  5.5.1 The Re-Acceleration of Galactic Cosmic-Ray Protons by the SWTS 97
  5.5.2 Acceleration of a Low-Energy Proton Source by the SWTS .... 102
  5.5.3 Combined Source Acceleration and Re-Acceleration of Protons by the SWTS 106
  5.5.4 A Linear Combination of Proton Source Acceleration and Re-Acceleration Solutions 110
5.6 Drift Solutions/Effects .................................. 111
  5.6.1 The Effects of Drift on Modulation Solutions .............. 112
  5.6.2 The Effects of Drift on Acceleration Solutions ............. 119
5.7 The Acceleration of Cosmic-Ray Oxygen by the SWTS ............ 128
  5.7.1 Solutions for Heavier Species .......................... 128
  5.7.2 Application to Model Results ........................... 133
  5.7.3 A Linear Combination ................................. 136

## 6 Limitations and Properties of the Solution

6.1 Introduction ........................................... 138
6.2 Limitations of the Model .................................. 138
  6.2.1 Theoretical Stability .................................. 139
  6.2.2 Theoretical Accuracy .................................. 140
  6.2.3 The Control Solution ................................... 141
  6.2.4 The Radial Grid and Domain ............................. 141
  6.2.5 The Polar Grid and Domain .............................. 146
  6.2.6 The Rigidity Grid and Domain ........................... 146
  6.2.7 The Time Grid and Domain .............................. 150
  6.2.8 Grid Ratios .......................................... 152
  6.2.9 Summary of Mesh and Domain Considerations .............. 153
6.3 Parameter Variations ...................................... 154
  6.3.1 Diffusion Coefficients ................................. 154
  6.3.2 The Outer-Boundary Radius .............................. 155
  6.3.3 The Position of the Shock and the Total Domain of Modulation 157
6.3.4 Drift Beyond the SWTS ........................................ 157
6.3.5 The Diffusion Coefficient Beyond the Shock .................. 158
6.3.6 Variation in the Compression Ratio .............................. 158
6.3.7 The Structure of the SWTS .................................... 159

7 Applications of the Solution .................................. 164
  7.1 Introduction .................................................. 164
  7.2 Acceleration of the Anomalous Component ........................ 164
     7.2.1 Acceleration and Modulation of Anomalous Helium and Oxygen in 1977 and 1987 .................................. 165
     7.2.2 Anomalous Hydrogen .................................... 167
  7.3 Time-Dependent Effects ........................................ 169
     7.3.1 Hysteresis Effects ..................................... 170
     7.3.2 Temporal Effects in Acceleration Models ................. 175
  7.4 Summary ...................................................... 180

8 Summary and Conclusions ....................................... 181

8 Bibliography .................................................... 183
List of Figures

2.1 The heliosphere. .............................................. 6
2.2 Solar Wind Solutions ........................................... 8
2.3 Single-fluid solar wind solution. ............................... 9
2.4 Single-fluid solar wind solution with shock. .................. 11
2.5 Structure of a planar hydrodynamic shock. ..................... 12
2.6 Deformations of the SWTS. ................................... 14
2.7 The photospheric magnetic field. .............................. 18
2.8 The wavy neutral sheet. ....................................... 19
3.1 Drift mechanisms. ............................................. 32
3.2 Drift along the wavy neutral sheet. ............................ 34
3.3 The wavy neutral sheet as function of $\phi$. .................. 34
3.4 Distribution of neutral sheet drift as function of $\theta$. ..... 35
3.5 $\lambda_{||}$ vs. $P$ ............................................. 37
4.1 Pure initial, initial-boundary, and pure boundary value problems. 42
4.2 Implementing the forward-difference explicit method. ........ 47
4.3 Implementing the backward-difference implicit method. ....... 48
5.1 A 1/4 heliosphere .............................................. 82
5.2 Solar Wind Profile ............................................. 82
5.3 Steady-state spectra. ......................................... 87
5.4 Steady-state radial profiles. ................................. 88
5.5 Steady-state polar profiles. ................................... 89
5.6 Time-asymptotic spectra. ..................................... 93
5.7 Time-asymptotic radial profiles. .............................. 94
5.8 Time-asymptotic polar profiles. ............................... 95
5.9 Re-acceleration spectra. ....................................... 98
5.10 Re-acceleration radial profiles. .............................. 99
5.11 Re-acceleration polar profiles. ............................... 100
5.12 Source acceleration spectra. ................................. 103
5.13 Source acceleration radial profiles. ............................................. 104
5.14 Source acceleration polar profiles. ............................................. 105
5.15 Combined source acceleration and re-acceleration spectra. ............... 107
5.16 Combined source acceleration and re-acceleration radial profiles. ......... 108
5.17 Combined source acceleration and re-acceleration polar profiles. ........... 109
5.18 Linear combination for protons. ............................................... 111
5.19 Time-asymptotic spectra for the positive drift case. ....................... 113
5.20 Time-asymptotic radial profiles for the positive drift case. ............... 114
5.21 Time-asymptotic polar profiles for the positive drift case. ............... 115
5.22 Time-asymptotic spectra for the negative drift case. ...................... 116
5.23 Time-asymptotic radial profiles for the negative drift case. ............... 117
5.24 Time-asymptotic polar profiles for the negative drift case. ............... 118
5.25 Source acceleration spectra for the positive drift case. ................... 120
5.26 Source acceleration radial profiles for the positive drift case. .......... 121
5.27 Source acceleration polar profiles for the positive drift case. .......... 122
5.28 Source acceleration spectra for the negative drift case. ................... 123
5.29 Source acceleration radial profiles for the negative drift case. .......... 124
5.30 Source acceleration polar profiles for the negative drift case. .......... 125
5.31 The Jokipii shock spectra. .................................................... 126
5.32 Re-acceleration spectra for $A/Z = 2$. .................................... 129
5.33 Scaled spectra of ACR He, C, N, Ne, Ar and O. ............................ 130
5.34 Source acceleration spectra for $A/Z = 16$. ................................ 131
5.35 $\beta P$ as function of energy. ............................................. 133
5.36 Scaling factors as function of energy. ...................................... 134
5.37 Energy scaling factor versus mass number. .................................. 135
5.38 Scaled model spectra of H+, He+ and O+. .................................. 136
5.39 Linear combination for Oxygen. ............................................ 137

6.1 Control spectra. ............................................................ 142
6.2 Control radial profiles. ...................................................... 143
6.3 Control polar profiles. ...................................................... 144
6.4 Radial grid effects. ......................................................... 145
6.5 Polar grid effects. ........................................................... 147
6.6 Coarse rigidity grid on no-drift solutions. .................................. 148
6.7 Instabilities arising from drift. ............................................ 149
6.8 Inaccuracies arising from a coarse time grid. ................................ 151
6.9 Boundary effects in the method of Characteristics .......................... 152
6.10 Inaccuracies due to insufficient convergence. ................................ 153
6.11 Doubled diffusion coefficients. ................................................. 155
6.12 Variation of the boundary, rt. .................................................. 156
6.13 Both dimensions and diffusion coefficients doubled. ....................... 157
6.14 Diffusion coefficients α r beyond the shock with a factor s = 4 drop for all θ. 159
6.15 A compression ratio s = 2. ......................................................... 160
6.16 Energy spectra for the smooth energy dependence of κ for discontinuous and continuous transitions. ....................................... 162
6.17 Energy spectra for the ‘kinked’ energy dependence of κ for discontinuous and continuous transitions. ....................................... 163

7.1 Observed and calculated spectra of He⁺ and O⁺. ............................... 166
7.2 Observed and calculated spectra of protons. .................................. 168
7.3 The effective radial κrr and the cyclic function as function of radial distance and time. ............................................................... 171
7.4 Hysteresis loops and temporal variation in intensity for an 11-year variation. 173
7.5 Temporal variation in intensity and hysteresis loop for a short pulse. .... 174
7.6 The 10, 100 and 1000 MeV modulated intensities at 1, 23, 42 and 90 AU for an 11-year variation in the diffusion coefficients for a time-dependent pure proton modulation model. .................................................. 176
7.7 The 10, 100 and 1000 MeV modulated intensities at 1, 23, 42 and 90 AU for an 11-year variation in the diffusion coefficients for a time-dependent source acceleration model for protons. .................................................. 177
7.8 The 10, 100 and 1000 MeV modulated intensities at 1, 23, 42 and 90 AU for an 11-year variation in the diffusion coefficients for a time-dependent re-acceleration model for protons. .................................................. 178
7.9 The 10, 100 and 1000 MeV modulated intensities at 1, 23, 42 and 90 AU for an 11-year variation in the diffusion coefficients for a time-dependent source acceleration/re-acceleration model for protons. .................................................. 179
Chapter 1
Introduction

The study of cosmic rays began with their discovery during the balloon flights of Victor F. Hess early in the second decade of this century (Hess, 1911, 1912). Even if the existence of cosmic rays was inferred before this, these measurements constituted the first real evidence of the existence of cosmic rays.

Initially, cosmic rays — named as such by Millikan in 1925 — were assumed to be electromagnetic rays with more penetration power than $\gamma$-rays. However, in the following decade it was proved that cosmic rays are energetic charged particles originating in space. The predominant component of these cosmic rays is protons, constituting approximately 90% of the particles. The most abundant component of the remaining 10% are $\alpha$-particles. The remainder is made up of heavier nuclei like Nitrogen, Oxygen, Carbon and Neon, and even nuclei as heavy as Iron ($Z = 26$).

Most of the cosmic rays originate outside the heliosphere. Where and how these cosmic rays are produced, is still not certain, although it is thought that first-order Fermi acceleration in astrophysical structures like supernova remnants in our galaxy are prime candidates. Some cosmic rays may even be of extra-galactic origin. However, since discussions on the origins of cosmic rays in the galaxy fall outside the scope of this thesis, it will not be discussed further, and the particles will be simply referred to as galactic cosmic rays.

Apart from these galactic cosmic rays, cosmic rays are also produced in the heliosphere. Some are of solar origin, being produced by solar flares, and some originate elsewhere in the heliosphere. This thesis will study this latter component in detail. Present knowledge indicates that those of heliospheric origin may be far more useful in probing the structure of the heliosphere than the others, since they are the most sensitive to the conditions inside the heliosphere. Examples of heliospheric cosmic rays are particles associated with co-rotating interaction regions and the so-called anomalous component of cosmic rays.

Of particular interest are the processes through which the intensity of all these types of cosmic rays is modulated by conditions inside the heliosphere. These conditions cause the cosmic rays in the heliosphere to suffer adiabatic energy losses and their trajectories are affected by pitch angle scattering, convection and drift in the large scale magnetic field of the sun. By studying the resultant energy spectra, spatial distributions, and temporal changes in the intensity of these modulated cosmic rays, the structure of the heliosphere and the processes operating therein may be deduced.

Initially, these cosmic rays could only be studied with ground-based neutron monitors and cosmic-ray detectors in airplanes and balloons. With the advent of the space age, cosmic rays could be studied in space itself with cosmic ray detectors on satellites and interplanetary spacecraft. In addition to this, the spacecraft carry instruments with which the solar wind and interplanetary magnetic field may be measured.

The Pioneer 10 and 11 spacecraft were launched in 1972 and 1973 respectively. The trajectory
of Pioneer 10 took it to Jupiter, and with the aid of a slingshot maneuver, using the gravitation of the giant planet, it was flung towards the outer reaches of the heliosphere where it is presently continuing well beyond the orbit of Pluto. Pioneer 11’s trajectory took it past Jupiter, from where it was flung towards Saturn. Like was done at Jupiter, a slingshot maneuver was employed to fling the spacecraft out of the heliosphere in the direction of the sun’s trajectory through interstellar space. At present, it is also continuing beyond the orbit of Pluto.

In 1976 and 1977 the Voyager 1 and 2 spacecraft were launched on similar missions. As was the case with Pioneer 11, Voyager 1’s trajectory carried it past Jupiter and Saturn before it headed out of the heliosphere, where it is still continuing beyond Pluto’s orbit. Voyager 2, on the other hand, passed not only Jupiter and Saturn, but also Uranus and Neptune from where it was flung towards interstellar space. In January 1995, Voyager 1 and 2, and Pioneer 10 will be about 58, 45 and 61 astronomical units from the sun. Unfortunately, Pioneer 11 will no longer transmit useful data after April 1995 and Pioneer 10 will cease to do so beyond 1998.

In 1990 the Ulysses spacecraft was launched. Like the Pioneer and Voyager spacecraft, the gravity of Jupiter was used to fling the spacecraft into its predetermined orbit. Instead of flinging it towards another planet, this spacecraft was put into a trans-polar solar orbit, which carried it over the south polar region of the sun in June 1994, and it is presently on a trajectory towards the northern-pole in June 1995.

The cosmic-ray observations transmitted back from these spacecraft are supplemented by satellites in earth orbit, such as the IMP series of satellites. The combined results of Pioneer 10 and 11, Voyager 1 and 2, Ulysses and IMP8 over the last two decades have given the greatest strides in the observations of the properties of the heliosphere and the transport of cosmic rays therein.

One of the most fascinating discoveries made with the aid of these spacecraft and satellites, is the discovery of the anomalous component of cosmic rays when García-Munoz et al. (1973a, 1973b, 1973c) measured anomalously high intensities of Helium, which could not be explained within the scope of solar modulation theory. Investigations by Hovestadt et al. (1973) and McDonald et al. (1974) yielded similar anomalously high intensities at low energies in the Oxygen and Nitrogen spectra.

Fisk et al. (1974) realized that these three elements (He, O, N) all have high first ionization potentials. They proposed that, due to these high first ionization potentials, these elements exist in interstellar space as neutral atoms, which can enter the heliosphere with a relative speed of 25 km/s due to the sun’s movement through interstellar space. These neutrals may then penetrate deeply into the heliosphere, before they become singly ionized in the inner heliosphere. This ionization can be due to photo-ionization near the sun, charge exchange with the solar wind plasma or electron collisions.

These singly-ionized atoms are then convected to the outer heliosphere where they are somehow accelerated to cosmic-ray energies. Thus, the anomalous component of cosmic rays is not of galactic origin but originates inside the heliosphere itself. This has some extraordinary consequences, since these anomalous cosmic rays are more responsive to heliospheric conditions than the higher energy galactic cosmic rays, which makes them an invaluable tool with which the processes in the heliosphere can be probed. In addition to this, the anomalous component may have a higher energy density in the outer heliosphere than galactic cosmic rays.

Further experiments detected anomalous components in the \(^{4}\text{He}, \, ^{15}\text{N}, \, ^{16}\text{O}\) and Ne spectra (García-Munoz et al., 1975; Mewaldt et al., 1975; von Rosenvinge and McDonald, 1975). Anomalous components were also discovered in the Ar and C spectra (Cummings and Stone, 1987).

Christian et al. (1988) reported the existence of anomalous Hydrogen. However, conclusive evidence about the extent that anomalous Hydrogen can be detected in the proton spectra does
not yet exist, since it was demonstrated by Reinecke and Moraal (1992) that simple proton modulation can account for the bulges in the proton spectra that led Christian et al. to believe that they had detected anomalous Hydrogen. However, Möbius et al. (1985) and Gloeckler et al. (1993) did detect the pick-up of freshly ionized Hydrogen ions with experiments on board the Ulysses spacecraft, which are the source of the anomalous component.

Pesses et al. (1981) proposed that the primary mechanism by which the above ions can be accelerated is diffuse shock acceleration at the solar wind termination shock. To test this theory, one needs solutions of the cosmic-ray transport equation (TPE). Since the TPE cannot be solved analytically for physically realistic coefficients, it is necessary to build numerical models with which it can be solved.

The first of these numerical models was developed by Fisk (1971) who solved the steady-state, spherically symmetric TPE, yielding a distribution function as function of energy and distance from the sun. The second model was also designed by Fisk (1973) who expanded it into another spatial dimension, the polar angle. Moraal and Gleeson (1975) improved on Fisk’s two-dimensional model by adding more physics to the diffusion coefficients. At the same time Cecchini and Quenby (1975) presented an independently developed two-dimensional energy dependent model.

The year 1977 saw the addition of particle drifts when Jokipii et al. (1977) performed the first drift calculations with a two-dimensional steady state model. Moraal et al. (1979), and Jokipii and Kopriva (1979a,b) simultaneously presented drift models at the 16th International Cosmic Ray Conference. Jokipii and Davila (1981) improved their two-dimensional drift model with the addition of more physically accurate energy dependences of the diffusion coefficients. Jokipii and Thomas (1981) modelled the transport equation with a warped neutral sheet.

Kóta and Jokipii (1983) expanded their two-dimensional steady state drift model into all three spatial dimensions. In the same year, Perko and Fisk (1983) published results from a time-dependent spherically symmetric model with which time-dependent effects may be studied.

Potgieter (1984) improved the two-dimensional steady-state drift model of Moraal et al. (1979) with more versatile handling of the neutral sheet drift problem. Another successful two-dimensional steady state drift model was made by Kadokura and Nishida (1986).

Jokipii (1986) published the first numerical model that can model the acceleration of particles at the solar wind termination shock with the aid of a fully time-dependent two-dimensional solution of the TPE. A second acceleration model was a spherically symmetric steady-state acceleration model developed by Potgieter and Moraal (1988).

Le Roux (1990) expanded Perko and Fisk’s (1983) technique to build a two-dimensional time-dependent model with which the modulation of cosmic rays could be studied as function of time. However, unlike Jokipii’s model, this model cannot accelerate particles at the solar wind termination shock. Kóta and Jokipii (1991) succeeded to build a fully time-dependent, three-dimensional model that can accelerate and modulate particles simultaneously.

Thus, the only acceleration models to date are those developed by the Jokipii-Kóta group, and the one-dimensional model of Potgieter and Moraal (1988). An attempt by Moraal to expand the steady-state Potgieter and Moraal acceleration model to two dimensions failed due to unforeseen numerical problems. Therefore, it is the purpose of this thesis to independently develop a fully time-dependent two-dimensional acceleration model with which the acceleration and modulation of the anomalous component of cosmic rays can be studied.

This numerical model is developed in Chapter 4 of this thesis with a complete discussion of the numerical mathematics on which this solution is ultimately based. The physical basis of this solution is described in Chapters 2 and 3. In Chapter 2 the structure of the heliosphere is de-
scribed, with specific reference to the solar wind and interplanetary magnetic field. In Chapter 3 the fundamentals of cosmic-ray transport, such as convection, diffusion, drifts, adiabatic energy losses, and particle acceleration at shocks are described in terms of the cosmic-ray TPE.

The basic properties of the numerical solution of the TPE are discussed in Chapter 5, where the validity of the model is demonstrated. A comprehensive summary of the modulation and acceleration processes in the heliosphere is given in this chapter. Chapter 6 explores the limitations and some properties of the numerical model and finally, in Chapter 7, some applications of the model are shown when the acceleration and modulation of anomalous Helium, Oxygen and Hydrogen are modelled and compared with experimental data. This is followed by a preliminary study of some time-dependent effects.
Chapter 2
Structure of the Heliosphere

2.1 Introduction

Main-sequence stars have an atmosphere in dynamic equilibrium, consisting of a supersonic outflow of a tenuous, highly ionized gas. Combined with the star's magnetic field, this stellar wind creates a region in which the star's influence dominates the physical processes in the vicinity of that star. Inside this sphere of influence, there exist conditions which are significantly different from that of the ambient interstellar medium.

Our own star, the sun, is no exception. It also possesses a stellar wind, called the solar wind, and a magnetic field, the interplanetary magnetic field or IMF, which dominate the local region of space. This region, in which the sun's influence changes the conditions of the interstellar medium, is called the heliosphere. Since the sun is moving through the interstellar medium at a speed of roughly 25 km/s, there exists a so-called interstellar wind in the sun's frame of reference. Figure 2.1 is adapted from Suess (1990), who recently gave a thorough description of the structure of the heliosphere.

The solar wind is a supersonic outflow of the solar atmosphere or corona. This outflow of plasma meets little resistance in the inner heliosphere, but further out the ram pressure of the interstellar gas causes a shock transition, where the solar wind velocity drops to subsonic values. This shock transition is called the solar wind termination shock (SWTS).

Around the earth there exists a so-called bow shock, created by the sudden deceleration and change in direction of the solar wind due to the magnetosphere of the earth. Likewise, it can be inferred that a similar (but very much larger) bow shock may exist around the heliosphere, where the interstellar wind is suddenly decelerated by the presence of the heliosphere and the outward flow of solar wind plasma.

Between the termination shock and the probable heliospheric bow shock lies a region of which very little is known. Since the subsonic plasma flow of the solar wind is initially directed more or less radially outward, it should meet the subsonic interstellar flow head-on on one side of the heliosphere. The two plasmas will not flow through one another and, therefore, the solar plasma will gradually change direction to flow around the heliosphere, while the interstellar plasma will flow around the heliosphere like air around a falling raindrop.

The two different plasma flows are separated by a contact surface that may be considered the true edge of the heliosphere. This contact surface is called the heliopause. On the opposite side of the heliosphere the subsonic flow of solar plasma will still be more or less radial, since the flow of interstellar plasma that has flowed around the heliosphere will be more or less parallel to that of the solar plasma. The result is that the heliosphere is not spherical at all, but rather blunt at the end facing the interstellar wind and elongated on the opposite end. Again, the analogy of a falling raindrop, shaped by the air moving past it, comes to mind.
In this thesis we study the transport, modulation and acceleration of cosmic rays in this heliosphere. Thus, in the following sections we briefly describe the profile of the solar wind plasma and the IMF in this heliosphere, as they will be used in this thesis.

2.2 The Solar Wind

2.2.1 The Parker Solution

The supersonic outflow of coronal plasma from the surface of the sun was called the solar wind by E.N. Parker, who first deduced its existence in 1958. He showed that if the temperature in an atmosphere, surrounded by a vacuum, declines less rapidly than $1/r$, it is not possible for the atmosphere of a star to be in hydrostatic equilibrium at radial distances very much larger than the radius of the star. The only steady equilibrium state is an expansion to supersonic velocities at large distances from its point of origin (Parker, 1963).

Parker determined the state of dynamic equilibrium of the solar corona (or any stellar atmosphere, for that matter) by using the following model:
- a single-fluid model for the coronal plasma, which treats the protons and electrons simultaneously;
- a temperature determined by the temperature function $T = T_0 (r_0 / r)^{2/7}$, where $\odot$ denotes the value on the surface of the sun;
- the continuity equation for the plasma density:
  $$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho V) = 0,$$
  with $V$ the plasma flow velocity and $\rho$ its density;
- the force of gravity per unit volume given by
  $$f = -\frac{GM_0 \rho}{r^2} e_r;$$
- the equation of motion of the plasma,
  $$\rho \left[ \frac{\partial V}{\partial t} + (V \cdot \nabla) V \right] + \nabla P - f = 0;$$
- the equation of state of the plasma as $P(r) = 2nkT(r)$ and $kT(r) = \frac{1}{2}mv^2(r)$. Therefore,
  $$P(r) = \rho v^2(r), \quad (2.1)$$
  with $v$ the random particle velocity.

To simplify the problem, the steady-state, spherically symmetric case is considered and, therefore,
$$\frac{\partial}{\partial t} = 0$$
and for spherical symmetry,
$$\nabla = e_r \frac{1}{r^2} \frac{\partial}{\partial \theta} r^2.$$

In addition, the plasma flow velocity is assumed to be radial, i.e., $V = V(r)e_r$. With this, the equation of continuity becomes
$$r^2 \rho V = \rho \frac{2GM_0 \rho}{r^2}, \quad (2.2)$$
while the equation of motion becomes
$$\rho V \frac{dV}{dr} + \frac{dP}{dr} + \frac{GM_0 \rho}{r^2} = 0. \quad (2.3)$$
If the escape velocity of a particle on the surface of the sun is
$$v_{\text{esc}} = \sqrt{\frac{2GM_0}{r_\odot}}, \quad (2.4)$$
the equation of motion becomes
$$\frac{dV^2}{dr} = \frac{R(r)}{1 - v^2/V^2}, \quad (2.5)$$
with
$$R(r) = -\frac{v_{\text{esc}}^2 r_\odot}{r^2} - 2r \frac{d}{dr} \left( \frac{v^2}{r^2} \right), \quad (2.6)$$

7
where equation (2.1) has been used to eliminate the density $\rho$. This equation of motion has a family of possible solutions as shown in Figure 2.2. It is physically unacceptable that there are two distinct values for the solar wind velocity, $V$, at a single position. This eliminates the possible solutions in regions A and B. Solution I, as well as those in region C, results in very large solar wind velocities as the surface of the sun is approached. However, large plasma flow velocities are not observed in the lower corona and, therefore, this set of possible solutions is not acceptable either.

Thus, the so-called 'solar breeze' solutions in region D and the so-called 'solar wind' solution, solution II, remain. The solar breeze solutions are not as easily dismissed as those in regions A, B, and C. However, the gas pressure is required to be zero very far from the surface of the sun, i.e.,

$$\lim_{r\to\infty} P = 0. \quad (2.7)$$

From (2.1), (2.3) and (2.4) it follows for the solutions in region D, as well as the discarded solution I, that

$$\frac{1}{P} \frac{dP}{dr} = -\frac{1}{2} \frac{v_{\text{esc}}^2 r_\odot}{v^2 r^2} - \frac{1}{v^2} \frac{d}{dr} \left( \frac{V^2}{2} \right).$$

This must be integrated from the sun towards infinity and, therefore, we have that

$$\frac{P_\infty}{P_\odot} = e^{-\left(I_1 + I_2\right)}, \quad (2.8)$$

with

$$I_1 = \frac{v_{\text{esc}}^2 r_\odot}{2} \int_{r_\odot}^{\infty} \frac{dr}{v^2 r^2},$$

and

$$I_2 = \int_{r_\odot}^{\infty} \frac{1}{v^2} \frac{d}{dr} \left( \frac{V^2}{2} \right) dr.$$

For (2.7) to be valid, $I_1$ and $I_2$ must go to $+\infty$ if $r \to \infty$. This is not the case, however, and, therefore, solution II is the only physically acceptable solution of (2.6).

This solution describes a plasma flow that continues to accelerate from a small value at $r_\odot$ until the flow velocity becomes supersonic at some critical radius, $r_c$, i.e., $V$ becomes greater than
the velocity of an acoustic wave (longitudinal pressure waves) in the coronal plasma

\[ c_s = \sqrt{\frac{\gamma P}{\rho}} , \]  

(2.9)

where \( \gamma = C_P/C_V \) and \( P \) the scalar pressure of the solar wind plasma. At this critical radius, the right-hand side of the equation of motion (2.5) becomes zero since \( R(r_c) = 0 \).

If an isothermal corona is assumed, the individual particle velocities must also be uniform and all have the same isothermal value, \( v_{iso} \). Since \( v = v_{iso} \) we may write

\[ R(r) = -\frac{v_{esc}^2 \sigma}{r^2} + \frac{4v_{iso}}{r} \]

and because \( R(r_c) = 0 \), and using (2.4), it follows that

\[ r_c = \frac{1}{4} \left( \frac{v_{esc}^2}{v_{iso}} \right)^2 = \frac{1}{2} \frac{GM_\odot}{v_{iso}^2} . \]

(2.10)

For a temperature of \( T \approx 10^6 \) K this yields \( r_c \approx 6r_\odot \). For this form of \( R(r) \), the equation of motion, (2.5), has the solution

\[ \frac{V^2}{v_{iso}^2} - \ln \left( \frac{V^2}{v_{iso}^2} \right) = 4 \frac{r_c}{r} + 4 \ln \left( \frac{r}{r_c} \right) - 3. \]

For large values of \( r (r \gg r_c \text{ or } V \gg v_{iso}) \) this yields

\[ V \approx 2v_{iso} \sqrt{\ln \left( \frac{r}{r_c} \right)} \]

(2.11)

which is a very slowly rising function of \( r \) at large radial distances, as shown in Figure 2.3.

Figure 2.3: Solution of the solar wind based on an isothermal single-fluid solution.

Even if more accurate two-fluid derivations exist, they do not significantly change the results of this single-fluid isothermal approach. Observations in the ecliptic plane verify the qualitative validity of this model. At the position of the earth the solar wind velocity has already started to flatten out, as predicted by equation (2.11), to a value of \( \approx 400 \text{ km/s} \).

2.2.2 Solar Wind Observations

Although the Parker solution of the previous section correctly describes the origin and the overall dynamics of the solar wind, it does not contain its detailed properties.
First of all, the solar wind contains irregularities of all scales and sizes, which contribute to the irregularities in the IMF (to be discussed later in this chapter), and these field irregularities scatter cosmic rays.

On the average, it seems that in the equatorial plane the solar wind speed stays remarkably constant out to radial distances of about 60 AU, as observed by the Pioneer and Voyager spacecraft.

On time scales of the order of days, however, the wind is not smooth but it contains high and low speed streams. High speed streams originate in so-called coronal holes, e.g., Zirken (1977). Hundhausen (1993) showed that slow solar wind regions are associated with coronal mass ejections (CMEs) occurring in closed field regions on the solar surface.

Coronal holes are semi-permanent over the solar poles, while according to Hundhausen, CMEs are correlated with the position of the heliomagnetic equator. During solar minimum periods at least, this equator nearly coincides with the heliographic equator or the ecliptic plane. Thus, it has long been suspected that the solar wind velocity will rise towards the poles.

Barnes (1989) indeed observed that in 1987 the solar wind blew some 250 km/s harder at 18° North than in the ecliptic plane. In addition, interplanetary scintillation (IPS) measurements by, e.g. Sheeley et al. (1991), also gave indirect evidence of this increase in solar wind velocity with heliolatitude.

These predictions have recently been directly confirmed on the Ulysses mission. McKibben et al. (1994) quote, for instance, that the solar wind speed increases from about 400 km/s to about 750 km/s between latitudes 15° and 40° South.

With increasing solar activity, the heliomagnetic equator becomes strongly inclined relative to the ecliptic plane. Thus, successions of high- and low-speed streams will probably occur over a much wider range of heliolatitudes during solar maximum. Due to solar rotation, however, one expects that during solar maximum the 26-day solar rotation average of the wind speed will be much more uniform with heliolatitude than during solar minimum.

Burlaga et al. (1993), pointed out that successive fast and slow solar wind streams form co­rotating interaction regions (CIRs). These CIRs collide and merge as they propagate outwards to form merged interaction regions (MIRs). When they have sufficient latitudinal and/or azimuthal extent, they develop into structures called global merged interaction regions (GMIRs).

These structures in the wind, including their tendency to form traveling interplanetary shocks, have profound modulation effects on cosmic rays. In this thesis, however, the influence of these structures in the solar wind on cosmic-ray transport will not be discussed further.

2.2.3 The Solar Wind Termination Shock

A supersonic flow cannot decelerate into subsonic flow in a continuous way. Thus, the supersonic flow energy must be dissipated discontinuously. This discontinuity in supersonic to subsonic flow is called a shock.

Consider a household sink with a running tap. The stream of water coming from the tap hits the bottom of the sink and the water flows more or less radially away from that point. Practically all the kinetic and potential energy that the stream of water had when it exited from the tap is now converted to kinetic energy and this fluid flow on the sink bottom is faster than the spread of small amplitude waves on the water surface, i.e., 'supersonic'. This flow energy in the sink is now dissipated by cohesive and viscous frictional forces and the flow energy drops so low that the flow has to undergo a shock transition to become subsonic. The surplus flow energy is
converted into turbulence beyond the shock.

The solar wind presents us with a similar problem: The flow energy of the coronal plasma is dissipated by the ram pressure of the interstellar gas in the outer heliosphere. At the point where the flow energy of the solar wind is dissipated sufficiently for supersonic flow to become impossible, the solar wind undergoes a shock transition where its velocity drops to subsonic values. This is shown in Figure 2.4 where the solar wind solution II, derived in Section 2.2.1, suddenly drops to subsonic values at radial distance $r_s$. As in the hydrodynamic equivalent, the surplus flow energy is converted into thermal energy and turbulence.

![Figure 2.4: Solution of the solar wind based on an isothermal single-fluid solution, with a shock transition at $r_s$ due to the ram pressure of the interstellar gas.](image)

**A Hydrodynamic Analysis**

Astrophysical shocks have much in common with hydrodynamic shocks due to the fluid-like models describing highly ionized plasmas. Many physical quantities and characteristics are therefore shared between astrophysical and hydrodynamic shocks. Thus, to define and derive the shock properties, it is instructive to start with an ordinary hydrodynamic shock. See Boyd and Sanderson (1969) and Jones and Ellison (1991) for more a more detailed discourse on the subject.

The velocity of a disturbance in such a fluid is conveniently described by the Mach number

$$M = \frac{V}{c_s},$$

with $c_s$ the velocity of longitudinal pressure waves as given by (2.9). Consider a plane, stationary shock. In Figure 2.5 it is placed at $x = 0$ and, therefore, the *upstream* (unshocked) flow at $x < 0$ is into the shock with supersonic velocity $V_1$, or Mach number

$$M_1 = \frac{V_1}{c_s} > 1.$$  

The *downstream* (shocked) plasma, at $x > 0$, recedes with subsonic Mach number

$$M_2 = \frac{V_2}{c_s} < 1.$$
Figure 2.5: The general structure of a simple planar hydrodynamic shock.

The mass, momentum, and energy fluxes across the shock must be conserved. These can be written as

\[
\frac{\partial}{\partial x}(\rho V) = 0, \quad (2.12)
\]

\[
\frac{\partial}{\partial x}(\rho V^2 + P) = 0 \quad (2.13)
\]

and

\[
\frac{\partial}{\partial x} \left( \frac{1}{2} \rho V^3 + \frac{\gamma}{\gamma - 1} VP \right) = 0, \quad (2.14)
\]

respectively, with \( \rho \) the gas density and \( P \) the gas pressure. If these equations are integrated across the shock, the so-called Rankine-Hugoniot boundary conditions on the shock are obtained:

\[
\rho_1 V_1 = \rho_2 V_2 \quad (2.15)
\]

\[
\rho_1 V_1^2 + P_1 = \rho_1 V_1^2 + P_1 \quad (2.16)
\]

\[
\frac{1}{2} \rho_1 V_1^3 + V_1 P_1 + \rho_1 V_1 U_1 = \frac{1}{2} \rho_2 V_2^3 + V_2 P_2 + \rho_2 V_2 U_2 \quad (2.17)
\]

with

\[
U = \frac{P}{\rho(\gamma - 1)}
\]

the internal energy of the fluid.

The compression ratio or shock ratio of the shock is defined to be

\[
s = \frac{\rho_2}{\rho_1},
\]

and from (2.15) it follows that

\[
s = \frac{V_1}{V_2} \quad (2.18)
\]

After a considerable amount of algebra with (2.15), (2.16) and (2.17), the compression ratio may be written in terms of the upstream Mach number as

\[
s = \frac{(\gamma + 1) M_1^2}{(\gamma - 1) M_1^2 + 2}.
\]
For a strong shock \((M_1 \to \infty)\) this reduces to

\[ s = \frac{\gamma + 1}{\gamma - 1}. \]  

(2.19)

For monatomic gasses, such as the solar wind plasma, \(\gamma = C_P/C_V = 5/3\) and, therefore, the compression ratio of a strong shock is \(s = 4\). For relativistic flows \(\gamma\) reduces to \(4/3\) and the compression ratio may increase to \(s = 7\).

**Astrophysical MHD Shocks**

MHD or *magnetohydrodynamic shocks* are defined as shocks in media (usually plasmas) that contain magnetic fields. In this case the relevant magnetic field pressure and energy terms must be added to (2.16) and (2.17). When this field is parallel to the normal on the shock front, the MHD shock is called parallel. In this case the field is continuous across the shock and it has no effect on the shock structure. The general problem is described by, e.g., Boyd and Sanderson (1969) and in the case of non-parallel shocks the field leads to modifications of the standard Rankine-Hugoniot conditions and the shock parameters.

In astrophysics there are many examples of systems that project large amounts of plasma at supersonic velocities. Some of these are stellar and galactic winds, and the shell of matter projected by a supernova explosion. Due to energy considerations, the ram pressure of the interstellar and intergalactic media must force these supersonic flows to subsonic velocities through shocks. More locally, astrophysical shocks may be found in places such as the bow shock of the earth in the solar wind, as well as travelling interplanetary shocks at the leading and trailing edges of CIRs, MIRs and GMIRs.

Since all systems in astrophysics contain magnetic fields, these are all examples of MHD shocks.

**Deformation of the Solar Wind Termination Shock**

A uniform and spherically symmetric solar wind, as well as a uniform interstellar gas pressure on all sides of the heliosphere, would result in a spherical shock at a constant radius around the sun. Since the solar wind velocity rises towards the polar directions, the flow energy is therefore larger in these regions. Assuming a uniform interstellar gas pressure, this flow energy will not be dissipated as quickly as in the ecliptic plane. This will effectively destroy the spherically symmetric nature of the shock, turning it into something that could be better described by an ellipsoid with its major axis through the solar poles and its two minor axes in the solar rotational plane (Suess, 1993). A cartoon of this is shown in Figure 2.6(a).

The localized high-velocity solar wind streams will cause localized ‘bulges’ in the shock face where it reaches the termination shock, pushing the shock further back. This will give the shock an uneven character (Suess, 1993). In addition to this, the solar wind is not constant with time, and, therefore, the whole structure probably oscillates back and forth.

The motion of the heliosphere through the interstellar medium creates a larger ram pressure due to the interstellar medium in the direction of motion and a smaller pressure in the opposite direction. This means that the flow energy of the solar wind will be dissipated faster in the direction of motion than in the opposite direction, and thus the shock face will be nearer to the sun in the direction of motion and further out on the other side. A spherical shock will thus be deformed into an ovoid, as shown in Figure 2.6(b).

All these factors contribute to the possible asymmetric nature of the termination shock structure. Such asymmetries will, however, not be included in our model, which will have a purely spherical
termination shock.

2.2.4 Structure Outside the Shock

For the flow velocity of an incompressible gas it follows from (2.2) that

\[ V \propto \frac{1}{r^2}. \]  

(2.20)

This means that a fully radial flow will be divergence free. This is a situation that cannot continue for large distances beyond the shock, since the influence of the interstellar medium will certainly become felt.

In the direction in which the sun is moving through interstellar space the opposing interstellar gas and plasma flow due to the motion of the sun will cause the solar wind to start to change its flow direction in the same way that a thin jet of water will bend and turn around in a gust of wind. The now subsonic (and more or less powerless) solar wind will therefore change direction such that it will flow around and towards the back of the heliosphere so that its flow direction ultimately ends up in the same direction as that of the background interstellar wind.

The point where the flow line in the direction of the heliospheric motion turns its direction about is called the stagnation point. According to Nerney and Suess (1993) this stagnation point is in all likelihood located at \( 2r_s \), i.e., twice the distance between the sun and termination shock.

On the opposite side of the heliosphere the solar wind velocity meets little resistance beyond the shock since the interstellar gas flow will be in the same direction. Therefore, the interstellar gas will not influence the solar wind in this region. In the intermediate regions, between these two extreme cases, the solar wind will also change direction in such a way that the final direction will be more or less the same as that of the interstellar wind, as shown in Figure 2.1.

2.3 The Interplanetary Magnetic Field

The solar wind is a highly ionized plasma with a large electrical conductivity (or small resistivity). In the case where an idealized plasma has no resistivity, a magnetic field cannot move relative to this plasma. It can, therefore, be said that the magnetic field is 'frozen' into the plasma. In the heliosphere this frozen-in condition leads to the so-called Parker spiral magnetic field.
2.3.1 The Parker Spiral Field

The formal derivation starts with Maxwell's equations and Ohm's law to give two partial differential equations containing the magnetic field, that can be solved with a solar wind flow, \( \mathbf{V} \), to obtain a vector expression for the IMF. More detailed derivations can be found in Webber and Davis (1967) and Jokipii and Kóta (1989).

From Ohm's law,
\[
\mathbf{J} = \sigma(\mathbf{E} + \mathbf{V} \times \mathbf{B}),
\]
it follows that for large conductivity, \( \sigma \to \infty \), the electric field must be \( \mathbf{E} = -\mathbf{V} \times \mathbf{B} \). Thus, Faraday's law,
\[
\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t},
\]
in the steady-state gives
\[
\nabla \times \mathbf{E} = \nabla \times \mathbf{V} \times \mathbf{B} = 0. \tag{2.21}
\]
If the heliosphere is axisymmetric (\( \partial / \partial \phi = 0 \)), the radial and \( \theta \)-components of (2.21) give
\[
V_r B_\theta - V_\theta B_r = \frac{f_1(r)}{\sin \theta} \tag{2.22}
\]
and
\[
V_r B_\theta - V_\theta B_r = \frac{f_2(\theta)}{r} \tag{2.23}
\]
These two conditions give
\[
V_r B_\theta - V_\theta B_r = \frac{C}{r \sin \theta}, \tag{2.24}
\]
with \( C \) a constant. Assume that the wind has no \( \theta \)-component. Then it follows from \( \nabla \cdot \mathbf{B} = 0 \) and (2.24) that
\[
\frac{1}{r^2} \frac{\partial}{\partial r} (r^2 B_r) = -\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} (r \sin \theta B_\theta) = -\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \frac{C}{V_r} \right) = 0,
\]
if \( V_r \) is independent of \( \theta \). Thus,
\[
B_r = B_r \Theta \left( \frac{r \Theta}{r} \right)^2, \tag{2.25}
\]
where the symbol \( \Theta \) denotes values on the surface of the sun.

Meanwhile, the \( \phi \)-component of (2.21) gives
\[
r(V_\phi B_r - V_r B_\phi) = r \Theta V_\phi \Theta B_r \Theta, \tag{2.26}
\]
assuming that \( V_r \Theta = 0 \).

Conservation of angular momentum density, \( L = r \times \rho \mathbf{V} \), in the solar wind gives
\[
r V_\phi = r \Theta V_\phi \Theta, \]
and \( V_\phi \Theta \) is the co-rotation velocity, \( \Omega r \Theta \sin \theta \), on the solar surface. Thus, (2.26) becomes
\[
B_\phi = -B_r \Theta \left( \frac{r \Theta}{r} \right)^2 \tan \psi \left[ 1 - \left( \frac{r \Theta}{r} \right)^2 \right], \tag{2.27}
\]
where
\[
\tan \psi = \frac{\Omega r \sin \theta}{V_r}.
\]
is the so-called spiral or garden hose angle of the field. The last term in (2.27) is insignificant beyond a few solar radii. Furthermore, it is convenient to normalize the expression in terms of the average magnitude of the field at Earth, which has a value \( B_e \approx 5-10 \) nT. Thus, the field expressions as used in this thesis are

\[
B = B_e \left( \frac{r_e}{r} \right)^2 \cos \psi_e [e_r - \tan \psi e],
\]

with magnitude

\[
B = B_e \left( \frac{r_e}{r} \right)^2 \cos \psi_e \cos \psi,
\]

where

\[
\tan \psi = \frac{\Omega r \sin \theta}{V},
\]

and where \( V = V_r \) is the radial solar wind velocity.

At this point it is also useful to note the following convenient system of units: The basic unit of distance is the Astronomical Unit (AU) with

\[
1 \text{ AU} = 1.5 \times 10^{11} \text{ m}.
\]

The basic unit of time is abbreviated as \( \mathcal{S} \), related to the angular velocity of the sun:

\[
1 \mathcal{S} = \frac{27.26 \text{ days}}{2\pi \text{ radians}} = 3.75 \times 10^5 \text{ s}.
\]

It then follows from (2.31) and (2.32) that the unit of velocity is \( 1 \text{ AU/1} \mathcal{S} = 400 \text{ km/s} \), which is the typical solar wind speed. In these units, \( V/\Omega = 1 \text{ AU} = r_e \) and the expression for the garden hose angle simplifies to

\[
\tan \psi = \frac{r}{r_e} \sin \theta.
\]

Thus \( \tan \psi = 1 \) and \( \cos \psi = 1/\sqrt{2} \) if \( V = 400 \text{ km/s} \) and \( \Omega = 2\pi/27.26 \text{ days}^{-1} \).

According to observations, the average magnetic field strength at the position of the earth is \( \sim 5 \text{ nT} \) (nanotesla) and the time-averaged value for the garden hose angle, \( \psi \), 45°. According to Thomas and Smith (1980) the field directions observed by Pioneer 10 and 11 between 1 and 8.5 AU conform on average to within 1.1° with the Parker spiral field. It is therefore a realistic approximation.

There also exist temporal variations in magnetic field strength (Winterhalter et al., 1990) in conjunction with the geometric changes described above. The magnetic field strength at the earth changed from \( \sim 5.5 \text{ nT} \) in 1976 to \( \sim 9.5 \text{ nT} \) in 1982 and back again to \( \sim 5.5 \text{ nT} \) in 1985. From this it seems that there exists a correlation between solar activity and magnitude of IMF. It is useful to note that, with the above system of units,

\[
5.55 \text{ nT} = 5.55 \times 10^{-9} \text{ V s m}^{-2} = \frac{1}{3} \text{ GV} \mathcal{S} (\text{AU})^{-2},
\]

where \( 1 \text{ GV} = 10^9 \text{ V} \).

The transport of cosmic rays in the heliosphere is strongly determined by the geometry of the IMF. In particular, the IMF above the solar poles is quite uncertain. Several models have been proposed to affect field modifications near the poles. In the next three subsections we discuss such modifications.
2.3.2 The Jokipii-Kóta Modification

Deviations from a pure Parker spiral, especially away from the ecliptic plane, may occur. In particular, the radial field lines at the poles are in a state of unstable equilibrium. Therefore, the smallest perturbation can cause the collapsing of the field line. The surface of the sun — at the ‘feet’ of the field lines — is not a smooth surface, but a granular, turbulent surface that keeps changing with time. This causes the ‘feet’ of the polar field lines to move randomly, creating transverse components in the field. The net effect of this is a highly irregular and compressed field line. In other words, the magnitude of the mean magnetic field at the poles is greater than in the case of the smooth magnetic field of a pure Parker spiral.

Jokipii and Kóta (1989) suggested that the Parker spiral field (2.28) may be generalized by the introduction of a parameter, \( \delta(\theta, \phi) \), which amplifies the field strength at large radial distances with \( \delta(\theta, \phi)(r/r_\odot) \). With this modification, the Parker spiral field, (2.28), becomes

\[
B = B_e \left( \frac{r_e}{r} \right)^2 \cos \psi_e \left[ e_r + \delta(\theta, \phi) \left( \frac{r}{r_\odot} \right) e_\theta - \tan \psi e_\phi \right].
\]

(2.35)

The magnitude of this modified field then becomes

\[
B = B_e \left( \frac{r_e}{r} \right)^2 \cos \psi_e \sqrt{\sec^2 \psi + \left( \frac{\delta(\theta, \phi) r}{r_\odot} \right)^2}.
\]

(2.36)

The effect of this modification is to increase the field in the polar regions in such a way that for large \( r \) it drops off as \( 1/r \) instead of \( 1/r^2 \). In the ecliptic regions of the outer heliosphere, where \( \sec \psi \approx 1 \), this modification has little effect on the field. It should be noted, however, that this leaves the field divergence free only if \( \delta \propto 1/\sin \theta \).

2.3.3 The Moraal Modification

Moraal (1990) suggested that the Parker spiral field may, alternatively, be modified by the introduction of an arbitrary function, \( \Theta(\theta) \), to incorporate the same physical effects that the Jokipii-Kóta modification compensates for:

\[
B = B_e \left( \frac{r_e}{r} \right)^2 \cos \psi_e \Theta(\theta) \left[ e_r - \tan \psi e_\phi \right].
\]

(2.37)

This function is chosen to have a value of one in the ecliptic (\( \theta = 90^\circ \)) and an arbitrary value such that \( \Theta > 1 \) at the poles. With this function, the magnetic field strength may be increased towards the poles, leaving the Parker spiral field unaffected in the ecliptic plane. The magnitude of this field then becomes

\[
B = B_e \left( \frac{r_e}{r} \right)^2 \Theta(\theta) \frac{\cos \psi_e}{\cos \phi}.
\]

(2.38)

This modification does not change the geometry of the Parker spiral field, but only the magnitude of the magnetic field strength in the polar regions.

In our numerical programs we employ both the Jokipii-Kóta and the Moraal modifications,

\[
B = B_e \left( \frac{r_e}{r} \right)^2 \cos \psi_e \left[ \Theta(\theta) e_r + \delta(\theta, \phi) \left( \frac{r}{r_\odot} \right) e_\theta - \Theta(\theta) \tan \psi e_\phi \right]
\]

(2.39)

with magnitude

\[
B = B_e \left( \frac{r_e}{r} \right)^2 \cos \psi_e \sqrt{\left( \frac{\Theta}{\cos \psi} \right)^2 + \left( \frac{\delta(\theta, \phi) r}{r_\odot} \right)^2}
\]

(2.40)

where the modification parameters \( \delta \) and \( \Theta \) can be chosen independently.
2.3.4 The Smith and Bieber Modification

Although granulation and supergranulation on the surface of the sun cause random changes in the magnetic field direction through the introduction of small random azimuthal/tangential fluctuations in the field lines, the mean field direction averages to the Parker spiral field geometry:

\[
\tan \psi = \frac{\Omega(r - b) \sin \theta}{V(r)}
\]

with \(b\) the inner boundary.

Smith and Bieber (1991) put forward a field modification based on magnetic field data that change the geometry of the magnetic field and predominantly affects the field strength over the poles. They proposed that the magnetic field is not fully radial below the Alfvèn radius (below which the magnetic field and the solar corona co-rotate in phase) taken to be in the order of 10–30 \(r_\odot\). The modification is parameterized by the ratio of the tangential (azimuthal) component of the magnetic field to that of the radial component as is found at the Alfvèn radius:

\[
\tan \psi = \frac{\Omega(r - b) \sin \theta}{V(r)} - \frac{B_T(b) V(b) r}{B_R(b) V(r) b}
\]  (2.41)

For the inner boundary, \(b\), an approximation of the Alfvèn radius was taken (\(b = 20r_\odot\)) and then \(B_T(b)/B_R(b) \approx -0.02\), according to an estimate by Smith and Bieber.

2.3.5 The Wavy Neutral Sheet

It has become clearly established that the IMF is of a bipolar nature, with the field in the Northern and Southern Hemisphere pointing in opposite directions. Thus, it follows that a thin ‘surface’, separating the two polarity sectors, must exist (Hoeksema et al., 1983). This ‘surface’ is called the neutral or current sheet. If the line separating the two polarities on the sun coincides with the rotational equator, the neutral sheet in the IMF will be a flat plane in the ecliptic, as shown in Figure 2.7(a). If, however, the ‘magnetic equator’ is skewed with an angle \(\alpha\) relative to the rotational equator, as in Figure 2.7(b), solar rotation will cause the neutral sheet to have the form

\[
\theta_{ns} = \frac{\pi}{2} + \sin^{-1} \left[ \sin \alpha \sin \left( \phi - \phi_\odot + \frac{\Omega r}{V} \right) \right],
\]  (2.42)

where \(\phi_\odot = r_\odot \Omega/V\) is an arbitrary phase constant. This surface is shown in Figure 2.8 for a tilt angle \(\alpha = 10^\circ\). An observer in the ecliptic plane will, therefore, see a polarity change every 1/2 solar rotation.
In view of this bipolar nature of the field with a neutral sheet, the Parker spiral (2.28) must formally be generalized to

\[ B = B_\text{e} \left( \frac{r_s}{r} \right)^2 \cos \psi_e [e_r - \tan \psi_e [1 - 2H(\theta - \theta_{\text{ns}})]], \]

(2.43)

with \( H \) the Heaviside step function.

2.3.6 Solar Cycle Variations

The tilt angle, \( \alpha \), of the neutral sheet changes dramatically with the 11-year solar cycle (Hoeksema, 1989). During solar minimum, the magnetic equator nearly coincides with the rotational equator and the neutral sheet exhibits a moderate amount of waviness. Towards solar maximum the tilt angle increases, causing an enhanced waviness in the neutral sheet, until the neutral sheet dominates the IMF near solar maximum. Approximately at solar maximum, the tilt angle increases so much that the magnetic equator on the photosphere rotates through the vertical (Webb, et al., 1984) over a period of 1 to 1.5 years, effectively changing the polarity of the two hemispheres around. In this way, the polarity of the two hemispheres is changed about every 11 years, creating a 22-year magnetic solar cycle.

Therefore, the normalization constant \( B_\text{e} \) in (2.43) must be taken as a signed quantity, changing every 11 years. In the periods \( \sim 1970 \) to \( \sim 1980 \) and from \( \sim 1990 \) onwards it was positive, leading to an outward pointing field in the Northern hemisphere and inward in the Southern hemisphere. From \( \sim 1960 \) to \( \sim 1970 \) and from \( \sim 1980 \) to \( \sim 1990 \) the directions were reversed.

2.3.7 The IMF at and beyond the Termination Shock

At and beyond the SWTS, the solar wind speed drops off according to (2.20):

\[ V' = \frac{V - \left( \frac{r_s}{r} \right)^2}{s}, \]

(2.44)
with \( r_s \) the shock radius and

\[
V^- = \lim_{r \to r_s} V.
\]

Consequently, beyond the SWTS the garden hose angle of the Parker spiral field becomes

\[
\tan \psi' = \frac{\Omega r \sin \theta}{V'} = \frac{\Omega r \sin \theta}{V^-} s \left( \frac{r}{r_s} \right)^2 = s \left( \frac{r}{r_s} \right)^2 \tan \psi
\]

i.e., the field becomes even more tightly wound. Thus, it follows from (2.29) that the field strength beyond the shock is

\[
B' = B_e \left( \frac{r_s}{r} \right)^2 \cos \psi \sqrt{1 + s^2 \left( \frac{r}{r_s} \right)^4 \tan^2 \psi}.
\] (2.45)

For large values of \( \tan \psi \), i.e., away from the polar regions, this field increases \( \propto r \), which cannot be realistic. However, at present very little is known of the way this field must merge into the interstellar field, and we will use this expression.

On the termination shock, the ratio of the upstream to downstream field is simply

\[
\frac{B^+}{B^-} = \sqrt{\frac{1 + s^2 \tan^2 \psi^-}{1 + \tan^2 \psi^-}}.
\] (2.46)

At the poles, where \( \tan \psi = 0 \), the radial field is continuous as required by Maxwell's equations. In the equatorial regions, where \( \tan \psi \gg 1 \), the field strength jumps with the compression ratio, \( s \).
Chapter 3
Cosmic Rays in the Heliosphere

3.1 Introduction

Probably the most important event in the study of the transport of cosmic rays in the heliosphere came in 1965 when E.N. Parker first wrote down the so-called transport equation (Parker, 1965), treating the process as a modified diffusion problem. This is a valid approach since the IMF does not have a smooth structure. Superimposed on the background field, $B_0$, there are irregularities, $\delta B$, such that the total field is $B = B_0 + \delta B$.

These irregularities are effective scattering centers for particles with gyroradii, $r_g$, of the same order of magnitude as the spatial scale length, $L$, of the fluctuations.

The Parker transport equation (TPE) contains terms describing the diffusion of cosmic rays by the scattering centers in the IMF, convection due to the radial outward propagation of the scattering centers, frozen into the solar wind, as well as energy changes.

The equation has subsequently been rederived, and its detailed properties studied by, e.g., Gleeson and Axford (1968a,b,c), Jokipii and Parker (1970), and Webb (1975).

In this chapter the Parker TPE is written down, together with the cosmic-ray streaming vector. Its terms are written out in detail in heliocentric spherical coordinates, as it will be used in the subsequent numerical calculations.

The central theme of this thesis is to study the acceleration of particles by the SWTS through the process of first-order Fermi acceleration. Therefore, a simplified, analytical solution of the TPE is given to demonstrate the basic properties of this shock acceleration process.

The last four sections of this chapter then discuss the properties of the drift and diffusion coefficients in the equation.

3.2 The Parker Transport Equation

The differential cosmic-ray number density, $U_p$, is defined as the number of particles per unit volume in the momentum interval $(p, p + dp)$. The continuity equation for this number density is

$$\frac{\partial U_p}{\partial t} + \nabla \cdot S_p + \frac{\partial}{\partial p}(<p > U_p) = Q(r, p, t),$$

(3.1)

with $S_p$ the differential particle current density, or streaming, and $Q$ a source of particles. The term

$$\frac{\partial}{\partial p}(<p > U_p)$$
can be taken as the streaming in momentum space due to the mean rate of change in momentum, \(< \dot{\mathbf{p}} >\) (notation: \(\dot{\mathbf{p}} = \partial \mathbf{p} / \partial t\)). According to Webb (1975), its value is
\[
< \dot{\mathbf{p}} > = \frac{\mathbf{V}}{3} \cdot \mathbf{g},
\]  
with
\[
\mathbf{g} = \frac{1}{U_p} \nabla U_p,
\]
the density gradient. For diffusive flow, the streaming is given by (a generalization of) the well-known Fick's law:
\[
S_p = -\mathbf{K} \cdot \nabla U_p,
\]
with \(\mathbf{K}\) the diffusion tensor. The scattering centers in the solar wind, however, move with velocity \(\mathbf{V}\) relative to a stationary observer. Thus, the stationary observer records a streaming
\[
S_p = CVU_p - \mathbf{K} \cdot \nabla U_p,
\]
where
\[
C = 1 - \frac{1}{3U_p} \frac{\partial}{\partial p}(pU_p),
\]
is the so-called Compton-Getting factor (e.g., Gleeson and Axford, 1968b).
Inserting (3.2) and (3.3) into (3.1) gives
\[
\frac{\partial U_p}{\partial t} + \nabla \cdot (CVU_p - \mathbf{K} \cdot \nabla U_p) + \frac{\mathbf{V}}{3} \cdot \nabla \frac{\partial}{\partial p}(pU_p) = Q
\]
or, combining the convection and momentum change terms,
\[
\frac{\partial U_p}{\partial t} + \nabla \cdot (\mathbf{V}U_p - \mathbf{K} \cdot \nabla U_p) - \frac{1}{3} (\nabla \cdot \mathbf{V}) \frac{\partial}{\partial p}(pU_p) = Q.
\]
This is the so-called Parker Equation or Transport Equation (TPE), first written down by Parker (1965).
To analyze the physical content of this equation, rearrange it by splitting the diffusion tensor into its symmetrical and antisymmetrical parts:
\[
\mathbf{K} = \mathbf{K}^{(S)} + \mathbf{K}^{(A)}
\]
\[
= \begin{bmatrix}
\kappa_{||} & 0 & 0 \\
0 & \kappa_0 & 0 \\
0 & 0 & \kappa_{\perp}
\end{bmatrix}
+ \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & \kappa_T \\
0 & -\kappa_T & 0
\end{bmatrix}.
\]
Thus, the third term in (3.6) becomes
\[
\nabla \cdot (\mathbf{K} \cdot \nabla U_p) = \nabla \cdot (\mathbf{K}^{(S)} \cdot \nabla U_p) + \nabla \cdot (\mathbf{K}^{(A)} \cdot \nabla U_p),
\]
i.e., the divergence of the diffusive and the drift flux. The symmetric part contains elements for diffusion parallel (\(\kappa_{||}\)) and perpendicular (\(\kappa_{\perp}\)) to the field, while the form of the antisymmetric part follows from standard drift theory as follows:
Standard drift analysis (e.g., Chen, 1985) of charged particles in magnetic fields gives the drift streaming
\[
S_p(\text{drift}) = \frac{peu}{3qB^2} B \times \nabla U_p = \frac{\beta P}{3B^2} B \times \nabla U_p,
\]
\[
\frac{\mathbf{V}}{3} \cdot \nabla U_p = \frac{\beta P}{3B^2} B \times \nabla U_p,
\]
where \( q \) is particle charge. Its divergence is

\[
\nabla \cdot S_p(\text{drift}) = \left( \nabla \times \frac{\beta p}{3B}e_B \right) \cdot \nabla U_p = v_d \cdot \nabla U_p,
\]

where the drift velocity, \( v_d \), is given by

\[
v_d = \nabla \times \kappa_T e_B,
\]

and

\[
\kappa_T = \frac{\beta p}{3B}.
\]

This drift velocity is the average value of the standard gradient and curvature drift for all particles with momentum between \( p \) and \( p + dp \), as was shown by Jokipii et al. (1977) and elaborated on by Burger et al. (1985).

Now,

\[
v_d \cdot \nabla U_p = (\nabla \times \kappa_T e_B) \cdot \nabla U_p = \nabla \cdot (\kappa_T e_B \times \nabla U_p).
\]

Thus,

\[
\nabla \cdot (\mathcal{K}^{(A)} \cdot \nabla U_p) = v_d \cdot \nabla U_p
\]

if

\[
\mathcal{K}^{(A)} \cdot \nabla U_p = \kappa_T e_B \times \nabla U_p.
\]

This is true if \( \mathcal{K}^{(A)} \) has the form as given in (3.7). With this definition of the drift velocity, the TPE (3.6) can alternatively be written as

\[
\frac{\partial U_p}{\partial t} - \nabla \cdot (\mathcal{K}^{(s)} \cdot \nabla U_p) + (V + v_d) \cdot \nabla U_p + (\nabla \cdot V) U_p - \frac{1}{3}(\nabla \cdot V) \frac{\partial}{\partial p}(pU_p) = Q.
\]

(3.10)

It is often useful to use the omni-directional distribution function of particles, which is related to the differential number density by

\[
f(r, p, t) = \frac{U_p(r, p, t)}{4\pi p^2}.
\]

With this, the Compton-Getting factor (3.4) becomes

\[
C = -\frac{1}{3} \frac{\partial \ln f}{\partial \ln p},
\]

(3.11)

and the TPE (3.6) becomes

\[
\frac{\partial f}{\partial t} + \nabla \cdot (Vf - \mathcal{K} \cdot \nabla f) - \frac{1}{3p^2}(\nabla \cdot V) \frac{\partial}{\partial p}(p^3f) = Q_f
\]

(3.12)

where

\[
Q_f = \frac{Q}{4\pi p^2}.
\]

(3.13)

In the same way, (3.10) can be written as

\[
\frac{\partial f}{\partial t} - \nabla \cdot (\mathcal{K}^{(s)} \cdot \nabla f) + (V + v_d) \cdot \nabla f - \frac{1}{3}(\nabla \cdot V) \frac{\partial f}{\partial \ln p} = Q_f.
\]

(3.14)

In this notation, the streaming density (3.3) becomes

\[
S_p = -4\pi p^2 \left( \frac{V \cdot \partial f}{3 \partial \ln p} + \mathcal{K} \cdot \nabla f \right)
\]

\[
= -4\pi p^2 \left( \frac{V \cdot \partial f}{3 \partial \ln p} + \mathcal{K}^{(s)} \cdot \nabla f + \kappa_T e_B \times \nabla f \right).
\]

(3.15)
Curvature and gradient drifts are included in the TPE, but the $E \times B$-drift due to the $E = -V \times B$ electric field seems to be conspicuously absent. This is, fortunately, not true. According to Forman and Gleeson (1975), this effect is indeed present in the TPE: The component of non-diffusive streaming parallel to the magnetic field is $CV_{\|}$, due to the convection velocity, $V_{\|}$, in that direction. The $E \times B$-drift velocity (for a stationary observer) is given by

\[
E \times \frac{B}{B^2} = -(V \times B) \times \frac{B}{B^2} = \frac{1}{B^2} [B^2V - (V \cdot B)B] = \frac{1}{B^2} [B^2V - B^2(V \cdot e_B)e_B] = V - V_{\|} = V_{\perp},
\]

where $(V \cdot e_B)e_B$ gives the component of $V$ in the direction of $B$. If every particle experiences a drift velocity, $V_{\perp}$, then the streaming density of the cosmic ray gas is $CV_{\perp}$. The sum $CV_{\|} + CV_{\perp}$ is precisely the convection streaming density, $CVU_p$ in (3.5).

For completeness, a term describing so-called second-order Fermi acceleration may be added to the TPE. This describes diffusion in momentum space due to randomly moving magnetic irregularities, causing a net acceleration because particles have a slightly higher chance to collide with approaching scattering centers than with receding ones. The process is called second-order, because it can be shown that the average momentum gain per scattering is proportional to $(\text{plasma speed/particle speed})^2$. With second-order Fermi acceleration included, the TPE thus becomes

\[
\frac{\partial f}{\partial t} = \mathbf{\nabla} \cdot (K^{(S)} \mathbf{\nabla} f) \quad \text{(diffusion)}
- v_d \cdot \mathbf{\nabla} f \quad \text{(guiding-center drift)}
- V \cdot \mathbf{\nabla} f \quad \text{(convection)}
+ \frac{1}{3} (\nabla \cdot V) \frac{\partial f}{\partial \ln p} \quad \text{(energy or momentum change)}
+ \frac{1}{p^2} \frac{\partial}{\partial p} \left( p^2 D_{pp} \frac{\partial f}{\partial p} \right) \quad \text{(diffusion in momentum space)}
+ Q_f(r, p, t) \quad \text{(source/sink)}
\]

The $D_{pp}$ factor in the momentum diffusion term is given by

\[
D_{pp} = \frac{p^2 V_A^2}{4 \zeta K_{||}},
\]

Here $V_A$ is the Alfvén velocity in the solar wind plasma, given by

\[
V_A = \frac{B}{\sqrt{\rho}}
\]

with $\zeta \geq 1$ (e.g., Gombosi et al., 1989).

This term is potentially important in cosmic ray transport, especially in regions with strong magnetic fields and small plasma densities (such as near the sun or in the downstream region of a MHD shock). Even so, this term is quite small in comparison to the others and, therefore, it is usually ignored in most heliospheric cosmic-ray transport models.

### 3.3 First-Order Fermi (Shock) Acceleration

Since $D_{pp}$ is so small, second-order Fermi acceleration cannot account for the effective acceleration of cosmic rays in the heliosphere. However, even if first-order Fermi acceleration does
not appear explicitly in the TPE, it does accelerate particles with great effectiveness if a MHD shock is present in a diffusive-convective system. This basic effect was discovered in 1977 (Axford et al., 1977; Bell, 1978a,b; Blandford and Ostriker, 1978; Krymski, 1977), and the spectrum resulting from this acceleration is derived below.

The fundamental starting point in such an analysis concerns the continuity properties of the density and streaming across the shock. Since particles have mobility across the shock, the first of these is that the density, intensity, or distribution function must be continuous, i.e.,

\[ U_p^- = U_p^+ \]

or

\[ f^- = f^+ \]  \hspace{1cm} (3.18)

with the notation that, if \( r_s \) is the position of the shock,

\[ f^- = \lim_{r \to r_s^-} f(r) \]

and

\[ f^+ = \lim_{r \to r_s^+} f(r). \]

Similarly, the condition on the streaming must be such that

\[ \nabla \cdot \mathbf{S_p} = Q, \]  \hspace{1cm} (3.19)

i.e., the flux that diverges from the shock must have its origin in a source on the shock. In a one-dimensional problem, or where the flux is perpendicular to the shock face, this second condition simply reduces to

\[ S^+ - S^- = \lim_{\epsilon \to 0} \int_{r_s - \epsilon}^{r_s + \epsilon} Q \, dr \]  \hspace{1cm} (3.20)

Consider a simple, steady, plane MHD shock at \( x = 0 \) with a structure as in Figure (2.5). Assume that sufficient scattering centers exist on both sides of the shock to keep the distribution function, \( f(r, p, t) \), isotropic to first order. Under sufficient simplifying assumptions, the TPE can be solved across this shock:

In the one-dimensional steady state (\( \partial f/\partial t = 0 \)), ignoring drifts, and for isotropic scattering, the TPE (3.12) reduces to

\[ \frac{\partial}{\partial x} \left[ V f - \kappa \frac{\partial f}{\partial x} \right] - \frac{1}{3p^2} \left( \frac{\partial V}{\partial x} \right) \frac{\partial}{\partial p} (p^3 f) = Q_J(x, p) \]

and the streaming density (3.15) is

\[ S_p = -4\pi p^2 \left[ \frac{V p}{3} \frac{\partial f}{\partial p} + \kappa \frac{\partial f}{\partial x} \right]. \]

If the upstream flow velocity and diffusion coefficient are independent of \( x \), the TPE becomes a simple linear differential equation,

\[ \left( \frac{\partial^2}{\partial x^2} - \frac{V}{\kappa} \frac{\partial}{\partial x} \right) f = \frac{Q_J(x, p)}{\kappa}. \]  \hspace{1cm} (3.21)

This equation has a simple solution, which depends on how the source is treated. Assume that \( f \) is known far upstream of the shock, denoted by \( f(-\infty, p) \), and that on the shock it is \( f(0, p) \).
Then, if the source is a delta function on the shock, i.e., \( Q_f(x,p) = Q_\ast(p) \delta(x) \), the solution in the downstream medium is
\[
f(x,p) = f(-\infty,p) + [f(0,p) - f(-\infty,p)] e^{Vz/\kappa}. \tag{3.22}
\]
If, for some reason, the source is distributed throughout the upstream medium, the solution is
\[
f(x,p) = f(-\infty,p) + [f(0,p) - f(-\infty,p)] e^{Vz/\kappa} + \frac{1}{V} \left[ (1 - e^{Vz/\kappa}) \int_{-\infty}^{x} Q_f \, dz - e^{Vz/\kappa} \int_{0}^{x} Q_f \left( 1 - e^{-Vz/\kappa} \right) \, dz \right]. \tag{3.23}
\]
The shape of the spectrum on the shock, \( f(0,p) \), is determined by the two continuity conditions (3.18) and (3.20).

For a delta function source on the shock, these conditions imply
\[
\frac{V_1 p \left( \frac{\partial f(0,p)}{\partial p} \right)}{\kappa_1} + \frac{V_2 p \left( \frac{\partial f(0,p)}{\partial x} \right)}{\kappa_2} = \frac{V_2 p \left( \frac{\partial f(0,p)}{\partial p} \right)}{\kappa_2} + \frac{\partial f(0,p)}{\partial x} + Q_\ast, \tag{3.24}
\]
while for the distributed source at \( x < 0 \) it is
\[
\frac{V_1 p \left( \frac{\partial f(0,p)}{\partial p} \right)}{\kappa_1} + \frac{V_2 p \left( \frac{\partial f(0,p)}{\partial x} \right)}{\kappa_2} = \frac{V_2 p \left( \frac{\partial f(0,p)}{\partial p} \right)}{\kappa_2} + \frac{\partial f(0,p)}{\partial x}.
\]
To solve the problem analytically, one assumes a no-diffusion situation beyond the shock, or \( \kappa_2 = 0 \). Under these circumstances the matching condition becomes
\[
\frac{V_1 - V_2}{p} \frac{\partial f(0,p)}{\partial p} = -\frac{\partial f(0,p)}{\partial x}. \tag{3.25}
\]
From equation (3.23) we can now calculate \( \frac{\partial f(x,p)}{\partial x} \):
\[
\frac{\partial f(x,p)}{\partial x} = \frac{V}{\kappa} [f(0,p) - f(-\infty,p)] e^{Vz/\kappa} - \frac{1}{\kappa} e^{Vz/\kappa} \left[ \int_{-\infty}^{x} Q_f \, dz - \int_{0}^{x} Q_f e^{-Vz/\kappa} \, dz \right].
\]
To obtain an expression for \( \left( \frac{\partial f(0,p)}{\partial x} \right)^- \) take \( \kappa = \kappa_1 \) at \( x < 0 \):
\[
\left( \frac{\partial f(0,p)}{\partial x} \right)^- = \frac{V}{\kappa_1} [f(0,p) - f(-\infty,p)] - \frac{1}{\kappa_1} \int_{0}^{-\infty} Q_f \, dz.
\]
From (3.25) we, therefore, have
\[
\frac{V_1 - V_2}{p} \frac{\partial f(0,p)}{\partial p} = -V_1 [f(0,p) - f(-\infty,p)] + \int_{0}^{-\infty} Q_f \, dz.
\]
If \( s = V_1/V_2 \) and \( q = 3s/(s - 1) \) we now have
\[
\frac{\partial f(0,p)}{\partial x} + \frac{q}{p} f(0,p) = \frac{q}{p} \left[ f(-\infty,p) + \frac{1}{V_1} \int_{0}^{-\infty} Q_f \, dz \right],
\]
with the solution
\[
f(0,p) = q p^{-q} \int_{0}^{p} \left[ f(-\infty,p^\prime) + \frac{1}{V_1} \int_{0}^{-\infty} Q_f(x,p^\prime) \, dx \right] p^\prime q - 1 \, dp^\prime. \tag{3.26}
\]
Similarly, if one takes the simpler case of a source on the shock, (3.22) and (3.24) give
\[
f(0,p) = q p^{-q} \int_{0}^{p} \left[ f(-\infty,p^\prime) + \frac{Q_\ast(p^\prime)}{V_1} \right] p^\prime q - 1 \, dp^\prime. \tag{3.27}
\]
Since equation (3.21) is linear, (3.26) and (3.27) may be added, or the problem might have been done in unison. The kernel of the combined expression is

$$f(-\infty, p') + \frac{1}{V_1} \left[ \int_0^{-\infty} Q_f(x, p') \, dx + Q_0 \right].$$

This shows that a pre-existing spectrum far ahead of the shock, a source in the upstream medium (produced for instance by travelling interplanetary shocks) or a source on the shock itself, will produce the same accelerated spectrum.

The spectrum in the downstream medium is homogenous and equal to the spectrum on the shock, i.e. \( f(x, p) = f(0, p), x > 0 \). The spectrum in front of the shock falls off exponentially to the value of the spectrum at \( x = -\infty \), i.e.

$$f(x, p) = \begin{cases} f(0, p) & \text{if } x \geq 0 \\ f(-\infty, p) + [f(0, p) - f(-\infty, p)] e^{\frac{V_1 x}{\kappa}} & \text{if } x < 0 \end{cases}$$

Consider as an example a mono-energetic source function at the shock:

$$\frac{Q_0(p)}{V_1} = \frac{N_0}{4\pi p^2} \delta(p - p_0) \text{ and } f(-\infty, p) = 0.$$

If one keeps in mind that

$$\int_0^{\infty} 4\pi p^2 \frac{Q_0(p)}{V_1} \, dp = N_0,$$

the accelerated spectrum (3.27) becomes

$$f(0, p) = \frac{N_0}{4\pi p_0^2} q \left( \frac{p}{p_0} \right)^{-q},$$

i.e., a power law with the spectral index

$$q = \frac{3s}{s-1}.$$  (3.30)

Diffusive shock acceleration due to an infinite plane shock always gives rise to power law spectra with the spectral index given by equation (3.30), that depends only on the compression ratio of the shock. For a strong MHD shock in a highly ionized plasma the compression ratio, \( s \), is 4 and therefore \( q = 4 \).

In practice, shocks are seldom plane or stationary. The power law (3.29) can only be achieved up to such a value of momentum as there is time for the particles to reach this momentum. From, e.g., Axford (1981), Drury (1983), and Lagage and Césparsky (1983) it follows that the acceleration time needed for the establishment of the steady-state solution from momentum \( p_0 \) to \( p \) is given by

$$\tau_a = \frac{3}{V_1 - V_2} \int_{p_0}^{p} \left( \frac{\kappa_1}{V_1} + \frac{\kappa_2}{V_2} \right) \frac{dp'}{p'}. $$

(3.31)

Above that momentum, the spectrum cuts off sharply. Similarly, in a curved shock, there is a curvature cutoff in the spectrum. This occurs at the point where the diffusive length scale \( \kappa/V \), becomes larger than the shock radius, \( r_s \), or

$$\frac{\kappa_1}{V_1} + \frac{\kappa_2}{V_2} > r_s.$$  (3.32)
3.4 Pick-Up Ions in the Solar Wind and the Acceleration of ACR

The primary aim of this thesis is to develop a numerical solution of the cosmic-ray transport equation, alongside the solution of Jokipii (1986) to study the acceleration and modulation of the anomalous cosmic-ray component in the heliosphere.

The discovery of the so-called anomalous component of cosmic rays (ACR) by Garciá-Munoz et al. (1973a, 1973b, 1973c) provided a powerful new tool with which the heliosphere can be probed. Soon thereafter, in addition to the Helium discovered by Garciá-Munoz, anomalous Oxygen (Hovestadt et al., 1973), Nitrogen (McDonald et al., 1974), Neon (von Rosenvinge and McDonald, 1975), was observed. Fisk et al. (1974) recognized that these elements all have high first ionization potentials and, therefore, they proposed that these elements enter the heliosphere as interstellar neutrals due to the movement of the heliosphere in its trajectory through interstellar space. These elements then penetrate deeply into the heliosphere before they become singly ionized by charge exchange with the solar wind ions, electron collisions, or photo-ionization.

These singly-ionized atoms are then picked up by the solar wind and convected outwards towards the outer heliosphere, where they undergo acceleration. Some of these accelerated cosmic rays may then diffuse into the heliosphere, where they are modulated by the same processes as the galactic component, to form the anomalous component of cosmic rays.

Möbius et al. (1985) obtained the first conclusive evidence of the solar wind picking up singly-ionized interstellar Helium (He$^+$), using a time-of-flight spectrometer. According to Möbius (1986) the kinetic energies of these so-called pick-up ions vary from basically zero to approximately four times the flow energy of the solar wind.

The specific mechanism through which these ions are accelerated, from about four times the solar wind flow energy to those at which the ACR are observed, is not certain. The leading and trailing shocks of CIRs and MIRs do indeed accelerate these pick-up ions to higher energies, as was dramatically demonstrated by Gloeckler et al. (1993). These travelling interplanetary phenomena may be regarded as examples of the source function, $Q$, in the upstream region, as described in the previous section. This cannot account for most of the acceleration, however, since both the curvature cutoff (3.32) and the acceleration time scale (3.31) limit the acceleration efficiency of these phenomena. However, Pesses et al. (1981) proposed that the solar wind termination shock, as described in Section 2.2.3, may accelerate these pick-up ions to sufficiently high energies. This remains the most plausible source of ACR, even though Fisk (1986) reviewed other acceleration mechanisms.

Thus, in the remainder of this chapter, we will develop the theoretical basis for a numerical shock acceleration model (to be developed in Chapter 4) with which this hypothesis may be explored.

3.5 The TPE in a Parker Spiral Field

In Chapter 4 the TPE will be solved in heliospheric polar coordinates ($r, \theta, \phi$). In these coordinates (3.12) is

$$\frac{\partial f}{\partial t} = \kappa_{rr} \frac{\partial^2 f}{\partial r^2} + \kappa_{\theta \theta} \frac{\partial^2 f}{\partial \theta^2} + \frac{\kappa_{\phi \phi}}{r^2 \sin^2 \theta} \frac{\partial^2 f}{\partial \phi^2}$$
\[ + \frac{1}{r} (\kappa_{r\theta} + \kappa_{\theta r}) \frac{\partial^2 f}{\partial \theta^2} + \frac{1}{r \sin \theta} (\kappa_{r\phi} + \kappa_{\phi r}) \frac{\partial^2 f}{\partial \phi^2} + \frac{1}{r^2} \frac{\partial^2 f}{\partial \phi \partial \theta} \]
\[ + \left[ \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 \kappa_{rr}) + \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta \kappa_{r\theta}) + \frac{1}{r \sin \theta} \frac{\partial \kappa_{r\theta}}{\partial \phi} - V_r \right] \frac{\partial f}{\partial r} \]
\[ + \left[ \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 \kappa_{\theta\theta}) + \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta \kappa_{\theta\theta}) + \frac{1}{r \sin \theta} \frac{\partial \kappa_{\theta\theta}}{\partial \phi} - \frac{V_\theta}{r} \right] \frac{\partial f}{\partial \theta} \]
\[ + \left[ \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 \kappa_{r\phi}) + \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta \kappa_{r\phi}) + \frac{1}{r \sin \theta} \frac{\partial \kappa_{r\phi}}{\partial \phi} - \frac{(\partial^2 \phi)}{r} \right] \frac{\partial f}{\partial \phi} \]
\[ + \frac{1}{3} \left[ \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 V_r \right) + \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta V_\theta) + \left( \frac{\partial^2 V_\phi}{\partial \phi^2} \right) \right] + Q_f (r, \theta, \phi, p, t). \]

(3.33)

The streaming expression (3.15) becomes,
\[ \frac{S_p (r, \theta, \phi)}{4\pi p^2} = - \frac{V}{3 \frac{\partial f}{\partial \ln p}} - \mathbf{K} \cdot \nabla f \]
\[ = \left[ \begin{array}{ccc} V_r \frac{\partial f}{\partial r} - \kappa_{rr} \frac{\partial f}{\partial r} - \kappa_{r\theta} \frac{\partial f}{\partial \theta} - \kappa_{r\phi} \frac{\partial f}{\partial \phi} \\ V_\theta \frac{\partial f}{\partial \theta} - \kappa_{\theta r} \frac{\partial f}{\partial r} - \kappa_{\theta\theta} \frac{\partial f}{\partial \theta} - \kappa_{\theta\phi} \frac{\partial f}{\partial \phi} \\ V_\phi \frac{\partial f}{\partial \phi} - \kappa_{r\phi} \frac{\partial f}{\partial r} - \kappa_{\theta\phi} \frac{\partial f}{\partial \theta} - \kappa_{\phi\phi} \frac{\partial f}{\partial \phi} \end{array} \right], \]
(3.34)

while the drift velocity components of (3.8) are
\[ \mathbf{v}_d = \frac{\beta P}{3r^2 \sin \theta} \left[ \begin{array}{c} e_r \cos \theta e_\phi \\ e_\theta \\ e_\phi \end{array} \right] = \left[ \begin{array}{c} \frac{B}{B_\perp} \cos \phi \\ \frac{B}{B_\perp} \sin \phi \end{array} \right] \]
(3.35)

If the field is a Parker spiral, given by (2.28) and (2.29), the elements of the diffusion tensor are
\[ \mathbf{K} (r, \theta, \phi) = \left[ \begin{array}{ccc} \cos \psi & 0 & \sin \psi \\ 0 & 1 & 0 \\ -\sin \psi & 0 & \cos \psi \end{array} \right] \left[ \begin{array}{ccc} \kappa_{\|} & 0 & 0 \\ 0 & \kappa_{\perp} & \kappa_T \\ 0 & -\kappa_T & \kappa_{\perp} \end{array} \right] \left[ \begin{array}{c} \cos \psi & 0 & -\sin \psi \\ 0 & 1 & 0 \\ -\sin \psi & 0 & \cos \psi \end{array} \right] \]

with \( \psi \) the angle between the magnetic field and the radial direction.

If, furthermore, we assume azimuthal symmetry (\( \partial / \partial \phi = 0 \)) and a radial solar wind, \( \mathbf{V} = V(r, \theta) e_r \), then (3.33) and (3.34) reduce to
\[ \frac{\partial f}{\partial t} = \left[ \frac{\partial^2 f}{\partial r^2} + \frac{\kappa_{\theta\theta}}{r^2} \frac{\partial^2 f}{\partial \theta^2} \right] + \left[ \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 \kappa_{rr}) + \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta \kappa_{r\theta}) - V \right] \frac{\partial f}{\partial r} \]
\[ + \left[ \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 \kappa_{r\theta}) + \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta \kappa_{r\theta}) \right] \frac{\partial f}{\partial \theta} \]
\[ + \frac{1}{3r^2} \frac{\partial}{\partial r} (r^2 V_r) \frac{\partial f}{\partial \ln P} + Q_f (r, \theta, P, t), \]
(3.36)
and
\[
S_p(r, \theta, \phi) \frac{4\pi p^2}{3} = \begin{bmatrix}
-\frac{V}{3} \frac{\partial f}{\partial \ln p} - \kappa_{rr} \frac{\partial f}{\partial r} - \kappa_{r\theta} \frac{\partial f}{\partial \theta} \\
- \kappa_{\theta r} \frac{\partial f}{\partial r} - \frac{\kappa_{r\phi}}{r} \frac{\partial f}{\partial \phi} \\
- \kappa_{\phi r} \frac{\partial f}{\partial r} - \frac{\kappa_{\theta\phi}}{r} \frac{\partial f}{\partial \theta}
\end{bmatrix},
\] (3.37)
respectively, while the drift velocity (3.35) becomes
\[
v_{dr} = \frac{\kappa_T}{rB} \left[ \cot B_\phi + \frac{\partial B_\phi}{\partial \theta} - 2 \frac{B_\phi}{B} \frac{\partial B}{\partial \theta} \right],
\] (3.38)
\[
v_{d\theta} = \frac{\kappa_T}{B} \left[ \frac{2}{B} \frac{\partial B}{\partial r} - \frac{B_\phi}{r} \frac{\partial B_\phi}{\partial r} - \frac{\partial B_\phi}{\partial r} \right],
\] (3.39)
and
\[
v_{d\phi} = \frac{2 \kappa_T}{r^2 B^2} \frac{\partial B_\phi}{\partial \theta}.
\] (3.40)
Using the Parker spiral, \(v_{dr}\) and \(v_{d\theta}\) can also be written as
\[
v_{dr} = -\frac{1}{r \sin \theta} \frac{\partial}{\partial r} \left( \sin \theta \kappa_T \sin \psi \right) = -\frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \kappa_\theta \right)
\] (3.41)
and
\[
v_{d\theta} = \frac{1}{r \sin \theta} \frac{\partial}{\partial r} \left( r \kappa_T \sin \psi \right) = \frac{1}{r \sin \theta} \frac{\partial}{\partial r} \left( r \kappa_\theta \right).
\] (3.42)
Comparing these last two expressions with (3.36), the TPE can thus, equivalently be written as:
\[
\frac{\partial f}{\partial t} = \kappa_T \frac{\partial^2 f}{\partial r^2} + \kappa_{\theta\phi} \frac{\partial^2 f}{\partial \theta^2} + \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \kappa_{rr} \right) - v_{dr} - \frac{V}{3} \frac{\partial f}{\partial \ln p} - \frac{\kappa_{r\theta}}{r} \frac{\partial f}{\partial \theta} - \frac{\kappa_{\theta\phi}}{r} \frac{\partial f}{\partial \phi} + \frac{1}{3r^2 \sin \theta} \frac{\partial}{\partial \phi} \left( \sin \theta \kappa_\phi \right) - \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \kappa_\theta \right)
\] (3.43)
In Chapter 5, numerical solutions of this equation will be displayed, subject to the continuity condition (3.18) on the SWTS, together with
\[
\left( \frac{\partial f}{\partial r} \right)^+ - \frac{\kappa_T^+}{\kappa_T} \left( \frac{\partial f}{\partial r} \right)^- = \frac{V^- - V^+}{3 \kappa_T} \frac{\partial f}{\partial \ln p} - \frac{\kappa_{r\theta}}{r \sin \kappa_T} \frac{\partial f}{\partial \theta} + \frac{Q_*}{\kappa_T},
\] (3.44)
which follows from (3.20) if \(Q_f = Q_*(p) \delta(r - r_s)\) is a delta function on the shock.

### 3.6 Drift Velocities in a Parker Spiral Field

#### 3.6.1 The Drift Velocities between the Sun and the SWTS

In a Parker spiral field the drift velocities (3.38), (3.39) and (3.40) become
\[
v_{dr} = \frac{\kappa_T}{r} \left[ \frac{\sin \psi}{\tan \theta} + \cos \psi (2 \sin^2 \psi - 1) \frac{\partial \tan \psi}{\partial \theta} \right],
\] (3.45)
\[
v_{d\theta} = \kappa_T \left[ \cos \psi (1 - 2 \sin^2 \psi) \frac{\partial \tan \psi}{\partial r} + \frac{3 \sin \psi}{r} \right],
\] (3.46)
respectively. Under the assumption that \( V \) is independent of \( r \) and \( \theta \), these may be further simplified to

\[
v_{dr} = -\frac{2\kappa T}{r} \sin \psi \cos^2 \psi \frac{\partial \tan \psi}{\partial \theta},
\]

(3.48)

\[
v_{d\phi} = \frac{2\kappa T}{r} \sin(1 + \cos^2 \psi)
= \frac{2\kappa T e}{\tan \psi_e} \left( \frac{r}{r_e} \right)^2 \sin \psi \cos \phi (1 + \cos^2 \psi),
\]

(3.49)

and

\[
v_{d\phi} = \frac{2\kappa T}{r} \sin^2 \psi \cos \psi \frac{\partial \tan \psi}{\partial \theta}
= \frac{2\kappa T e}{\tan \psi_e} \left( \frac{r}{r_e} \right)^2 \sin^2 \psi \cos^2 \psi
\]

(3.50)

where \( r_e, \kappa T_e, \) and \( \psi_e \) are the values of \( r, \kappa T, \) and \( \psi \) at Earth. When \( r \gg r_e \), and for \( \tan \theta \gg 1 \) (in the near-ecliptic regions), the latitudinal component \( v_{d\theta} \) dominates. Under these circumstances, the drift velocity relative to the solar wind velocity is

\[
\frac{v_d}{V} = \frac{2\kappa T e}{r_e \cos \psi_e V} \left( \frac{r}{r_e} \right)^2 \cos \psi = 2\sqrt{2} \beta P \left( \frac{r}{r_e} \right)^2.
\]

(3.51)

For \( r \gg r_e \) and \( \tan \theta \ll 1 \), in the polar regions,

\[
\frac{v_d}{V} = -\frac{2\kappa T e}{r_e \cos \psi_e V} \left( \frac{r}{r_e} \right)^3 = -2\sqrt{2} \beta P \left( \frac{r}{r_e} \right)^3.
\]

(3.52)

Thus, the outer heliosphere is strongly drift-dominated in the sense that the drift speed greatly exceeds the solar wind convection speed.

### 3.6.2 The Drift Velocities Beyond the SWTS

Consider a heliosphere with a shock at \( r = r_s \) with compression ratio \( s \), and a solar wind given by (2.44):

\[
V' = \frac{V}{s} \left( \frac{r_s}{r} \right)^2
\]

beyond the shock. Now

\[
\tan \psi' = \frac{\Omega r \sin \theta}{V'}
\]

and

\[
\frac{\partial \tan \psi'}{\partial r} = \frac{3}{r} \tan \psi'.
\]

With this the drift velocity components, (3.45), (3.46) and (3.47) become

\[
v_{d\phi}' = -\frac{2\kappa T_e}{r e \cos \psi_e} \left( \frac{r}{r_e} \right)^2 \sin \phi' \cos^3 \psi'
\]

(3.53)

\[
v_{d\theta}' = \frac{6\kappa T e}{r e \cos \psi_e} \left( \frac{r}{r_e} \right)^2 \sin \phi' \cos^3 \psi',
\]

(3.54)

and

\[
v_{d\phi}' = \frac{2\kappa T e}{r e \cos \psi_e} \left( \frac{r}{r_e} \right)^2 \sin^2 \phi' \cos^2 \psi'.
\]

(3.55)
Thus, the $\theta$-component outside the shock has a different form than inside the shock, as can be seen by comparing (3.54) with (3.49). Outside the shock, at $r > r_s \gg r_e$ and in the near-ecliptic regions where $\tan \theta \gg 1$, the $\theta$-component dominates again and

$$
\frac{\mathbf{v}_d^2}{V} = \frac{6\sqrt{2} \beta P}{s^3} \left( \frac{r_s}{r} \right)^6 \left( \frac{r_e}{r} \right)^2.
$$

(3.56)

This drops off as $r^{-8}$. From (3.51) and (3.56) it follows that the drop across the shock is $3/s^3$. For a shock with $s = 4$ and $r_s = 60$ AU, this is only $\sim 5\%$ of the pre-shock value.

At the poles for $r > r_s \gg r_e$ we have that

$$
\frac{\mathbf{v}_d^2}{V} = -2\sqrt{2} \beta P s \left( \frac{r}{r_e} \right)^2 \left( \frac{r}{r_s} \right)^2.
$$

(3.57)

In other words, the drift velocity increases with $r$ even more strongly than inside the shock. On the shock (at $r_s = 60$ AU, say) there is a moderate jump of

$$
\frac{2\sqrt{2}}{15} \beta P s \approx 0.75 \beta P,
$$

before it grows again $\propto r^4$.

From (3.53) and (3.39) it follows that the radial and $\theta$-components are equal when $\tan \theta = 1/3$ or $\theta = 18.4^\circ$. Thus, for all regions more than about $20^\circ$ from the poles, drift effects are small outside the shock. Since the wind is also divergence free ($\nabla \cdot \mathbf{V} = 0$) there, the TPE outside the shock is well-approximated by

$$
\mathbf{V} f - \mathbf{K} \cdot \nabla f = \text{constant}.
$$

It must be borne in mind, however, that these small drift velocities beyond the shock are valid for a Parker spiral field. In Section 2.3.7 it was shown that this field is unrealistic beyond the shock. For the demonstration purposes of this thesis it will, however, be used.

### 3.7 Neutral Sheet and Shock Drift

The heliosphere contains two discontinuities in $\mathbf{B}$. At the shock the field increases by the compression ratio $s$, while at the neutral sheet it switches direction. This leads to the basic drift patterns of Figure 3.1. The shock drift is easily handled, because Jokipii (1987) has shown that

![Figure 3.1: Drift mechanisms across two types of discontinuities in the magnetic field: (a) shock drift, and (b) neutral sheet drift.](image)

it is completely and correctly incorporated in the shock boundary condition (3.44).
A similar simple expression in principle exists for the neutral sheet drift, where the drift flux across the sheet must be zero. If it is a flat sheet, aligned along the ecliptic plane, the $\theta$-component of (3.37) must be zero, or

$$\frac{\partial f}{\partial \theta} = -r \frac{\kappa_{\theta \theta}}{\kappa_{\theta \theta}} \frac{\partial f}{\partial r}.$$  

In practice, however, this condition is difficult to implement because the neutral sheet is wavy. The wavelength is $(\text{solary wind speed}) \times (\text{solar rotation period})$, or 

$$VT \approx 400 \text{ km s}^{-1} \times 27 \text{ days} \approx 6 \text{ AU}.$$  

Now, in a numerical code, one can typically afford $\sim 100$ radial grid points. If the heliosphere is $\sim 100 \text{ AU}$ in extent, this means one radial grid point per AU. This implies that the wavy sheet can only be 'sampled' about 6 times per wave. Outside the termination shock, where the total drift velocity decreases to about 1.5 AU, there is almost one sample per wave. Thus, this procedure becomes extremely inaccurate.  

For this reason Potgieter and Moraal (1985), Burger and Potgieter (1989) and Hattingh (1994) have developed methods to simulate the drift effects along a wavy neutral sheet. In the next section the model used in this thesis will be described.

### 3.8 The Neutral Sheet Drift Model

The neutral sheet (2.42) may be written as

$$\theta_{ns} = \frac{\pi}{2} + \sin^{-1} \left[ \sin \alpha \sin (\phi - \phi_0 + R) \right],$$  

(3.58)

with

$$R = \frac{\Omega r}{V}$$  

(3.59)

and $\phi_0$ an arbitrary phase constant.

The total drift velocity may be decomposed into two parts: $v_d = v_{dm} + v_{ns}$. The first term gives the drift due to gradients and curvatures in the large-scale magnetic field. The second term gives the drift along the neutral sheet. Figure 3.2 shows the geometry of the neutral sheet drift problem. Since the sheet (3.58) is three-dimensional in nature, it is impossible to treat the full neutral sheet drift problem in a two-dimensional model and, therefore, an appropriate azimuthal average must be calculated. Such averages were calculated by Potgieter and Moraal (1985), Burger and Potgieter (1989), and recently, by Hattingh (1994). In this thesis we use a model based on the principles of the Hattingh model, but with a few technical differences.

#### 3.8.1 Azimuthal Averages of Gradient and Curvature Drift

When $\pi/2 - \alpha < \theta < \pi/2 + \alpha$, Figure 3.3 shows that the drift velocity will switch sign, depending on the value of $\phi$, e.g. as shown by points a and b. For a particular value of $\theta$ the drift state will be 'North' except for the fraction

$$\frac{\Delta \phi}{2\pi} = \frac{1}{2} \frac{1}{\pi} \sin^{-1} \left[ \cos \theta \right].$$  

Thus, on the line $\theta^*$ the drift states are mixed according to

$$\left(1 - \frac{\Delta \phi}{2\pi} \right) N + \frac{\Delta \phi}{2\pi} S.$$  

33
Figure 3.2: Drift along the wavy neutral sheet.

Since the drift state $S = -N$, this is $(1 - \Delta \phi / \pi)N$. Thus the drift velocity must be scaled down with a factor

$$\frac{2}{\pi} \sin^{-1} \left( \frac{\cos \theta}{\sin \alpha} \right)$$

when $\pi/2 - \alpha < \theta < \pi/2 + \alpha$ to get the proper $\phi$-averaged value.

Figure 3.3: The wavy neutral sheet as function of $\phi$. 
3.8.2 Azimuthal Average of Neutral Sheet Drift

If $\cos \beta$ gives the projection of the drift velocity $v_{ns}$ into the $(r, \theta)$-plane, Figure 3.2 shows that the drift velocity components are

\[
v_{nsr} = v_{ns} \cos \beta \sin \psi, \quad v_{nsth} = v_{ns} \sin \beta, \quad v_{ns} = v_{ns} \cos \beta \cos \psi.
\]

In an azimuthal symmetric model the $\phi$-average, $<v_{nsth}>_\phi$ is of no consequence, $<v_{nsth}>_\phi = 0$, and the problem reduces to finding the average value of the radial component, i.e.,

\[
<v_{nsr}>_\phi = <v_{ns} \cos \beta \sin \psi>.
\]  

(3.61)

The average first requires a spatially averaged, and properly redistributed value of the magnitude, $v_{ns}$. Burger (1987) and Burger and Potgieter (1989) showed that particles within two gyroradii, $2r_g$ on either side of the sheet can partake in neutral sheet drift, and that these particles, on average, drift with a velocity $v_{ns} = v/6$, with $v$ the particle velocity, along the sheet. Now, $v/6 \times 4r_g = 2\beta P/(3B) = 2\kappa_T$. Thus, the correct neutral sheet drift flux will be maintained if we center the drift along $\theta = \pi/2$, let it drop linearly to zero at $\theta = \pi/2 - \alpha$ and $\theta = \pi/2 + \alpha$, and choose the velocity such that the area under this profile remains $2\kappa_T$ as shown in Figure 3.4.

\[
\frac{2\kappa_T}{\alpha^* r}[1 - (\frac{\pi}{2} - \theta)/\alpha^*]
\]

Figure 3.4: Distribution of neutral sheet drift as function of $\theta$.

The quantity $\alpha^* = \alpha$, unless the tilt angle is so small that the wave of the sheet covers less than five points on the numerical $\theta$-grid (see Section 6.2.5). In this case $\alpha^*$ is given such a value that the drift is distributed over five grid points, at $\theta = \pi/2$, and two points grid points on either side of it.

Thus (3.61) becomes

\[
<v_{nsr}> = \frac{2}{\alpha^* r} \left(1 - \frac{\pi/2 - \theta}{\alpha^*}\right) <\kappa_T \cos \beta \sin \psi> \phi .
\]  

(3.62)

Noting that $\kappa_T \propto 1/B$, and for a Parker spiral,

\[
B = B_e \left(\frac{r_e}{r}\right) \frac{2 \cos \psi_e}{\cos \psi},
\]

(3.62) can be written as

\[
<v_{nsr}> = \frac{2\kappa_T e}{\alpha^* r} \left(\frac{\pi}{r_e} \right)^2 \frac{1}{\cos \psi_e} \left[1 - \frac{\pi/2 - \theta}{\alpha^*}\right] <\cos \beta \cos \psi \sin \psi> \phi .
\]  

(3.63)
The quantity in \(< \cdots >\) must be evaluated at the position of the sheet, i.e., at \(\theta = \theta_{ns}\) for each value of \(\phi\). We notice that

\[
\tan \beta = -r \frac{d\theta_{ns}}{dr} = -\frac{r d\theta_{ns}}{d\cos \theta_{ns}} \frac{d \cos \theta_{ns}}{dr} = -\frac{r}{\sin \theta_{ns}} \frac{d \cos \theta_{ns}}{dr}
\]

Thus

\[
< \cos \beta \sin \psi \cos \psi > = \left( \frac{R^2 \sin^2 \theta_{ns}}{1 + R^2 \sin^2 \theta_{ns}} \frac{1}{\sqrt{R^2 \sin^2 \theta_{ns} + R^4 \left(1 - \frac{R dV}{dR}\right)^2 (\sin^2 \alpha - \cos^2 \theta_{ns})}} \right)
\]

This expression must be evaluated numerically. At large radial distances, it gives that \(< v_{nsr} > \propto 1/r\), which implies that the total flux carried by the sheet, which is \(\propto < v_{nsr} > \propto r\) is independent of \(r\). Beyond the termination shock, where

\[
V = \frac{V^-}{\frac{r_a}{r}},
\]

where \(V^-\) is the value immediately inside the shock, the term \((R/V)(dV/dR) = -2\), and this is automatically handled.

### 3.9 The Diffusion Coefficients

The diffusion tensor (3.7) contains two distinct elements. Firstly, \(\kappa_\parallel\) describes the process of pitch angle scattering by irregularities on the magnetic field lines. Secondly, \(\kappa_\perp\) describes the migration of particles from one field line to the adjacent one (cross-field diffusion) due to the random walk of the magnetic field lines, as developed by Jokipii and Parker (1970). These diffusion coefficients have associated diffusion mean free paths given by \(\lambda_\parallel = 3\kappa_\parallel /v\) and \(\lambda_\perp = 3\kappa_\perp /v\).

The magnitudes, and spatial and energy dependence of these individual elements in the diffusion tensor are still a relatively open question. Advances in quasi-linear theory (QLT), as well as advances in the particle transport in magnetic fields with helicity and adiabatic focussing do shed more light on these aspects of the diffusion tensor (Bieber et al., 1987; Bieber and Burger, 1990; Bieber et al., 1994). However, the resulting spatial and energy dependences are complex, and are, therefore, not of much use for demonstration purposes. Simpler approximations to these energy and spatial dependences are, thus, needed.

#### 3.9.1 Diffusion Parallel and Perpendicular to the Magnetic Field

According to early QLT (Jokipii, 1966), the energy dependence of the parallel diffusion mean free path, \(\lambda_\parallel\), should be \(\propto P^{1/3}\) below a particle rigidity of about 10 GV. Above this, the rigidity dependence of \(\lambda_\parallel\) steepens towards \(P^{1.5}\). Observational evidence in the rigidity range of 30 kV to 20 GV indicates that the magnitudes are much larger than predicted by standard QLT as is shown in Figure 3.5. In the analysis by Bieber et al. (1994) they have shown, however, that a model based on scattering in dynamic turbulence with composite slab/two-dimensional geometry, yields parallel diffusion mean free paths for protons that are slightly smaller at low energies and considerably larger (by almost a factor 5) than that predicted by Jokipii (1966).
Figure 3.5: Parallel diffusion mean free path, $\lambda_{\|}$, as function of particle rigidity. The circles are actual observed values and the triangles are lower-limit values. The filled symbols denote electrons and the open symbols protons. The dotted line is the prediction of standard QLT (Jokipii, 1966). The enclosed area is the Palmer observational consensus (adapted from Bieber et al., 1994).

According to the Palmer consensus (Palmer, 1982) the value of the parallel diffusion mean free path, $\lambda_{\|}$, is in the range $0.08 \text{ AU} \leq \lambda_{\|} \leq 0.3 \text{ AU}$ for particle rigidities $\leq 5 \text{ GV}$. From this and $\kappa_{\|} = \nu \lambda_{\|}/3$ it follows that the magnitude of $\kappa_{\|}$ must be in the range $(1-5) \times 10^{22} \text{ cm}^2/\text{s}$. The consensus value for $\kappa_{\perp}$ requires that $\kappa_{\perp}$ must be such that $\kappa_{\perp}/\beta \approx 10^{21} \text{ cm}^2/\text{s}$. This requires that $\kappa_{\perp}/\kappa_{\|} < 0.1$ at Earth.

Morfill and Völk (1979) did detailed calculations of the spatial dependence, using results from a WKB analysis of fluctuations and quasi-linear theory to obtain the latitudinal dependence of the parallel diffusion mean free path, $\lambda_{\|}$. According to Jokipii and Davila (1981), Morfill and Völk's calculations lead to an approximate $B^{-1}$ dependence.

Jokipii and Davila used the spatial dependence $\kappa_{\|} \propto B_e/B$ for $r > 1 \text{ AU}$ while for $r \leq 1 \text{ AU}$ $\kappa_{\|}$ was independent of $r$ but with a latitudinal dependence $\propto B_e/B$. This was done to enhance the stability of their numerical code, but since our code(s) do not have this problem, a $B_e/B$ dependence is generally assumed for all $r$.

In view of these uncertainties, the present code was programmed with the general forms

$$\kappa_{\|} = (\kappa_{\|})_0 \left( \frac{B_e}{B} \right)^{a_{\|}} \beta \left( \frac{P}{P_0} \right)^{b_{\|}}$$  \hspace{1cm} (3.65)

and

$$\kappa_{\perp} = (\kappa_{\perp})_0 \left( \frac{B_e}{B} \right)^{a_{\perp}} \beta \left( \frac{P}{P_0} \right)^{b_{\perp}}$$  \hspace{1cm} (3.66)

as suggested by Reinecke and Moraal (1992). For demonstration purposes, $a_{\|}$, $a_{\perp}$, $b_{\|}$ and $b_{\perp}$ are all set to 1, while $(\kappa_{\|})_0 = 1.8 \times 10^{22} \text{ cm}^2/\text{s}$ and $(\kappa_{\perp})_0 = 1.8 \times 10^{20} \text{ cm}^2/\text{s}$. Thus, the rigidity
dependence approximately agrees (on average) with QLT predictions, while the magnitudes agree with the Palmer consensus.

### 3.10 Summary

The purpose of this chapter was to write down a cosmic-ray transport model to be implemented in a numerical model. Below follows a short summary of the results of this chapter. The differential equation to be solved is the two-dimensional time-dependent TPE

\[
\frac{\partial f}{\partial t} = \frac{\kappa_{rr}}{r^2} \frac{\partial^2 f}{\partial r^2} + \frac{\kappa_{\theta\theta}}{r^2} \frac{\partial^2 f}{\partial \theta^2} + \left[ \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \kappa_{rr} \right) - v_{dr} - V \right] \frac{\partial f}{\partial r} + \frac{1}{r^2} \frac{\partial}{\partial \theta} \left( \sin \theta \kappa_{\theta\theta} \right) - \frac{v_{d\theta}}{r} \frac{\partial f}{\partial \theta} + \frac{1}{3r^2} \frac{\partial}{\partial r} \left( r^2 V \right) \frac{\partial f}{\partial \ln P} + Q(r, \theta, P, t). \tag{3.43}
\]

The drift velocities are given by

\[
v_{dr} = \frac{\kappa_{TT}}{rB} \left[ \cot \theta B_\phi + \frac{\partial B_\phi}{\partial \theta} - 2 \frac{B_\phi}{B} \frac{\partial B}{\partial \theta} \right], \tag{3.38}
\]

and

\[
v_{d\theta} = \frac{\kappa_{TT}}{B} \left[ 2 \frac{B_\phi}{r} \frac{\partial B}{\partial r} - \frac{B_\phi}{r} - \frac{\partial B_\phi}{\partial r} \right]. \tag{3.39}
\]

The drift in the influence zone of the wavy neutral sheet is calculated by

\[
<v_{dm}> = \begin{cases} v_{dm} & \text{if } 0 \leq \theta < \frac{\pi}{2} - \alpha, \\ -v_{dm} \frac{2}{\pi} \sin^{-1} \left( \frac{\cos \beta}{\sin \alpha} \right) & \text{if } \frac{\pi}{2} - \alpha \leq \theta \leq \frac{\pi}{2} + \alpha, \\ -v_{dm} & \text{if } \frac{\pi}{2} + \alpha < \theta \leq \pi. \end{cases} \tag{3.67}
\]

The azimuthal average of the radial component of the neutral sheet drift velocity is given by

\[
<v_{nsr}>_\phi = <v_{ns} \cos \beta \sin \psi > \tag{3.68}
\]

where this average is given by (3.64).

At the shock the matching conditions

\[
\left( \frac{\partial f}{\partial r} \right)^- = \frac{\kappa_{rr}^+}{\kappa_{rr}} \left( \frac{\partial f}{\partial r} \right)^+ - \frac{V^- - V^+}{3 \kappa_{rr}} \frac{\partial f}{\partial \ln P} - \frac{\kappa_{r\theta}^- - \kappa_{r\theta}^+}{r \kappa_{rr}} \frac{\partial f}{\partial \theta} + \frac{Q^*}{\kappa_{rr}} \tag{3.44}
\]

and

\[
f^- = f^+ \tag{3.18}
\]

must be valid.

With this information at hand, one can now proceed to develop a solution technique with which equation (3.43) can be solved.
Chapter 4

Numerical Solution of Partial Differential Equations with Finite Difference Methods

4.1 Introduction

In this chapter some of the fundamental concepts of finite difference techniques are outlined and the basic discretization techniques are discussed. Eventually, these techniques are applied to equation (3.43), the two-dimensional transport or Parker equation, to derive a specific discretization method and solve it for cosmic-ray transport in the heliosphere.

Sufficient theory and basics will be described to form a continuous line from the basic concepts to the methods ultimately used. The primitive discretization techniques, on which the more sophisticated methods are based, will provide the uninitiated reader with insight into the basic philosophy and history of the numerical techniques on which the model is based. Techniques to solve elliptical, ultra-hyperbolic and second-order hyperbolic partial differential equations will not be shown because they are not relevant to this particular field of study. Unless otherwise referenced, the development of this chapter is based on the book of Lapidus and Pinder (1982).

Thus, the aim of this chapter is to develop a numerical technique with which equation (3.43) can be solved. This equation is a partial differential equation (PDE) of the form

$$\frac{\partial f}{\partial t} = a_0 \frac{\partial^2 f}{\partial r^2} + b_0 \frac{\partial^2 f}{\partial \theta^2} + c_0 \frac{\partial f}{\partial r} + d_0 \frac{\partial f}{\partial \theta} + e_0 \frac{\partial f}{\partial \ln r} + Q$$

(4.1)

or

$$f_t = a_0 f_{rr} + b_0 f_{\theta\theta} + c_0 f_r + d_0 f_\theta + e_0 f_{\ln r} + Q$$

with coefficients

$$a_0 = \kappa_{rr}$$
$$b_0 = \frac{\kappa_{\theta\theta}}{r^2}$$
$$c_0 = \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \kappa_{rr} \right) + V_d r - V$$
$$d_0 = \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \kappa_{\theta\theta} \right) + \frac{V_d \theta}{r}$$
$$e_0 = \frac{1}{3r^2} \frac{\partial}{\partial r} \left( r^2 V \right).$$

This is a linear PDE of order two in four variables, which is an example of the general form

$$\sum_{i,j=1}^{n} a_{ij} u_{x_i x_j} + \sum_{i=1}^{n} b_i u_{x_i} + cu = d.$$  

(4.2)
Second-order PDEs are classified in four distinct families, namely parabolic, elliptic, hyperbolic and ultra-hyperbolic equations. Before a PDE can be solved, either analytically or numerically, it must first be classified to determine to which family it belongs, since the equations in each of these four families have different characteristics, and different types of solution methods are required.

4.2 Classification of PDEs

4.2.1 Classification of Second-Order PDEs

The cosmic-ray transport equation is a linear second order PDE in four variables (two spatial variables \( r \) and \( \theta \), particle rigidity \( P \), and time \( t \)) and should be classified before any attempt can be made to solve it.

Consider a general linear PDE of order two in \( n \) variables such as equation (4.2):

If \( U_{x;xj} = U_{x;x;} \), then the principal part (that which contains only second-order derivatives) can always be arranged so that \( a_{ij} = a_{3ii} \) therefore, the \( n \times n \) matrix \( A = [aij] \) is symmetrical. Every real, symmetric \( n \times n \) matrix has \( n \) real eigenvalues. These eigenvalues are the (possibly repeated) zeros of an \( n \)th-degree polynomial in \( \lambda \), \( \det(A - \lambda I) \), where \( I \) is the \( n \times n \) identity matrix. Let \( P \) denote the number of positive eigenvalues, and \( Z \) the multiplicity of the eigenvalue zero, of the matrix \( A \). Then the PDE is

- hyperbolic if \( Z = 0 \) and \( P = 1 \) or \( Z = 1 \) and \( P = n - 1 \),
- elliptic if \( Z = 0 \) and \( P = n \) or \( Z = 0 \) and \( P = 0 \),
- ultra-hyperbolic if \( Z = 0 \) and \( 1 < P < n - 1 \), and
- parabolic if \( Z > 0 \) (equivalently, if \( \det A = 0 \)).

If \( a_{ij} \) depend on one or more of the variables, the type of the PDE under consideration may vary with position in the domain (DuChateau and Zachmann, 1986, Chapter 2, §2.1).

If equations (4.1) and (4.2) are compared it is apparent that \( a_{ij} = 0 \) if \( i \neq j \) and \( a_{11} = a_{0} \), \( a_{22} = b_{0} \) and \( a_{33} = a_{44} = 0 \). Furthermore, \( b_{1} = c_{0} \), \( b_{2} = d_{0} \), \( b_{3} = e_{0} \), \( b_{4} = 1 \), \( c = 0 \) and \( d = Q \). The different variables are \( x_{1} = r \), \( x_{2} = \theta \), \( x_{3} = \ln P \) and \( x_{4} = t \). Thus,

\[
A = \begin{bmatrix}
a_{0} & 0 & 0 & 0 \\
0 & b_{0} & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

and \( \det A = 0 \). Therefore, the TPE is of parabolic form.

4.2.2 Classification of First-Order Equations

In Section 4.7 it will be shown that the detailed numerical solution of the TPE (4.1) also leads to a first-order PDE. Thus, it is also necessary to discuss the classification of first-order PDEs.

The general quasi-linear system of \( n \) first-order PDEs in \( n \) functions of two independent variables is

\[
\sum_{j=1}^{n} a_{ij} \frac{\partial u_{j}}{\partial x} + \sum_{j=1}^{n} b_{ij} \frac{\partial u_{j}}{\partial y} = c_{j}, \quad i = 1, 2, \ldots, n
\]

where \( a_{ij}, b_{ij} \) and \( c_{j} \) may depend on \( x, y, u_{1}, u_{2}, \ldots, u_{n} \). If each \( a_{ij} \) and \( b_{ij} \) is independent of \( u_{1}, u_{2}, \ldots, u_{n} \), the above system is called almost linear. If, in addition, each \( c_{i} \) depends linearly on \( u_{1}, u_{2}, \ldots, u_{n} \), the system is said to be linear.
In terms of the \( n \times n \) matrices \( A = [a_{ij}] \) and \( B = [b_{ij}] \), and the column vectors \( u = [u_1, u_2, \ldots, u_n]^T \) and \( c = [c_1, c_2, \ldots, c_n]^T \), the system above can be expressed as

\[
Au_x + Bu_y = c.
\]

If \( A \) or \( B \) is non-singular, it is usually possible to classify the above system according to type. Suppose \( \det(B) \neq 0 \). Define a polynomial of degree \( n \) in \( \lambda \) by

\[
P_n(\lambda) \equiv \det(A^T - \lambda B^T) = \det(A - \lambda B).
\]

The system is classified as

- **elliptic** if \( P_n(\lambda) \) has no real zeros;
- **hyperbolic** if \( P_n(\lambda) \) has \( n \) real, distinct zeros; or if \( P_n(\lambda) \) has \( n \) real zeros, at least one of which is repeated, and the generalized eigenvalue problem \((A^T - \lambda B^T)t = 0\) yields \( n \) linearly independent eigenvectors, \( t \);
- **parabolic** if \( P_n(\lambda) \) has \( n \) real zeros, at least one of which is repeated, and the above generalized eigenvalue problem yields fewer than \( n \) linearly-independent eigenvectors.

An exhaustive classification cannot be carried out when \( P_n(\lambda) \) has both real and complex zeros. Since \( a_{ij} \) and \( b_{ij} \) may depend on \( x, y, u_1, u_2, \ldots, u_n \), the above classification may again be position- and/or solution-dependent (DuChateau and Zachmann, 1986, §5.1).

In the special case of \( n = 1 \), the system of equations reduces to one equation:

\[
a_{11}\frac{\partial u_1}{\partial x} + b_{11}\frac{\partial u_1}{\partial y} = c_1
\]

and the polynomial \( P_n(\lambda) \), is simply \( a_{11} - \lambda b_{11} \) with only one real zero. Therefore, such an equation is always of hyperbolic form.

### 4.3 Basic Numerical Techniques

#### 4.3.1 Introduction

A PDE in a set of \( n \) variables (co-ordinates) is always solved on a specific region of interest in the \( n \)-dimensional co-ordinate space. This region is called the **domain** of the PDE. Since a specific PDE has an infinite number of possible solutions on this domain, sufficient information must be specified on most (preferably all) boundaries of this domain to determine the specific, one to one, solution. These sets of information that fix the solution at the boundaries of the domain are called the **boundary conditions** of the solution. Among numerous finer distinctions, one may distinguish between four simple types of initial and boundary conditions. These are

- **Dirichlet** conditions, where the solution is specified at the boundary;
- **Neumann** conditions, where the slope is specified on the boundary;
- **Cauchy** conditions, where both the solution value and the slope are specified on the boundary in two distinct equations, and
- **Robbins** conditions, where the solution value and the slope are specified on the boundary in the form of a first-order differential equation.
In the case where a boundary condition is specified on one side of the domain but none on the opposite side, the boundary condition (for obvious reasons) is referred to as an initial value. With this distinction, one may therefore distinguish between three kinds of problems in solving PDEs:

- the pure boundary value problem, where matched boundary conditions are specified on all boundaries;
- the initial-boundary value problem, where initial conditions are imposed on some boundaries, and matched boundary conditions on at least one pair of opposing boundaries, and
- the so-called pure initial value problem, where all boundary conditions are initial conditions.

The details are shown schematically in Figure 4.1.

4.3.2 Finite Difference Formulae

Finite difference methods are discrete techniques that are used when the domain of interest is represented by a set of points or nodes, forming a structured grid with definite boundaries. Information between these nodes or points is then obtained by Taylor series expansions. These expansions are used to approximate the PDE with a discrete difference equation. The transformation of the PDE into the discrete difference equation is frequently called the discretization of the PDE. In order to discretize a PDE, one must start by replacing the partial derivatives with discrete approximations, called finite difference formulae.

In this section the finite difference approximations of first- and second-order derivatives will be derived briefly. Since the solution of the TPE may change very rapidly in certain regions in the heliosphere, it is desirable to choose a fine mesh in those areas, in order to keep the numerical code stable. In other regions the solution may change very slowly and a fine mesh is unnecessary in these regions.

The standard solution of such a problem is to transform the spatial coordinates in the TPE in such a way that a linear (evenly spaced) grid in these new spatial coordinates will yield smaller steps in the untransformed coordinates in the desired regions. This is an elaborate process that must be repeated each time a new coordinate transformation is desired. To alleviate this problem, the well-known finite difference formulae (e.g. Lapidus and Pinder, 1982, §2.1) can be written in terms of an uneven grid (Moraal, personal communication, 1991). To the best of our knowledge, this approach is nowhere to be found in the literature.

Consider an uneven grid in one dimension:
The following conventions are used to denote the solution at different grid points:

\[
\begin{align*}
  f(x - \Delta x_2 - \Delta x_4) &= f_{i-2} \\
  f(x - \Delta x_2) &= f_{i-1} \\
  f(x) &= f_i \\
  f(x + \Delta x_1) &= f_{i+1} \\
  f(x + \Delta x_1 + \Delta x_3) &= f_{i+2}
\end{align*}
\]

The derivatives are represented as

\[
\frac{\partial f}{\partial x} = f' \quad \text{and} \quad \frac{\partial^2 f}{\partial x^2} = f''.
\]

The basis of finite difference techniques is the Taylor series expansion about a small interval called a smallness parameter, \( \Delta x \):

\[
f(x \pm \Delta x) = f(x) \pm \Delta x f' + \frac{(\Delta x)^2}{2} f'' + \ldots
\]  \hspace{1cm} (4.3)

The truncation error after \( n \) terms is

\[
\sum_{i=n}^{\infty} \frac{(\Delta x)^i}{i!} f^{(i)}
\]

where \( f^{(i)} \) is the \( i \)-th derivative with respect to \( x \).

Truncating the series (on an uneven grid) after two terms gives

\[
\begin{align*}
  f_{i+1} &\approx f_i + \Delta x_1 f' \\
  f_{i-1} &\approx f_i - \Delta x_2 f'
\end{align*}
\]

from which follow the first-order one-sided derivatives

\[
f' = \frac{f_{i+1} - f_i}{\Delta x_1} \quad \text{(forward difference)} \]  \hspace{1cm} (4.4)

and

\[
f' = \frac{f_i + f_{i-1}}{\Delta x_2} \quad \text{(backward difference)} \]  \hspace{1cm} (4.5)

which are correct to first order in \( \Delta x \), written as \( O(\Delta x) \).

In a similar fashion, the second-order truncation on an uneven grid gives the following four expansions:

\[
\begin{align*}
  f_{i+1} &= f + \Delta x_1 f' + (\Delta x_1)^2 f''/2 + O((\Delta x_1)^3), \hspace{1cm} (4.6) \\
  f_{i+2} &= f + (\Delta x_1 + \Delta x_3) f' + (\Delta x_1 + \Delta x_3)^2 f''/2 + O((\Delta x_1 + \Delta x_3)^3), \hspace{1cm} (4.7) \\
  f_{i-1} &= f - \Delta x_2 f' + (\Delta x_2)^2 f''/2 + O((\Delta x_2)^3), \hspace{1cm} (4.8) \\
  f_{i-2} &= f - (\Delta x_2 + \Delta x_4) f' + (\Delta x_2 + \Delta x_4)^2 f''/2 + O((\Delta x_2 + \Delta x_4)^3), \hspace{1cm} (4.9)
\end{align*}
\]
which are correct to $O((\Delta x)^2)$.

To obtain the *central derivative* formulae, the operations $(4.6) \times (\Delta x_2)^2 - (4.8) \times (\Delta x_1)^2$ and $(4.6) \times \Delta x_2 + (4.8) \times \Delta x_1$ yield the following expressions for the first

$$f' = \frac{-\Delta x_1}{\Delta x_2(\Delta x_1 + \Delta x_2)} f_{i-1} + \frac{\Delta x_1 - \Delta x_2}{\Delta x_1 \Delta x_2} f_i + \frac{\Delta x_2}{\Delta x_1(\Delta x_1 + \Delta x_2)} f_{i+1} \quad (4.10)$$

and second

$$f'' = \frac{2}{\Delta x_2(\Delta x_1 + \Delta x_2)} f_{i-1} - \frac{2}{\Delta x_1 \Delta x_2} f_i + \frac{2}{\Delta x_1(\Delta x_1 + \Delta x_2)} f_{i+1} \quad (4.11)$$

derivative, respectively.

To obtain the *second-order one-sided* differences, the operations $(4.8) \times (\Delta x_2 + \Delta x_4)^2 - (4.9) \times (\Delta x_2)^2$ and $(4.6) \times (\Delta x_1 + \Delta x_3)^2 - (4.7) \times \Delta x_1^2$ yield

$$f' = -\frac{2\Delta x_1 + \Delta x_3}{\Delta x_1(\Delta x_1 + \Delta x_3)} f_i + \frac{\Delta x_1 + \Delta x_3}{\Delta x_1 \Delta x_3} f_{i+1} - \frac{\Delta x_1}{\Delta x_3(\Delta x_1 + \Delta x_3)} f_{i+2} \quad \text{(forward)} \quad (4.12)$$

and

$$f' = \frac{2\Delta x_2 + \Delta x_4}{\Delta x_2(\Delta x_2 + \Delta x_4)} f_i - \frac{\Delta x_2 + \Delta x_4}{\Delta x_2 \Delta x_4} f_{i-1} + \frac{\Delta x_2}{\Delta x_4(\Delta x_2 + \Delta x_4)} f_{i-2} \quad \text{(backward)} \quad (4.13)$$

In the case of an even grid, where $\Delta x_1 = \Delta x_2 = \Delta x_3 = \Delta x_4 = \Delta x$, equations (4.4) and (4.5) reduce to their well-known even-grid equivalents,

$$f' = \frac{f_{i+1} - f_i}{\Delta x}$$

and

$$f' = \frac{f_i - f_{i-1}}{\Delta x}$$

respectively. Similarly, equations (4.10), (4.11), (4.12) and (4.13), respectively, become

$$f' = \frac{f_{i+1} - f_{i-1}}{2\Delta x} \quad , \quad (4.14)$$

$$f'' = \frac{f_{i-1} - 2f_i + f_{i+1}}{(\Delta x)^2} \quad , \quad (4.15)$$

$$f' = \frac{-3f_i + 4f_{i+1} - f_{i+2}}{2\Delta x} \quad (4.16)$$

and

$$f' = \frac{3f_i - 4f_{i-1} + f_{i-2}}{2\Delta x} \quad . \quad (4.17)$$

### 4.3.3 Implementation

In order to solve ordinary differential equations (ODEs) and PDEs numerically, the above-mentioned formulae are employed to replace the differentiations in the equations. This discretized differential equation (difference equation) may now be rewritten in another way, grouping grid points rather than derivatives in such a way that the solution in one or more grid points is written in terms of the solution in other grid points that are known beforehand. Therefore, if one specifies an initial condition on one boundary and boundary conditions on the boundaries orthogonal to the first, one may step away from the boundary on which the initial condition
is specified, and calculate the solution on the rest of the grid. In the case of parabolic and hyperbolic PDEs, the variable which increases as one steps away from the boundary on which the initial condition is specified, is called the *stepping parameter*.

At this point it is important to note that the direction in which one steps during the solution process (also called the direction of integration) must always be in the same direction that information in the physical system propagates. For example, in a typical time-dependent problem information propagates in the direction of increasing time from an initial condition at an early time to yield a result at a later time.

There exist many ways to discretize a differential equation, especially for PDEs. Some of these methods enable one to write the solution in one grid point in terms of the solution in all other adjacent grid points in the so-called computational molecule. These methods are called *explicit* methods. The calculation of the solution in this case becomes almost trivial. Although these methods are generally very simple, more complex and sophisticated explicit methods have been developed. Explicit methods have the useful property that one can directly march the solution forward in the stepping parameter from one level to the next. They can be applied to pure initial-value problems and initial-boundary value problems. Unfortunately, these methods suffer the drawbacks of limited accuracy and conditional stability.

Differential equations may be discretized in more involved ways, yielding discretized forms where adjacent (usually three) grid points are written in terms of other, previously known, grid points. In this case, a system of equations is built up from one boundary to the other and this is then solved to obtain more accurate and stable solutions than in the explicit case. This family of methods is usually called *implicit*. These implicit methods, unfortunately, require a matrix inversion at each time step and are, therefore, only applicable to initial-boundary value problems.

### 4.4 Numerical Solution of Parabolic PDEs

Since it was shown in Section 4.2 that the TPE is a parabolic PDE of the general form

\[
a u_t = \sum_{i=1}^{n} \left( b_i u_{x_ix_i} + c_i u_{x_i} \right) + du + e,
\]

(4.18)

the methods with which parabolic PDEs are solved must be thoroughly explored in order to decide on a solution technique. All the textbook discretization techniques for parabolic PDEs were developed under the condition that the coefficients \( b_i \) must be strictly positive. In all PDEs of this form, the logical choice for the stepping parameter is the variable \( t \) which will be taken to be a time co-ordinate. The other variables will be referenced as being 'spatial' co-ordinates. From equation (4.1) it is clear that the rigidity variable, \( P \), can be classified as such a 'spatial' variable.

Using (4.14) and (4.15), the following notations will be used to denote first-order

\[
\delta_x = \frac{1}{2\Delta x} (u_{i+1} - u_{i-1})
\]

and second-order

\[
\delta_x^2 = \frac{1}{(\Delta x)^2} (u_{i-1} - 2u_i + u_{i+1})
\]

derivatives, respectively.
4.4.1 PDEs in One Spatial Dimension

In the special case when \( n = 1, x_1 = x \) and \( a = c_1 = d = e = 0 \), equation (4.18) reduces to

\[
 u_t = b_1 u_{xx}.
\]  

Consider the solution of this equation on the grid \((t_s, x_i) = (s \Delta t, i \Delta x)\) and define

\[
 r = b_1 \Delta t / (\Delta x)^2.
\]

Equation (4.19) must be discretized in such a way that the solution one time-step \((s = 1)\) away from the boundary \((s = 0)\), where the initial condition is specified, can be calculated from the initial condition itself. Away from the boundary this means that the solution at the \( s + 1 \) level must be calculated from the solution at the \( s \) level. This implies that a first-order one-sided difference must be used to approximate the first-order derivative to \( t \), the stepping parameter.

Since no such limitations are imposed on the spatial co-ordinate, second-order accurate central differences can be used to approximate the spatial derivatives. There are different ways to do this. Three of these are discussed:

The Forward-Difference Explicit Method:

In the forward-difference explicit method, the spatial difference is calculated at the \( s \) level where the solution is known:

\[
 \frac{u_j^{s+1,i} - u_j^{s,i}}{\Delta t} = b_1 \delta x^2 u_j^{s,i} = b_1 \frac{u_j^{s,i+1} - 2u_j^{s,i} + u_j^{s,i-1}}{(\Delta x)^2}. 
\]

Since only one unknown, \( u_{s+1,i} \), appears in the resulting difference equation, this method is an explicit method. This difference equation represents a relationship between neighboring grid points in the domain. If the known solution values (represented by filled circles) and the unknown solution values (open circles) in this difference equation are drawn on a sample portion of the grid, we obtain what is known as a computational molecule, which makes the interpretation of the difference equation much easier:

From this diagram it is apparent that this method can be applied to pure initial-value problems as well as initial-boundary value problems. In the remainder of this chapter the computational molecule will be shown for every difference equation developed.

The implementation of this difference scheme is straightforward (see Figure 4.2). The index \( s \) denotes the stepping parameter, and \( i \) is the index of the spatial variable. In this scheme, there is only one unknown, \( u_{s+1,i} \). At \( s = 0 \), there is an initial condition and, therefore, one must start there. At \( i = 0 \) and \( i = n \) there are two boundary conditions that have to be taken into account. Starting at \( s = 0 \) and \( i = 0 \), the initial condition at \( s = 0 \) determines all the values \( u_{0,i}, i = 0,1,2,\ldots,n \). The values of \( u_{s,-1}, u_{s,0}, u_{s,n} \) and \( u_{s,n+1}, s = 1,2,\ldots \), are determined by
the boundary conditions there. While stepping through $i \forall s$, the solution, $u$, is found for each point on the grid.

The difference equation (4.21) has a truncation error $O(\Delta t)$ in time and $O[(\Delta x)^2]$ in $x$. In more compact form, this can be written as $O[\Delta t + (\Delta x)^2]$. Stability analyses have shown the method to be stable if and only if $r$, as defined by (4.20) satisfies $r \leq 1/2$ (Theorem 9.4, DuChateau and Zachmann, 1986:127; Lapidus and Pinder, 1982:155). Therefore, the method is said to be conditionally stable.

This conditional stability is undesirable in practice because one may encounter problems where this condition is either computationally expensive or physically impossible to meet. One way to solve this problem is to look at implicit methods.

The Backward-Difference Implicit Method:

In the backward-difference implicit method the spatial difference is calculated at the $s+1$ level where the solution is not known:

$$\frac{u_{s+1,i} - u_{s,i}}{\Delta t} = b_1 b_2 u_{s+1,i} = b_1 \frac{u_{s+1,i+1} - 2u_{s+1,i} + u_{s+1,i-1}}{(\Delta x)^2}.$$  

(4.22)

Computational molecule:

From the difference equation and the computational molecule, it is apparent that there are three unknown (and one known) solution values in this scheme. To implement this scheme, one must start at $s = 0$ and $i = 0$ (see Figure 4.3). The initial condition at $s = 0$ determines the value of $u_{0,0}$ and all the values of $u_{0,i}$, $i = 0, 1, 2, \ldots, n$. The values of $u_{1,-1}$, $u_{1,0}$ and $u_{1,n+1}$, $u_{1,n}$ are
determined by the boundary conditions at the left and right boundaries. While stepping up in
$i$ at $s = 0$, a system of linear equations,

$$Au = X \quad (4.23)$$

is compiled. The vector $X$ is determined by the known part of equation (4.22), i.e., those
terms containing $u_{0,i}$, $i = 1, 2, \ldots, n$, in this case. The vector $u$ contains all the unknowns,
$u_{1,i}, u_{1,2}, \ldots, u_{1,n-1}$, and the coefficient matrix, $A = [a_{i,j}]$, is a matrix with only the elements
$a_{i,i}$, $a_{i-1,i}$ and $a_{i,i-1}$ ($i = 1, 2, \ldots, n$) non-zero. Since this represents only the diagonal of $A$
and the diagonal rows immediately above and below the diagonal, this matrix is said to be a
tri-diagonal matrix.

The system (4.23) must be solved to obtain the solution, $u$, at the $s = 1$ level. The number of
calculations required to do this is the same as that required to invert $A$ and calculate

$$u = A^{-1}X.$$  

This operation must be repeated for each time level until the solution is known on the entire
domain.

There exist a number of efficient algorithms that can invert $A$ and calculate $u$, like the Jacobi
(Lapidus and Pinder, 1982:395-397) and Gauss-Seidel iterative methods (Lapidus and Pinder,
1982:397-399). The most popular (and accurate) method, however, is the so-called Thomas
algorithm (Lapidus and Pinder, 1982:216) that can invert these tri-diagonal matrices with a
so-called L-U decomposition technique, as will be shown in Section 4.4.2.

This method has a local truncation error $O[\Delta t + (\Delta x)^2]$, similar to the explicit method, but
it is unconditionally stable (Theorem 9.5, DuChateau and Zachmann, 1986:127; Lapidus and
Pinder, 1982:159). A major drawback of this and the previous method is the fact that they are
only accurate to first order in $t$. It can be applied to initial-boundary value problems but not
to pure initial value problems (DuChateau and Zachmann, 1986:128).

The Crank-Nicolson Implicit Method:

One way to improve the accuracy in $t$ is to take a weighted average of the previous two methods
by doing the operation $\lambda(4.22) + (1 - \lambda)(4.21)$:

$$\frac{u_{s+1,i} - u_{s,i}}{\Delta t} = b_1[\lambda\delta_x^2 u_{s+1,i} + (1 - \lambda)\delta_x^2 u_{s,i}]. \quad (4.24)$$
Computational molecule:

\[
\begin{align*}
& s + 1, i - 1 \quad s + 1, i \quad s + 1, i + 1 \\
& s, i - 1 \quad s, i \quad s, i + 1
\end{align*}
\]

If \( \lambda = 1 \), the Backward-Difference Implicit method is obtained, and if \( \lambda = 0 \), the Forward-Difference Explicit method. If \( \lambda \in [1/2, 1] \) the method is unconditionally stable. The optimum accuracy is obtained if \( \lambda = 1/2 \) with the accuracy steadily decreasing as \( \lambda \) rises towards \( \lambda = 1 \). The case of \( \lambda = 1/2 \) is called the Crank-Nicolson Implicit method (sometimes called the time centered implicit method):

\[
\frac{u_{s+1,i} - u_{si}}{\Delta t} = \frac{b_1}{2} \delta_x^2 (u_{s+1,i} + u_{si}).
\]  

(4.25)

This method is unconditionally stable with a local truncation error of \( O((\Delta t)^2 + (\Delta x)^2) \), i.e., accurate to second order in both the stepping parameter and the spatial variable (DuChateau and Zachmann, 1986:127; Lapidus and Pinder, 1982:160). This second order accuracy, however, only holds for the model parabolic PDE with \( b_1 = 1 \) (Lapidus and Pinder, 1982:160).

### 4.4.2 The Thomas Algorithm

Both the backward-difference implicit and the Crank-Nicolson method require the solution of tri-diagonal systems of equations. The so-called Thomas algorithm is deemed to be the most effective method with which this can be accomplished.

Consider a system of equations

\[
Ty = r
\]  

(4.26)

with a tri-diagonal coefficient matrix

\[
T = 
\begin{bmatrix}
  d_1 &  &  \\
  c_2 & d_2 &  \\
  & \ddots & \ddots & \ddots  \\
  &  & c_{n-1} & d_{n-1} & e_{n-1}  \\
  &  &  & c_n & d_n
\end{bmatrix},
\]  

(4.27)

solution vector

\[
y = 
\begin{bmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_n
\end{bmatrix},
\]

and answer vector

\[
r = 
\begin{bmatrix}
  r_1 \\
  r_2 \\
  \vdots \\
  r_n
\end{bmatrix},
\]
The system (4.26) can be solved if the matrix, \( T \), is invertible. Since \( T \) is a tri-diagonal matrix, the solving of equation (4.26) via matrix factoring is relatively simple. This method is sometimes called the Thomas algorithm. In order for (4.26) to be solvable, \( T \) must be a square \((n \times n)\) matrix.

Assume that \( T \) can be written as

\[
T = LU
\]

where \( L \) is left triangular and \( U \) right triangular. Since \( T \) is a tri-diagonal matrix, \( L \) and \( U \) will only have two diagonals each:

\[
L = \begin{bmatrix}
1 & & & \\
f_2 & 1 & & \\
& \ddots & \ddots & \\
& & f_{n-1} & 1 \\
& & & f_n & 1
\end{bmatrix}
\]

and

\[
U = \begin{bmatrix}
g_1 & e_1 & & & \\
g_2 & e_2 & & & \\
& \ddots & \ddots & & \\
& & \ddots & \ddots & \\
g_{n-1} & e_{n-1} & & & g_n
\end{bmatrix}
\]

With this information, we can write the product

\[
LU = \begin{bmatrix}
g_1 & e_1 & & & \\
f_2g_1 & f_2e_1 + g_2 & e_2 & & \\
& f_3g_2 & f_3e_2 + g_3 & e_3 & & \\
& & \ddots & \ddots & \ddots & \\
& & & f_ng_{n-1} & f_ne_{n-1} + g_n
\end{bmatrix}
\]

and compare it with \( T \) [equation (4.27)]. This comparison yields

\[
g_1 = d_1
\]

\[
f_ig_{i-1} = c_i, \quad i = 2,3,\ldots,n;
\]

\[
f_i \varepsilon_{i-1} + g_i = d_i, \quad i = 2,3,\ldots,n.
\]

Equations (4.31) and (4.32) can be recursively solved for \( f_i \) and \( g_i \):

\[
f_i = c_i / g_{i-1}, \quad i = 2,3,\ldots,n;
\]

\[
g_i = d_i - f_i \varepsilon_{i-1}, \quad i = 2,3,\ldots,n.
\]

Now we know \( L \), and \( U \) and \( Ty = r \) can be solved in two steps, namely

\[
Lz = r,
\]

and

\[
Uy = z.
\]

If equations (4.35) and (4.36) are written out it is found that

\[
z_1 = r_1
\]

\[
f_iz_{i-1} + z_i = r_i, \quad i = 2,3,\ldots,n
\]
The equations (4.37) to (4.40) can be recursively solved for $z_i$ and $y_i$ as follows:

\[
\begin{aligned}
    z_1 &= r_i; \\
    z_i &= r_i - f_i z_{i-1}, \quad i = 2, 3, \ldots, n; \\
    y_n &= z_n/g_n; \\
    y_i &= \frac{z_i - e_i y_{i+1}}{g_i}, \quad i = n - 1, n - 2, \ldots, 1.
\end{aligned}
\]  

In short: If the vectors $c, d, e$ and $r$ are known, then the vectors $g$ and $f$ can be calculated with equations (4.30), (4.33) and (4.34) and the vector $z$ can be calculated with equations (4.41) and (4.42). With $g, f$ and $z$ known, the solution vector $y$ can be calculated with equations (4.43) and (4.44).

### 4.4.3 Two-Dimensional Equations

All three of the previously discussed methods may be expanded into more than one dimension. Here, these three methods will be shown for two spatial dimensions. In the case when $n = 2$, $x_1 = x$, $x_2 = y$, and $a = c_i = d = e = 0$. Thus, equation (4.18) reduces to the following model PDE:

\[
    u_t = b_1 u_{xx} + b_2 u_{yy}.
\]  

The previous two-dimensional grid now becomes a three-dimensional grid: $(t_s, x_i, y_j) = (s \Delta t, i \Delta x, j \Delta y)$. As before, the variable $t$ is treated as the stepping parameter.

**The Forward-Difference Explicit Method:**

With expansion to two spatial dimensions, the forward-difference explicit method of Section 4.4.1 (equation 4.21) becomes

\[
    \frac{u_{s+1,i,j} - u_{s,i,j}}{\Delta t} = b_1 u_{s,i,j} + b_2 u_{s,i,j}.
\]  

Computational molecule:

\[
\begin{aligned}
    &s + 1, i, j \\
    &s, i - 1, j \quad s, i, j - 1 \quad s, i + 1, j \quad s, i, j + 1
\end{aligned}
\]

This method has a local truncation error of $O[\Delta t + (\Delta x)^2 + (\Delta y)^2]$ and is stable if $\Delta t [b_1 / (\Delta x)^2 + b_2 / (\Delta y)^2] \leq 1/4$ (Theorem 9.7, DuChateau and Zachmann, 1986:128; Lapidus and Pinder, 1982:234).
The Backward-Difference Implicit Method:

With expansion to two spatial dimensions, the backward-difference implicit method of Section 4.4.1 (equation 4.22) becomes

\[
\frac{u_{s+1,i,j} - u_{s,i,j}}{\Delta t} = \frac{b_1}{2} \Delta x^2 (u_{s+1,i,j} + u_{s,i,j}) + \frac{b_2}{2} \Delta y^2 (u_{s+1,i,j} + u_{s,i,j}).
\]  

(4.47)

Computational molecule:

This method has a local truncation error \(O[\Delta t + (\Delta x)^2 + (\Delta y)^2]\) but it is unconditionally stable (Theorem 9.8, DuChateau and Zachmann, 1986:128; Lapidus and Pinder, 1982:242).

The Crank-Nicolson Implicit Method:

As in the one-dimensional case (Section 4.4.1, equation 4.25), the Crank-Nicolson method is produced by a weighted average of equations (4.46) and (4.47) with a weight \(\lambda = 1/2\):

\[
\frac{u_{s+1,i,j} - u_{s,i,j}}{\Delta t} = \frac{b_1}{2} \Delta x^2 (u_{s+1,i,j} + u_{s,i,j}) + \frac{b_2}{2} \Delta y^2 (u_{s+1,i,j} + u_{s,i,j}).
\]  

(4.48)

Computational molecule:

This method has a local truncation error of \(O[(\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2]\) and is unconditionally stable (Theorem 9.9, DuChateau and Zachmann, 1986:128; Lapidus and Pinder, 1982:242).

For the two previously discussed two-dimensional implicit methods, equations (4.47) and (4.48), the matrices to be inverted when solving the systems of equations,

\[Au = X\]
are of penta-diagonal form and cannot be inverted by the quick Thomas algorithm. The Jacobi and Gauss-Seidel methods can be used to solve them, but they are very time-consuming in inverting matrices with more than three diagonals. There do, of course, exist sophisticated methods like the conjugate gradient method and the more modern preconditioned conjugate gradient method. These methods, however, require the storage of a large matrix with as many columns and rows as the product of all the dimensions in the domain on which the PDE must be solved. This requires computers with huge amounts of RAM.

It was this problem that motivated a group of mathematicians to develop other ways to discretize and solve equation (4.45).

The Alternating Direction Implicit Method:

Alternating Direction Implicit (ADI) methods are extensions of the Crank-Nicolson implicit method that preserve the tri-diagonal nature of the matrices to be inverted, by employing an operator splitting technique that produces two or more (depending on the number of spatial dimensions) separate, but interdependent, difference equations. With these equations one can solve sequences of equations in each of the spatial dimensions. The simplest of these ADI splitting schemes is the so called Peaceman-Rachford ADI method that produces two difference equations

\[
\frac{u_{s+1,i,j}^* - u_{s,i,j}}{\Delta t} = \frac{b_1}{2} \sigma_x^2 (u_{s+1,i+1,j}^* + u_{s+1,i,j}) + \frac{b_2}{2} \sigma_y^2 (u_{s+1,i,j+1} + u_{s+1,i,j})
\]

and

\[
\frac{u_{s+1,i,j} - u_{s,i,j}}{\Delta t} = \frac{b_1}{2} \sigma_x^2 (u_{s+1,i,j}^* + u_{s,i,j}) + \frac{b_2}{2} \sigma_y^2 (u_{s+1,i,j+1} + u_{s,i,j}).
\]

Computational molecule:

To implement this method one uses the first equation to solve implicitly in \(x\) for all \(y\) in order to obtain \(u_{s+1,i,j}^*\) from \(u_{s+1,i,j}\), and then the second equation to solve implicitly in \(y\) for all \(x\) in order to obtain \(u_{s+1,i,j}^*\) from \(u_{s+1,i,j}\) and \(u_{s+1,i,j}^*\) for each time step.

This method has a local truncation error \(O(\Delta t^2 + (\Delta x)^2 + (\Delta y)^2)\) and is unconditionally stable (Theorem 9.10; DuChateau and Zachmann, 1986:128; Lapidus and Pinder, 1982:249). Unfortunately, the expansion to higher dimensions (three for instance) will not be unconditionally stable (Lapidus and Pinder, 1982:248).

The Peaceman-Rachford scheme is but one of a family of ADI algorithms, with names like Douglas, Douglas-Rachford, Mitchell-Fairweather. Among these, there exist methods that can produce solutions of higher-order accuracy (like the Mitchell-Fairweather method).
The Locally One-Dimensional (LOD) Method:

The LOD and fractional splitting methods (developed in the former ‘Eastern-bloc’ countries at the same time as the family of ADI methods was developed in the West) represent another approach to keep the matrices that have to be inverted tri-diagonal.

Fractional splitting methods are based on the fact that the differential operator in any parabolic PDE is a linear operator. For equation (4.45) this linear differential operator is

$$\mathcal{L} = b_1 \frac{\partial^2}{\partial x^2} + b_2 \frac{\partial^2}{\partial y^2},$$

which is explicitly independent of \( t \). Therefore, the model equation may be written in terms of this linear differential operator (a linear transformation of the solution \( u \)):

$$u_t = \mathcal{L}u.$$

Since \( \mathcal{L} \) is a linear operator, it can be written as a linear combination of two other linear operators, \( \mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2 \), with \( \mathcal{L}_1 = b_1 \frac{\partial^2}{\partial x^2} \) and \( \mathcal{L}_2 = b_2 \frac{\partial^2}{\partial y^2} \) in the case of equation (4.45).

With an expansion to more spatial dimensions the situation stays the same, and the linear differential operator may be written as \( \mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2 + \cdots + \mathcal{L}_n \), where \( n \) is the number of spatial dimensions (or physical processes). Now, a difference scheme is developed which replaces \( \mathcal{L} \) with \( n\mathcal{L}_1, n\mathcal{L}_2, \ldots, n\mathcal{L}_n \), each operating for an interval \( \Delta t/n \). In other words, the PDE may be split into a set of one-dimensional PDEs which can easily be solved by simpler one-dimensional methods. As may be anticipated, a PDE may be split in various ways. Some fractional splitting methods may even result in identical composite expressions that have been derived in the family of ADI approaches (see Lapidus and Pinder, 1982:258 for a comparison of the LOD and Peaceman-Rachford ADI algorithm).

A special case of the fractional splitting technique is the case where the PDE is split into a set of parabolic one-dimensional equations by splitting with respect to its spatial variables (LOD). In this case equation (4.45) becomes

$$\frac{1}{2} u_t = b_1 u_{xx}$$

and

$$\frac{1}{2} u_t = b_2 u_{yy},$$

which can be solved individually with the Crank-Nicolson [equation (4.25)] method, which is attractive because of its unconditional stability and second-order accuracy. This gives two discretized equations

$$\frac{1}{2} \frac{u_{s+\frac{1}{2},i,j} - u_{s,i,j}}{\Delta t/2} = \frac{b_1}{2} \sigma_x^2 (u_{s+\frac{1}{2},i,j} + u_{s,i,j})$$

and

$$\frac{1}{2} \frac{u_{s+1,i,j} - u_{s+\frac{1}{2},i,j}}{\Delta t/2} = \frac{b_2}{2} \sigma_y^2 (u_{s+1,i,j} + u_{s+\frac{1}{2},i,j}).$$

Computational molecules:
The first equation is valid on the time interval $s \Delta t \leq t < (s + 1/2) \Delta t$, and the second on $(s + 1/2) \Delta t \leq t < (s + 1) \Delta t$.

This method has a local truncation error $O((\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2)$ and is unconditionally stable (Lapidus and Pinder, 1982:257). Unfortunately, the LOD inherits all the problems of the Crank-Nicolson and is therefore only accurate to second order in time for the model equation with $b_1 = b_2 = 1$. The LOD scheme implements in a way similar to the ADI, by first using the first equation to calculate an intermediate solution and then using the second to calculate a final solution for each time step. The major difference between the ADI and the LOD lies in the fact that the ADI calculates the final solution from the intermediate solution as well as the solution from the previous time step.

With the exception of the ADI, the extension of all the previous discretization algorithms to three and more dimensions are almost trivial and will not be discussed for a model equation in this thesis. A three-dimensional implementation of the LOD method will be discussed in detail when the TPE is discretized. It should be noted that, unlike the Peaceman-Rachford ADI algorithm, the LOD stays unconditionally stable with extension to higher dimensions.

### 4.4.4 Stability Considerations

In the previous subsections the conditions for stability were stated for each of the methods. This feature of numerical analysis has basically nothing to do with the PDE, but rather concerns the growth (unstable) or decay (stable) of errors in the arithmetic operations needed to solve the finite difference approximations.

Although there is an exact solution for a given difference approximation, errors are committed as the necessary calculations are carried out. Whether these errors amplify or decay characterizes the stability property of a given difference method.

The computational errors inside a computer are due to two machine-dependent parameters, called *machine precision* and *machine epsilon*. The first determines the number of significant digits used in a floating point operation in the floating point processor of the computer, while the second determines the largest number that can be added to any floating point number, such that the number stays unchanged. Unfortunately, these two parameters are machine-dependent. Some modern high level (fourth generation) programming languages like Mathematica have these two parameters fixed at software level in order to nullify this limitation. The only problem with this approach is that the overall runtime of a numerical model written in such a language will be much longer (up to ten times) than a similar model written in a third generation language like FORTRAN, C/C++ and Pascal.
There exist two popular ways to characterize the stability of a numerical scheme. The first and simplest technique is the Heuristic Stability analysis. This method is based upon a simple 'try-it-out' philosophy. An isolated error is introduced in one number and the behavior and propagation of this error is observed. This method, however, does not provide any information (unless an extensive empirical search is made) about the possible bounds on the stability.

The second and, perhaps, the most widely used stability condition is the von Neumann stability analysis. In this approach an initial line of errors (represented by a finite Fourier series) is introduced and the growth or decay of these errors is considered as the stepping parameter increases. A difference method for an initial-boundary value problem with a bounded solution is, therefore, called von Neumann stable if every solution $D[u_s] = 0$ of the form

$$u_{s,i} = \xi e^{\sqrt{-1} \beta s},$$

with $\beta$ real and $\xi = \xi(\beta)$ complex, has the property $|\xi| \leq 1$. For a problem with an unbounded solution this criterion becomes $|\xi| \leq 1 - O(\Delta t)$ (DuChateau and Zachmann, 1986:126).

This method applies strictly only to pure initial value problems with periodic initial data and it completely neglects the effects of boundary conditions. In practice, however, the results of von Neumann stability analyses are frequently used as a guideline for initial-boundary value problems. The results of a von Neumann stability analysis can, therefore, only be applied to linear PDEs with constant coefficients (Lapidus and Pinder, 1982:1170).

Therefore, these stability analyses are highly idealized due to their basic assumptions, and are only valid for unrealistically simple model equations with constant coefficients. Thus, the stability conditions stated previously are only rough guidelines, which may not successfully predict a given numerical scheme's stability for a realistic problem.

### 4.5 Methods for Solving First-Order Hyperbolic PDEs

It will be shown in Section 4.8 that the numerical solution of the TPE will require the discretization and solution of a first-order hyperbolic equation. Therefore, it is also necessary to be able to solve these first-order hyperbolic equations.

Consider the following model equation of first-order hyperbolic form (see Section 4.2.2)

$$a u_x + u_t = f$$  \hspace{1cm} (4.49)

where $f = f(x,t)$ and $a$ is a constant. Consider the grid $(t,s,x) = (s\Delta t, i\Delta x)$ and $r = \Delta t/\Delta x$.

Now we may write down the following methods to solve equation (4.49) numerically:

**The Forward-in-x Explicit Method:**

If the differentiations with respect to $x$ and $t$ are both approximated with first-order forward differences, we obtain the so-called forward-in-x explicit method:

$$a \frac{u_{s,i+1} - u_{s,i}}{\Delta x} + \frac{u_{s+1,i} - u_{s,i}}{\Delta t} = f_{s,i}.$$  \hspace{1cm} (4.50)

Computational molecule:
This method has a local truncation error $O(\Delta t + \Delta x)$ and is stable if and only if $-1 \leq ra \leq 0$ (Theorem 10.4, DuChateau and Zachmann, 1986:146).

The Backward-in-$x$ Explicit Method:

If the differentiation with respect to $x$ in the previous method is approximated by a first-order backward difference instead of a forward difference, we have the so-called backward-in-$x$ explicit method:

$$a \frac{u_{s+1,i} - u_{s,i}}{\Delta x} - \frac{u_{s+1,i} - u_{s,i}}{\Delta t} = f_{s,i}$$

(4.51)

Computational molecule:

This method has a local truncation error $O(\Delta t + \Delta x)$ and is stable if and only if $0 \leq ra \leq 1$ (Theorem 10.5, DuChateau and Zachmann, 1986:146).

The Modified Centered-in-$x$ Explicit Method:

To improve accuracy, one may combine the previous methods [(4.50) + (4.51)] and replace $u_{s,i}$ with the average between $u_{s,i+1}$ and $u_{s,i-1}, (u_{s,i+1} + u_{s,i-1})/2$, to obtain the modified centered-in-$x$ explicit method

$$a \frac{u_{s,i+1} - u_{s,i-1}}{2\Delta x} + \frac{u_{s+1,i} - \frac{1}{2}(u_{s,i+1} + u_{s,i-1})}{\Delta t} = f_{s,i}$$

(4.52)

Computational molecule:
This method has a local truncation error $O[\Delta t + (\Delta x)^2]$ and is stable if and only if $|\sigma| \leq 1$ (Theorem 10.6, DuChateau and Zachmann, 1986:146).

It should be noted that an unmodified centered-in-x method,

$$a \frac{u_{s,i+1} - u_{s,i-1}}{2\Delta x} + \frac{u_{s+1,i} - u_{s,i}}{\Delta t} = f_{s,i}$$

with a simple forward difference in time ($u_{s,i}$ not replaced by the average between $u_{s,i-1}$ and $u_{s,i+1}$) is always unconditionally unstable.

**Wendroff’s Implicit Method:**

To improve accuracy and stability, one may design a two-level implicit method by approximating both derivatives as averages of two forward differences. For the differentiation with respect to $x$, the derivative is approximated by the average between a forward difference at the $s$ time level and the $s+1$ level. For the differentiation with respect to $t$ the derivative is approximated by the average between a forward difference at the $i$ $x$-level and the $i+1$ level:

$$a \left( \frac{u_{s,i+1} - u_{s,i}}{2\Delta x} + \frac{u_{s+1,i} - u_{s,i}}{\Delta t} \right) + \frac{(u_{s+1,i+1} - u_{s+1,i}) + (u_{s+1,i+1} - u_{s,i+1})}{2\Delta t} = f_{s+\frac{1}{2},i+\frac{1}{2}}$$

**Computational molecule:**

![Diagram](image)

This method has a local truncation error $O[(\Delta t)^2 + (\Delta x)^2]$ and is unconditionally stable (DuChateau and Zachmann, 1986:147). Wendroff’s implicit method cannot be applied to an initial-boundary value problem. For a pure initial value problem it can be used in an explicit way, eliminating the need for matrix inversions.

### 4.6 Existing Solutions of the Transport Equation

This section contains a summary of existing solution techniques of the transport equation, which were studied to determine their characteristics and to decide on the most optimal method for the present solution.

For brevity the following notation will be used henceforth: If a function value is referenced in a difference equation, only the operative indices will be shown. If the non-operative indices are not shown it must be assumed to be in their simplest form. For example, if $f_{j+1} - f_j$ is referenced the reader must assume this is the same as $f_{s,i,j+1,k} - f_{s,i,j,k}$.

The $\delta_x^2$ and $\delta_x$ operators in Section 4.4 will be used such that $\delta_x^2 f_i$ yields equation (4.11) and $\delta_x f_i$ equation (4.10).
4.6.1 Steady-State Non-Acceleration Solutions

Fisk’s 1D Crank-Nicolson Discretization

Fisk (1971) was the first to solve the Parker TPE using numerical techniques. He solved the steady-state, spherically symmetric TPE,

$$a_o \frac{\partial^2 f}{\partial r^2} + c_o \frac{\partial f}{\partial r} + e_o \frac{\partial f}{\partial \ln P} = 0,$$

with the aid of an one-dimensional Crank-Nicolson algorithm. He specified a deduced energy spectrum at the highest rigidity as an initial condition, and used the rigidity variable, $\ln P$, as a stepping parameter. In this initial-boundary value problem the stepping parameter goes downward form the initial condition at the highest energy.

The basic Crank-Nicolson algorithm, however, requires the stepping parameter to go upward, away from the initial condition (see the computational molecule of Section 4.4.1). To be able to step downwards in rigidity from an initial condition at the upper boundary in energy, requires the computational molecule to be inverted. This can be accomplished either by using a backward difference in rigidity rather than a standard forward difference, or the standard forward difference must be multiplied by $-1$. The latter approach yields the discretization scheme that Fisk used:

$$e_o \frac{f_{k+1} - f_k}{\Delta \ln P} = \frac{a_o}{2} \delta_r^2 (f_{k+1} - f_k) + \frac{c_o}{2} \delta_r (f_{k+1} + f_k) + \frac{e_o}{2} \delta_{\ln P} (f_{k+1} + f_k).$$

Computational molecule:

```
   k, i-1  k, i  k, i+1
     \   /    /    /
      \ /    /    /
       \ /    /    /
        \ /    /    \\
         k+1, i-1 k+1, i k+1, i+1
```

Fisk’s 2D ADI Discretization

Fisk (1973) expanded his one-dimensional spherically symmetric model into a two-dimensional model by using a Peaceman-Rachford algorithm to discretize the steady-state two-dimensional TPE

$$a_o \frac{\partial^2 f}{\partial r^2} + b_o \frac{\partial^2 f}{\partial \theta^2} + c_o \frac{\partial f}{\partial r} + d_o \frac{\partial f}{\partial \theta} + e_o \frac{\partial f}{\partial \ln P} = 0.$$

The discretization is as follows:

$$\frac{f_{k+1}^* - f_k}{\Delta \ln P} = \frac{a_o}{2} \delta_r^2 (f_{k+1}^* + f_k) + \frac{c_o}{2} \delta_r (f_{k+1}^* + f_k) + b_o \delta_\theta^2 f_k + d_o \delta_\theta f_k$$

and

$$\frac{f_{k+1} - f_k}{\Delta \ln P} = \frac{a_o}{2} \delta_r^2 (f_{k+1} + f_k) + \frac{c_o}{2} \delta_r (f_{k+1} + f_k) + \frac{b_o}{2} \delta_\theta^2 (f_{k+1} + f_k) + \frac{d_o}{2} \delta_\theta (f_{k+1} + f_k).$$
To simplify the second equation, the first equation may be subtracted from it, yielding a new, equivalent form of the second equation:

\[
\frac{f_{k+1} - f_k}{\Delta \ln P} = \frac{b_0}{2} \delta_\theta (f_{k+1} - f_k) + \frac{d_0}{2} \delta_\phi (f_{k+1} - f_k).
\] (4.55)

The computational molecules for the two discretization equations are

respectively. Moraal and Gleeson (1975) extended and improved on Fisk’s initial model by adding more physics to the coefficients, but the basic discretization stayed the same. These results were presented at the 14th ICRC at the same time that Cecchini and Quenby (1975) presented an independently developed ADI model.

In 1979 Moraal et al. again used their 1975 discretization to present the first drift calculations at the 16th ICRC, at the same time that Jokipii and Kopriva (1979a) presented their separately developed drift model. In 1983 Nagashima and Munakata published an independently developed drift model. In this paper, however, they erroneously added \( E \times B \) drift even though it is already present in the energy change term of the TPE as was shown by Forman and Gleeson (1975).

These methods did not involve any deviation from Fisk’s initial discretization, since drift is added to the code by modifying the coefficients, \( c_0 \) and \( d_0 \).

Kóta & Jokipii’s 3D Discretization

Kóta and Jokipii (1983) presented the first discretization of the full three-dimensional steady-state TPE, which they wrote in the form:

\[
\frac{\partial f}{\partial \ln P} = \gamma_{rr} \frac{\partial}{\partial r} (\alpha_{rr}) + 2 \gamma_{r\phi} \frac{\partial^2 f}{\partial r \partial \phi} + \gamma_{\phi\phi} \frac{\partial^2 f}{\partial \phi^2} + \gamma_{\theta\theta} \frac{\partial f}{\partial \theta} (\alpha_{\phi\theta} \theta') + \beta_r \frac{\partial f}{\partial r} + \beta_\phi \frac{\partial f}{\partial \phi} + \beta_\theta \frac{\partial f}{\partial \theta}.
\]
Most notable in this equation is the existence of the $\partial^2 f/\partial r \partial \phi$ term which presents a new, but simply solved, problem. They discretized the TPE as follows:

$$\frac{f_{k+\gamma_3} - f_k}{\Delta \ln P} = \gamma_{rr} \frac{\delta}{\delta r} \left( \alpha_{rr} \frac{\delta}{\delta r} f_{k+\gamma_3} \right) + \beta_r \delta_r f_{k+\gamma_3} + \gamma_{r\phi} \frac{\delta^2}{\delta r \delta \phi} f_k,$$

$$\frac{f_{k+\gamma_3} - f_{k+\gamma_3}}{\Delta \ln P} = \gamma_{\theta \phi} \frac{\delta^2}{\delta \theta \delta \phi} f_{k+\gamma_3} + \beta_{\phi} \delta_\phi f_{k+\gamma_3} + \gamma_{\phi \phi} \frac{\delta^2}{\delta \phi \delta \phi} f_{k+\gamma_3}$$

and

$$\frac{f_{k+1} - f_{k+\gamma_3}}{\Delta \ln P} = \gamma_{\theta \theta} \frac{\delta}{\delta \theta} \left( \alpha_{\theta \theta} \frac{\delta}{\delta \theta} f_{k+1} \right) + \beta_{\phi} \delta_\theta f_{k+1}$$

where

$$\frac{\delta}{\delta \theta} \left( \alpha \frac{\delta}{\delta \theta} f \right) = \frac{1}{\Delta r} \left( \alpha_{i+\gamma} f_{i+1} - f_i \Delta r - \alpha f_i - f_{i-1} \right)$$

which reduces to

$$\alpha_{i+\gamma} f_{i-1} - 2 f_i + f_{i+1} \frac{(\Delta r)^2}{\Delta \ln P} = \alpha \delta^2 f$$

if $\alpha_{i+\gamma} = \alpha_{i-\gamma} = \alpha$, and

$$\frac{\delta^2}{\delta r \delta \phi} = \frac{f_{i+1,q+1} - f_{i+1,q-1} - f_{i-1,q+1} + f_{i-1,q-1}}{2 \Delta r \Delta \phi}$$

with $k$ denoting the rigidity index, $i$ the radial index and $q$ the azimuthal index. No computational molecule is given, since it is difficult to represent a four-dimensional structure on a two-dimensional surface.

4.6.2 Steady-State Acceleration Solutions

Potgieter & Moraal's Crank-Nicolson-Based 1D Solution

Potgieter and Moraal (1988) developed a steady-state spherically symmetric model with which particles can be accelerated. They used the simple Crank-Nicolson discretization of the spherically symmetric steady state TPE, pioneered by Fisk (see Section 4.6.1) to discretize the TPE. The effect of the shock was added in the form of a Neumann boundary condition, derived from the spherically symmetric form of the requirement $S_\gamma = S_\gamma^+$ (see Section 3.3):

$$\frac{\partial f}{\partial r} = \frac{1}{3} \left( \frac{V}{\kappa} \right) - \left( 1 - \frac{1}{s} \right) \frac{\partial f^{(s)}}{\partial \ln P},$$

with $f^{(s)}$ the solution on the shock. This model worked commendably well, but an attempt by Moraal in 1989 to extend this model to two dimensions did not succeed, when he encountered serious and apparently insurmountable numerical instabilities.

4.6.3 The TPE Dilemma with Time-Dependent Solutions

The aim of this chapter is to design a numerical scheme with which the time-dependent transport or Parker equation (TPE) can be solved in two spatial dimensions. This equation is a PDE of parabolic form, and the family of discretization methods designed for parabolic PDEs should be applicable. However, if the TPE is written in the general form

$$f_t = \sum_{i=1}^{3} (b_if_{x_i} + c_if_{x_i}) + e$$

$$= b_1 f_{rr} + c_1 f_r + b_2 f_{\theta \theta} + c_2 f_{\theta} + c_3 f_{ln P} + e,$$

61
(where we have used $x_1 = r$, $x_2 = \theta$ and $x_3 = \ln P$) it is clear that $b_3 = 0$. Since all solution schemes were developed under the condition that $b_i > 0 \ \forall i$ in equation (4.18), the above-mentioned techniques developed to solve parabolic PDE (ADI, LOD) cannot be used in solving the TPE in this form.

In Section 3.2, equation (3.16), it has been shown that the TPE does indeed contain a term with a non-zero coefficient of $\partial^2 f/\partial \ln P^2$. The addition of this term seems to be able to solve this numerical problem. This, unfortunately, is not the case. The physics of the heliosphere is such that the coefficient $D_{pp}$ is much smaller than $(\nabla \cdot V)/3$, resulting in a coefficient of $\partial^2 f/\partial \ln p^2$ that is practically zero relative to the coefficient of $\partial f/\partial \ln p$. This results in the inevitable rise of numerical instabilities, rendering the TPE unsolvable.

One must, therefore, consider an alternative approach, or attempt to construct a hybrid method containing elements of the standard methods.

### 4.6.4 Time-Dependent Non-Acceleration Solutions

**Perko & Fisk's Modified Crank-Nicolson 1D Solution**

One way to get around this problem is to consider both time and rigidity as stepping parameters and step up in time and down in rigidity at the same time. This is the approach of Perko and Fisk (1983). They discretized the spherically symmetric time-dependent TPE,

$$\frac{\partial f}{\partial t} = a_0 \frac{\partial^2 f}{\partial r^2} + c_0 \frac{\partial f}{\partial r} + e_0 \frac{\partial d}{\partial \ln p},$$

as follows:

\[
\begin{align*}
\frac{1}{2} \left( f_{s+1,k} - f_{s,k} \right) + & \left( f_{s+1,k+1} - f_{s,k+1} \right) \\
\frac{\Delta t}{2} & + \frac{e_0}{2} \left( f_{s,k+1} - f_{s,k} \right) + \left( f_{s+1,k+1} - f_{s+1,k} \right) \\
& = \frac{a_0}{4} \delta_r^2 \left( f_{s,k} + f_{s,k+1} + f_{s+1,k} + f_{s+1,k+1} \right) \\
& = \frac{c_0}{4} \delta_r \left( f_{s,k} + f_{s,k+1} + f_{s+1,k} + f_{s+1,k+1} \right).
\end{align*}
\]

Computational molecule:

![Diagram](https://via.placeholder.com/150)

This is a good algorithm with which spherically symmetric time-dependent modulation can be studied. The problem with this algorithm, however, is that the energy changes can only take place in one direction, rendering it unsuitable to produce acceleration solutions.
4.6.5 Le Roux’s Modified 2D ADI Solution

Le Roux (1990) expanded Fisk’s approach to the TPE dilemma into two spatial dimensions. Le Roux discretized the TPE as follows, using a modified ADI algorithm:

\[
\frac{1}{2} \left( \frac{(f_{s+1,k} - f_{s,k}) + (f_{s+1,k+1}^* - f_{s,k+1})}{\Delta t} \right) + \frac{e_0}{2} \left( \frac{(f_{s,k+1}^* - f_{s,k}) + (f_{s+1,k+1}^* - f_{s+1,k})}{\Delta \ln P} \right)
\]

\[
= \frac{a_0}{4} \delta_r^2 (f_{s,k} + f_{s,k+1}^* + f_{s+1,k} + f_{s+1,k+1}^*)
+ \frac{c_0}{4} \delta_t^2 (f_{s,k} + f_{s,k+1}^* + f_{s+1,k} + f_{s+1,k+1}^*)
+ \frac{b_0}{2} \delta_\theta^2 (f_{s,k} + f_{s+1,k})
+ \frac{d_0}{2} \delta_\theta (f_{s,k} + f_{s+1,k})
\]

and

\[
\frac{1}{2} \left( \frac{(f_{s+1,k} - f_{s,k}) + (f_{s+1,k+1} - f_{s,k+1})}{\Delta t} \right) + \frac{e_0}{2} \left( \frac{(f_{s,k+1} - f_{s,k}) + (f_{s+1,k+1} - f_{s+1,k})}{\Delta \ln P} \right)
\]

\[
= \frac{a_0}{4} \delta_r^2 (f_{s,k} + f_{s,k+1}^* + f_{s+1,k} + f_{s+1,k+1}^*)
+ \frac{c_0}{4} \delta_t^2 (f_{s,k} + f_{s,k+1}^* + f_{s+1,k} + f_{s+1,k+1}^*)
+ \frac{b_0}{4} \delta_\theta^2 (f_{s,k} + f_{s,k+1} + f_{s+1,k} + f_{s+1,k+1})
+ \frac{d_0}{4} \delta_\theta (f_{s,k} + f_{s,k+1} + f_{s+1,k} + f_{s+1,k+1})
\]

Again, no computational molecule is shown for this method since it is difficult to represent a four-dimensional structure in two dimensions.

4.6.6 Time-Dependent Acceleration Solutions

Jokipii’s 2D ADI Solution

In 1986 Jokipii presented the first acceleration results, employing a two-dimensional time-dependent model. The precise discretization that he used is not known to us. To the best of our knowledge Jokipii never published it in detail. From Jokipii (1986:2930):

“The differencing scheme used was quite standard, in that a two-dimensional alternating direction implicit (ADI) technique was used for the spatial differencing. In order to avoid numerical instabilities, an explicit scheme was used in energy (or momentum) with one-sided differences used depending on the direction of energy change at the spatial location being considered, to ensure that the solution obeyed causality.”

The most that we could learn from his subsequent papers was that he used the matching condition (3.44) to accelerate particles at the shock.
Kóta & Jokipii's 3D Solution

Kóta and Jokipii (1991) published a paper in which they studied the effects of CIRs on the long term modulation of cosmic rays with a three-dimensional, time-dependent model. Particle acceleration at the termination shock, as well as particle acceleration at CIR shocks was included in their analysis. No substantial information was made available on the discretization method used, with only the general statement (Kóta and Jokipii, 1991:1798): “The cosmic-ray distribution function as a function of time, position, and momentum is computed using a modified ADI finite difference scheme.”

4.7 LOD Solution of the Time-Dependent TPE

From the above discussion it is clear that until now only the group at the University of Arizona had an acceleration model with which acceleration at the solar wind termination shock can be studied. At the Space Research Unit at the Potchefstroom University for CHE we needed such a model to study the anomalous component of cosmic rays. It was decided to develop an independent acceleration model to do this. The first attempt to develop a two-dimensional steady-state acceleration model failed when the one-dimensional Potgieter-Moraal model was expanded into two dimensions, as described in Section 4.6.2. It was thus decided to develop a time-dependent mode.

This section, as well as Section 4.8, describe the numerical solution of this thesis in detail.

The transport equation is of the general form

\[
\frac{\partial f}{\partial t} = a_0 \frac{\partial^2 f}{\partial r^2} + b_0 \frac{\partial^2 f}{\partial \theta^2} + c_0 \frac{\partial f}{\partial r} + d_0 \frac{\partial f}{\partial \theta} + e_0 \frac{\partial f}{\partial \ln P} + Q. \tag{4.56}
\]

According to the theory of the Locally One-Dimensional (LOD) method, a PDE may be split into a system of equations, each containing only derivatives in one dimension:

\[
\begin{align*}
\frac{1}{3} \frac{\partial f}{\partial t} &= a_0 \frac{\partial^2 f}{\partial r^2} + c_0 \frac{\partial f}{\partial r} \tag{4.57} \\
\frac{1}{3} \frac{\partial f}{\partial t} &= b_0 \frac{\partial^2 f}{\partial \theta^2} + d_0 \frac{\partial f}{\partial \theta} \tag{4.58} \\
\frac{1}{3} \frac{\partial f}{\partial t} &= e_0 \frac{\partial f}{\partial \ln P} + Q. \tag{4.59}
\end{align*}
\]

The theory states that each equation written down above is valid on a third of the time step, i.e., equation (4.57) is valid on the interval \( t' < t < t' + \Delta t/3 \), equation (4.58) on \( t' + \Delta t/3 < t < t' + 2\Delta t/3 \) and equation (4.59) on \( t' + 2\Delta t/3 < t < t' + \Delta t \).

Subsequently, equation (4.57) will be referred to as the radial equation, (4.58) as the polar equation and (4.59) as the energy equation.

The radial and polar equations are of parabolic form and may be solved with the Crank-Nicolson algorithm, but the energy equation is of first-order hyperbolic form and must be solved differently.

4.7.1 Solving the Radial Equation

The radial equation (4.57), in time-centered implicit form (Section 4.4.1), is

\[
\frac{1}{3} \frac{f_{s+\gamma_3} - f_s}{\Delta t/3} = a_0 \frac{\delta^2}{2} (f_{s+\gamma_3} + f_s) + c_0 \frac{\delta r}{2} (f_{s+\gamma_3} + f_s). \tag{4.60}
\]
In expanded form this becomes
\[
\frac{1}{\Delta t}(f_{s+\gamma_3 i,j,k} - f_{i,j,k}) = \frac{1}{2}(a^- f_{s+\gamma_3 i-1,j,k} + a f_{s+\gamma_3 i,j,k} + a^+ f_{s+\gamma_3 i+1,j,k}) \\
+ \frac{1}{2}(a^- f_{s,i-1,j,k} + a f_{s,i,j,k} + a^+ f_{s,i+1,j,k}) \\
+ \frac{1}{2}(c^- f_{s+\gamma_3 i-1,j,k} + c f_{s+\gamma_3 i,j,k} + c^+ f_{s+\gamma_3 i+1,j,k}) \\
+ \frac{1}{2}(c^- f_{s,i-1,j,k} + c f_{s,i,j,k} + c^+ f_{s,i+1,j,k}).
\]

After grouping all the grid points together it becomes
\[
\left[ \frac{a^- + c^-}{2} \right] f_{s+\gamma_3 i-1,j,k} + \left[ \frac{a + c}{2} - \frac{1}{\Delta t} \right] f_{s+\gamma_3 i,j,k} + \left[ \frac{a^+ + c^+}{2} \right] f_{s+\gamma_3 i+1,j,k} \\
- \left[ \frac{a^- + c^-}{2} \right] f_{s,i-1,j,k} - \left[ \frac{a + c}{2} + \frac{1}{\Delta t} \right] f_{s,i,j,k} + \left[ \frac{a^+ + c^+}{2} \right] f_{s,i+1,j,k},
\]
or
\[
D_{r1} f_{s+\gamma_3 i-1,j,k} + D_{r3} f_{s+\gamma_3 i,j,k} + D_{r2} f_{s+\gamma_3 i+1,j,k} \\
= -D_{r1} f_{s,i-1,j,k} - D_{r4} f_{s,i,j,k} - D_{r2} f_{s,i+1,j,k}.
\]

With the aid of equations (4.10) and (4.11)
\[
D_{r1} = \frac{a^- + c^-}{2} = \frac{1}{\Delta r_2(\Delta r_1 + \Delta r_2)} \left[ a_0 - \frac{\Delta r_1 c_0}{2} \right] \\
D_{r2} = \frac{a^+ + c^+}{2} = \frac{1}{\Delta r_1(\Delta r_1 + \Delta r_2)} \left[ a_0 + \frac{\Delta r_2 c_0}{2} \right] \\
D_{rI} = \frac{a + c}{2} = D_{r1} - D_{r2} \\
D_{r3} = \frac{a^- + c^- - \frac{1}{\Delta t}}{2} = D_{rI} - \frac{1}{\Delta t} \\
D_{r4} = \frac{a^+ + c^+ + \frac{1}{\Delta t}}{2} = D_{rI} + \frac{1}{\Delta t}.
\]
The computational molecule is:

To implement equation (4.61) one needs two boundary conditions. It is useful to write these boundary conditions in the most general form possible, to make changes in boundary conditions easier to implement. These general boundary conditions are written as
\[
f_{s+\gamma_3 0,j,k} = v_1 f_{s+\gamma_3 1,j,k} + v_2 f_{s+\gamma_3 2,j,k} + v_3,
\]
and
\[
f_{s+\gamma_3 n,j,k} = v_4 f_{s+\gamma_3 n-1,j,k} + v_5 f_{s+\gamma_3 n-2,j,k} + v_6.
\]
For example, if \( v_1 = v_2 = 0 \) and \( v_3 \neq 0 \) in (4.62) we have a Dirichlet condition (see Section 4.3.1). If \( v_1 = v_3 = 0 \) and \( v_2 \neq 0 \) we have an example of a Neumann condition. A Robbins condition can be obtained if both \( v_2 \) and \( v_3 \) are non-zero. A non-zero \( v_1 \) may be used to construct a more flexible derivative in a Neumann or Robbins condition. This general form, however, cannot be used to construct a Cauchy boundary condition.

The boundary condition (4.62) represents the inner radial boundary of the heliosphere which will be taken to be on the surface of the sun, and condition (4.63) represents the outer radial boundary.

Equation (4.61) is used to compile a set of linear equations for every \( s, j \) and \( k \)

\[
D_{r_1} f_{s+i_1} + D_{r_3} f_{s+i_3,j,k} + D_{r_2} f_{s+i_2,j,k} = E_{s,i,j,k},
\]

with

\[
E_{s,i,j,k} = -D_{r_1} f_{s+i_1,j,k} - D_{r_4} f_{s,i,j,k} - D_{r_2} f_{s,i+1,j,k}
\]

and \( i = 1, 2, \ldots, n - 1 \).

Using the boundary conditions (4.62) and (4.63), the first and last equations of this system become

\[
(D_{r_3} + v_1 D_{r_1}) f_{s+i_1,j,k} + (D_{r_2} - v_2 D_{r_1}) f_{s+i_2,j,k} = E_{s,1,j,k} - v_3 D_{r_1}
\]

and

\[
(D_{r_{n-1}} - v_3 D_{r_{n-1}}) f_{s+i_3,j,k} + (D_{r_3} - v_4 D_{r_{n-1}}) f_{s+i_3,j,k} = E_{s,n-1,j,k} - v_3 D_{r_{n-1}},
\]

respectively.

This forms a system of equations in the form

\[
A f_{s+i,j,k} = F_{s,j,k}
\]

with the coefficient matrix, \( A \), a tri-diagonal matrix of the form

\[
A = \begin{bmatrix}
A_{1,1} & A_{1,2} & & & & \\
A_{2,1} & A_{2,2} & A_{2,3} & & & \\
& \cdots & \cdots & \cdots & & \\
& & \cdots & \cdots & \cdots & \\
A_{n-2,n-3} & A_{n-2,n-2} & A_{n-2,n-1} & & & \\
A_{n-1,n-2} & A_{n-1,n-1} & & & &
\end{bmatrix},
\]

\[
f_{s+i,j,k} = \begin{bmatrix}
f_{s+i_1,j,k} \\
f_{s+i_2,j,k} \\
& \cdots & \\
f_{s+i_3,j,k} \\
& & \cdots \\
f_{s+i_{n-1},j,k}
\end{bmatrix}
\]

and

\[
F_{s,j,k} = \begin{bmatrix}
E_{s,1,j,k} - v_3 D_{r_1} \\
E_{s,2,j,k} \\
& \cdots & \\
E_{s,n-2,j,k} \\
E_{s,n-1,j,k} - v_3 D_{r_{n-1}}
\end{bmatrix}.
\]

66
In this system \( A_{i,j} = (D_{r_3} + v_1 D_{r_1,j} \), \( A_{1,2} = (D_{r_2} - v_2 D_{r_1,j}) \), \( A_{n-1,n-2} = (D_{r_1} + v_3 D_{r_2,n-1}) \), \( A_{n-1,n-1} = (D_{r_3} - v_4 D_{r_2,n-1}) \) and \( A_{2,1} = D_{r_2} \), \( A_{2,2} = D_{r_3} \), \( A_{2,3} = D_{r_2} \), etc.

The \( A \) can be inverted by a streamlined Thomas algorithm: let \( X_{i} = 0 \) for \( i \geq 2 \), \( X_{-1} = -v_2 \), \( X_0 = -v_1 \) and \( Y_0 = v_3 \). For all \( j \) and \( k \) calculate the following:

\[
Z_i = D_{r_3} - D_{r_1} X_{i-1}, \quad i = 1, 2, \ldots, n - 1
\]

\[
X_i = \frac{D_{r_2} - D_{r_1} X_{i-1}}{Z_i}, \quad i = 1, 2, \ldots, n - 1
\]

and

\[
Y_i = \frac{E_{s,i,j,k} - D_{r_1} Y_{i-1}}{Z_i}, \quad i = 1, 2, \ldots, n - 1
\]

Then calculate

\[
f_{s+\gamma_5, n-1, i, j, k} = \frac{Y_{n-1} - X_{n-1}(v_4 + v_3 Y_{n-2})}{1 + X_{n-1}(v_4 - v_3 Y_{n-2})}
\]

and

\[
f_{s+\gamma_5, i, j, k} = Y_i - X_i f_{s+\gamma_5, i+1, i, j, k} \quad i = n - 2, n - 3, \ldots, 1
\]

from which \( f_{s+\gamma_5, i, j, k} \) follows.

### 4.7.2 Solving the Polar Equation

The polar equation (4.58), in time-centered implicit, form is

\[
\frac{1}{3} (f_{s+\gamma_5} - f_{s+\gamma_5}) - \frac{b_0}{3} \delta^2 (f_{s+\gamma_5} + f_{s+\gamma_5}) + \frac{d_0}{3} \delta (f_{s+\gamma_5} + f_{s+\gamma_5})
\]

which becomes

\[
\frac{1}{\Delta t} (f_{s+\gamma_5, i, j, k} - f_{s+\gamma_5, i, j, k}) = \frac{1}{2} \left( b^- f_{s+\gamma_5, i, j, k} - b^+ f_{s+\gamma_5, i, j, k} \right)
\]

after expansion. Grouping the grid points yields

\[
\begin{bmatrix}
\frac{b^- + d^-}{2} & f_{s+\gamma_5, i, j, k} + \frac{b^+ + d^+}{2} & f_{s+\gamma_5, i, j, k} + \frac{b^- + d^-}{2} & f_{s+\gamma_5, i, j, k} + \frac{b^+ + d^+}{2} & f_{s+\gamma_5, i, j, k}
\end{bmatrix}
\]

or

\[
D_{\theta_1} f_{s+\gamma_5, i, j, k} + D_{\theta_3} f_{s+\gamma_5, i, j, k} + D_{\theta_2} f_{s+\gamma_5, i, j, k} = -D_{\theta_1} f_{s+\gamma_5, i, j, k} - D_{\theta_3} f_{s+\gamma_5, i, j, k} - D_{\theta_2} f_{s+\gamma_5, i, j, k}
\]

(4.66)
The computational molecule is:

\[
\begin{align*}
D_{\theta 1} &= \frac{b^- + d^-}{2} = \frac{1}{\Delta \theta_2 (\Delta \theta_1 + \Delta \theta_2)} \left[ b_0 - \frac{\Delta \theta_1 d_0}{2} \right] \\
D_{\theta 2} &= \frac{b^+ + d^+}{2} = \frac{1}{\Delta \theta_1 (\Delta \theta_1 + \Delta \theta_2)} \left[ b_0 + \frac{\Delta \theta_2 d_0}{2} \right] \\
D_{\theta 3} &= \frac{b + d}{2} = -D_{\theta 1} - D_{\theta 2} \\
D_{\theta 4} &= \frac{b + d}{2} = D_{\theta 1} + \frac{1}{\Delta t}
\end{align*}
\]

The discretization of the polar equation follows the same basic pattern as the discretization of the radial equation, with only two major differences: the time-step lies in the second third of the time-interval and the polar discretization is orthogonal to the radial discretization.

Again, as in the radial case, the implementation of equation (4.66) requires two distinct boundary conditions in general form:

\[
f_{s+\gamma_3,i,j-1,k} = w_1 f_{s+\gamma_3,i,j,k} + w_2 f_{s+\gamma_3,i,j+1,k} + w_3
\]

and

\[
f_{s+\gamma_3,i,m+1,k} = w_4 f_{s+\gamma_3,i,m,k} + w_5 f_{s+\gamma_3,i,m-1,k} + w_6.
\]

As for the radial case, equation (4.66) is used to compile a set of linear equations for every \( s, i \) and \( k \)

\[
D_{\theta 1} f_{s+\gamma_3,i,j-1,k} + D_{\theta 3} f_{s+\gamma_3,i,j,k} + D_{\theta 2} f_{s+\gamma_3,i,j+1,k} = E_{s+\gamma_3,i,j,k},
\]

with

\[
E_{s+\gamma_3,i,j,k} = -D_{\theta 1} f_{s+\gamma_3,i,j-1,k} - D_{\theta 3} f_{s+\gamma_3,i,j-1,k} - D_{\theta 2} f_{s+\gamma_3,i,j+1,k},
\]

for \( j = 1, 2, \ldots, m \).

Using the boundary conditions (4.67) and (4.68), the first and last equations of this system become

\[
(D_{\theta 3} + w_1 D_{\theta 1}) f_{s+\gamma_3,i,1,k} + (D_{\theta 1} - w_2 D_{\theta 1}) f_{s+\gamma_3,i,2,k}
\]

\[
= E_{s+\gamma_3,i,1,k} - w_3 D_{\theta 1}
\]

68
and
\[
(D \Theta_1 - w_5 D \Theta_2) f_{s+ \gamma_3, i, m-1, k} + (D \Theta_3 - w_4 D \Theta_2) f_{s+ \gamma_3, i, m, k} = E_{s+ \gamma_3, i, m, k} - w_3 D \Theta_1,
\]
respectively.

This, again, forms a system of equations in the form
\[
\mathbb{B} f_{s+ \gamma_3, i, k} = F_{s+ \gamma_3, i, k}
\]
with the coefficient matrix, \(\mathbb{B}\), a tri-diagonal matrix
\[
\mathbb{B} = \begin{bmatrix}
B_{1,1} & B_{1,2} & \cdots & B_{1,m-2} & B_{1,m-1} & B_{1,m} \\
B_{2,1} & B_{2,2} & \cdots & B_{2,m-2} & B_{2,m-1} & B_{2,m} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
B_{m-1,m-2} & B_{m-1,m-1} & \cdots & \cdots & \cdots & B_{m,m}
\end{bmatrix},
\]
and
\[
f_{s+ \gamma_3, i, k} = \begin{bmatrix}
f_{s+ \gamma_3, i, 1, k} \\
f_{s+ \gamma_3, i, 2, k} \\
\vdots \\
f_{s+ \gamma_3, i, m-1, k} \\
f_{s+ \gamma_3, i, m, k}
\end{bmatrix},
\]
and
\[
F_{s+ \gamma_3, i, k} = \begin{bmatrix}
E_{s+ \gamma_3, i, 1, k} - w_3 D \Theta_1 \\
E_{s+ \gamma_3, i, 2, k} \\
\vdots \\
E_{s+ \gamma_3, i, m-1, k} \\
E_{s+ \gamma_3, i, m, k} - w_3 D \Theta_1
\end{bmatrix}
\]
Here \(B_{1,1} = (D \Theta_3 + w_1 D \Theta_1)\), \(B_{1,2} = (D \Theta_3 - w_2 D \Theta_1)\), \(B_{m,m-1} = (D \Theta_1 - w_5 D \Theta_2)\), \(B_{m,m} = (D \Theta_3 - w_4 D \Theta_2)\) and \(B_{2,1} = D \Theta_1\), \(B_{2,2} = D \Theta_3\), \(B_{2,3} = D \Theta_2\), etc.

Matrix \(\mathbb{B}\) can be inverted by a streamlined Thomas algorithm: let \(X_{j-} = 0\) for \(j \geq 2\), \(X_{-1} = -w_2\), \(X_0 = -w_1\) and \(Y_0 = w_3\). For all \(i\) and \(k\) calculate the following:
\[
Z_j = D \Theta_3 - D \Theta_1 X_{j-1}, \quad j = 1, 2, \ldots, m
\]
\[
X_j = \frac{D \Theta_3 - D \Theta_1 X_{j-}}{Z_j}, \quad j = 1, 2, \ldots, m
\]
and
\[
Y_j = \frac{E_{s+ \gamma_3, i, j, k} - D \Theta_1 Y_{j-}}{Z_j}. \quad j = 1, 2, \ldots, m
\]
Then calculate
\[
f_{s+ \gamma_3, i, m, k} = \frac{Y_m - X_m (w_6 + w_8 Y_{m-1})}{1 + X_m (w_4 - w_5 X_{m-1})}
\]
and
\[
f_{s+ \gamma_3, i, j, k} = Y_j - X_j f_{s+ \gamma_3, i, j+1, k} \quad j = m - 1, m - 2, \ldots, 1
\]
from which \(f_{s+ \gamma_3, i, k}\) follows.
4.7.3 Solving the Energy Equation

Write the energy equation (4.59) in the form

\[
\frac{1}{3} \frac{\partial f}{\partial t} - e_0 \frac{\partial f}{\partial \ln P} = Q. \tag{4.70}
\]

The solution, \( f \), of this first order hyperbolic PDE is constant along a set of characteristic curves (characteristics) in \((P-t-f)\)-space. To obtain an expression for these characteristics, we follow the approach of Lapidus and Pinder, (1982:5). If \( f = f(t, \ln P) \) then

\[
df = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial \ln P} d\ln P. \tag{4.71}
\]

Equations (4.70) and (4.71) form a system of equations in \( \partial f / \partial t \) and \( \partial f / \partial \ln P \):

\[
\begin{bmatrix}
\frac{1}{3} & -e_0 \\
\frac{\partial f}{\partial t} & \frac{\partial f}{\partial \ln P}
\end{bmatrix}
\begin{bmatrix}
\frac{\partial f}{\partial t} \\
\frac{\partial f}{\partial \ln P}
\end{bmatrix} =
\begin{bmatrix}
Q \\
df
\end{bmatrix}.
\]

If a square coefficient matrix for a set of \( n \) linear equations has a vanishing determinant, a necessary condition for finite solutions to exist is that when the right-hand side is substituted for any column of the coefficient matrix, the resulting determinants must also vanish. Thus the three determinants

\[
\begin{vmatrix}
\frac{1}{3} & -e_0 \\
\frac{\partial f}{\partial t} & \frac{\partial f}{\partial \ln P}
\end{vmatrix} = 0,
\]

and

\[
\begin{vmatrix}
\frac{1}{3} & Q \\
\frac{\partial f}{\partial t} & \frac{\partial f}{\partial \ln P}
\end{vmatrix} = 0,
\]

imply that the characteristics are

\[
\frac{dt}{\frac{1}{3}} = \frac{d\ln P}{-e_0} = \frac{df}{Q},
\]

or

\[
\frac{dt}{\frac{1}{3}} = \frac{d\ln P}{-e_0} \Rightarrow d\ln P = -e_0 dt \tag{4.72}
\]

and

\[
\frac{dt}{\frac{1}{3}} = \frac{df}{Q} \Rightarrow df = Q dt. \tag{4.73}
\]

After integration, equations (4.72) and (4.73) become

\[
\ln P = -3e_0 t + C_1 \tag{4.74}
\]

and

\[
f = 3Qt + C_2, \tag{4.75}
\]

respectively.

To determine the constants \( C_1 \) and \( C_2 \), consider the conditions at time \( t = t' + \Delta t \): The solutions of equations (4.74) and (4.75) are required on \( t = t' + \Delta t \), but the solution at a previous time \( t = t' + 2\Delta t / 3 \) is known. Consider the following set of boundary conditions:
1. $\ln P = \ln P_{\gamma_3}$ at $t = t' + 2\Delta t/3$ (known)
2. $\ln P = \ln P_{\gamma_3}$ at $t = t' + \Delta t$ (unknown)
3. $f = f_{s+\gamma_3}$ at $t = t' + 2\Delta t/3$ (known)
4. $f = f_{s+1}$ at $t = t' + \Delta t$ (unknown)

These conditions yield:

1. $\ln P_{\gamma_3} = -3e_0 t' - 2e_0 \Delta t + C_1 \Rightarrow C_1 = \ln P_{\gamma_3} + 3e_0 t' + 2e_0 \Delta t$
2. $\ln P_{\gamma_3} = -3e_0 t' - 3e_0 \Delta t + C_1 = -e_0 \Delta t + \ln P_{\gamma_3}$
3. $f_{s+\gamma_3} = 3Qt' + 2Q\Delta t + C_2 \Rightarrow C_2 = f_{s+\gamma_3} - 3Qt' - 2Q\Delta t$
4. $f_{s+1} = 3Qt' + 3Q\Delta t + C_2 = Q\Delta t + f_{s+\gamma_3}$

Therefore, in order to calculate the distribution function, $f_{s+1}$, at $t' + \Delta t$ at rigidity $P_{\gamma_3}$, one would need to know at what rigidity, $P_{\gamma_3}$, the distribution function, $f_{\gamma_3}$, at $t' + 2\Delta t/3$ is the same as the required value. To find this energy one uses the relation

$$\ln P_{\gamma_3} = e_0 \Delta t + \ln P_{\gamma_3}.$$ 

This rigidity generally does not lie on a grid point and, therefore, one needs to find the values of the distribution function on the grid points (say $k'$ and $k' + 1$) on both sides of the calculated energy, $\ln P_{\gamma_3}$, and do a simple interpolation to find the value for $f_{s+\gamma_3}(\ln P_{\gamma_3})$:

$$f_{s+1,i,j,k} = Q\Delta t + f_{s+\gamma_3,i,j,k'} + \frac{f_{s+\gamma_3,i,j,k' + 1} - f_{s+\gamma_3,i,j,k'}}{\Delta \ln P}(\ln P_{\gamma_3} - \ln P_{k'}) \quad (4.76)$$
Thus, the LOD solution follows from equations (4.64), (4.69) and (4.76).

The development of this section is sufficient to incorporate the first-order Fermi acceleration into a general modulation problem. Such a shock transition can be modelled by letting the solar wind speed (and other appropriate parameters) continuously change from upstream to downstream conditions over a series of grid points. This can be done sufficiently 'sharp' by employing a transformed radial grid with difference equations according to (4.10) and (4.11), and grid spacings only a small fraction of an AU in the vicinity of the shock.

This method works well, but it has a tendency to become unstable at high energies, where the diffusion mean free path of the cosmic rays becomes of the same order of magnitude as the size of the shock transition region. Therefore, a second approach was taken where the shock is modelled by a true discontinuity in the solar wind speed.

4.8 LOD Solution of the Time-Dependent TPE with a Discontinuity

Consider a heliosphere with a discontinuity in the outward flow of the solar wind at a radial distance of \( r_s \), where the flow velocity abruptly changes from supersonic to subsonic flow, across a true infinitesimal distance, i.e., a discontinuity. The handling of this discontinuity in the solution of the TPE presents a new problem.

From Section 3.3 we know that the radial streaming, \( S_r \), and the solution of the TPE, \( f \), must stay continuous across the discontinuity. This yields the boundary condition given in (3.44), which is a first-order hyperbolic PDE in two variables, \( \ln P \) and \( \theta \).

Therefore, to solve the TPE across a discontinuous solar wind transition, one must solve two different PDEs: equation (4.1) and equation (3.44), simultaneously. The streaming condition (3.44) across the shock is of the general form

\[
A_0 \left( \frac{\partial f}{\partial r} \right)_- = B_0 \left( \frac{\partial f}{\partial r} \right)_+ + C_0 \frac{\partial f}{\partial \ln P} + D_0 \frac{\partial f}{\partial \theta} + Q_0,
\]

(4.77)

with coefficients

\[
A_0 = 1, \\
B_0 = \frac{\kappa_{rr}^+}{\kappa_{rr}}, \\
C_0 = \frac{V^- - V^+}{3\kappa_{rr}}, \\
D_0 = \frac{\kappa_{\theta \theta}^- - \kappa_{\theta \theta}^+}{r_s \kappa_{rr}^+}, \\
Q_0 = \frac{Q_s}{\kappa_{rr}}.
\]

The TPE (4.1) is then valid in the domain \( r_0 \leq r < r_s \) and \( r_s < r \leq r_b \) \( \forall \theta, P \) and the matching condition (4.77) is valid at \( r = r_s \) \( \forall \theta, P \).

Note that the source/sink function, \( Q \), in (4.1) and (4.59) is non-zero only if there exist sources/sinks away from the discontinuity. At the discontinuity, \( r = r_s \), it can have no effect, since the TPE (4.1) is not valid in the discontinuity. The part of the source/sink function that is singular in the shock, \( Q_s \), is found in equation (4.77).
4.8.1 The Finite Difference Form of the Matching Condition

Using second-order one-sided differences [equations (4.12) and (4.13)] in \( r \), the matching condition becomes

\[
A f_{n_s,j,k} + A^- f_{n_s-1,j,k} + A^- f_{n_s-2,j,k} = B f_{n_s,j,k} + B^+ f_{n_s+1,j,k} + B^{++} f_{n_s+2,j,k} + C_0 \frac{\partial f}{\partial \ln P} + D_0 \frac{\partial f}{\partial \theta} + Q_0,
\]

or

\[
D_0 \frac{\partial f}{\partial \theta} + C_0 \frac{\partial f}{\partial \ln P} = (A - B) f_{n_s,j,k} + A^- f_{n_s-1,j,k} + A^- f_{n_s-2,j,k} - B^+ f_{n_s+1,j,k} - B^{++} f_{n_s+2,j,k} - Q_0 = F.
\]

This is a first-order hyperbolic PDE in two variables (\( \theta \) and \( \ln P \)) and may readily be solved with the second-order accurate Wendroff's Implicit method of Section 4.5:

\[
D_0 \left( \frac{f_{n_s,i+1,j} - f_{n_s,i,j}}{2\Delta \theta} + \frac{f_{n_s,i+2,j} - f_{n_s,i+1,j}}{2\Delta \ln P} \right) + C_0 \left( \frac{f_{n_s,i+1,j} - f_{n_s,i,j}}{2\Delta \ln P} + \frac{f_{n_s,i+2,j} - f_{n_s,i+1,j}}{2\Delta \ln P} \right) = F(\theta_i + \Delta \theta, \ln P_j + \Delta \ln P).
\]

(4.78)

To enhance the stability of this method, it is desirable to step the computational molecule in the direction in which drift takes place. Since shock drift (drift along the shock face) changes direction from one solar cycle to the next, i.e., from pole to ecliptic in one cycle and from ecliptic to pole in the other, the code must be written such that the Wendroff computational molecule can step up or down the shock face, depending on the sign of the drift cycle.

Case 1: Pole to Ecliptic

In the first case, one must derive a form of this discretization to step from the pole down to the ecliptic, i.e., from \( \theta = 0^\circ \) up to \( \theta = 90^\circ \).

Make the substitutions \( i = j - 1 \) and \( j = k - 1 \). Then (4.78) becomes

\[
D_0 \left( \frac{f_{n_s,i,j-1,k} - f_{n_s,i-1,j-1,k}}{2\Delta \theta} + \frac{f_{n_s,i,j-2,k} - f_{n_s,i-1,j-2,k}}{2\Delta \ln P} \right) + C_0 \left( \frac{f_{n_s,i-1,j-1,k} - f_{n_s,i-2,j-1,k}}{2\Delta \ln P} + \frac{f_{n_s,i,j-2,k} - f_{n_s,i-1,j-2,k}}{2\Delta \ln P} \right) = F_{j-1/2,k-1/2},
\]

or

\[
\mathcal{V} f_{n_s,j,k} + \mathcal{W} f_{n_s,j-1,k-1} - \mathcal{V} f_{n_s,j-1,k-1} - \mathcal{V} f_{n_s,j-1,k-1} = F_{j-1/2,k-1/2},
\]

with

\[
\mathcal{V} = \frac{1}{2} \left( \frac{D_0}{\Delta \theta} + \frac{C_0}{\Delta \ln P} \right),
\]

and

\[
\mathcal{W} = \frac{1}{2} \left( \frac{D_0}{\Delta \theta} - \frac{C_0}{\Delta \ln P} \right).
\]

73
This yields the finite difference form of the matching condition:

\[
(A - B - V)f_{n_\perp,j,k} = B^+ f_{n_{\perp}+1,j,k} + B^{++} f_{n_{\perp}+2,j,k} - A^- f_{n_{\perp}-1,j,k} - A^{--} f_{n_{\perp}-2,j,k} + W(f_{n_{\perp},j,k-1} - f_{n_{\perp},j-1,k}) - V f_{n_{\perp},j-1,k-1} + Q_0, \tag{4.79}
\]

with the computational molecule:

Note that the coefficients, \( W, V, A, A^-, A^{--}, B, B^+, B^{++} \) and \( Q_0 \) must be evaluated at \( (\theta_j - \Delta \theta/2, \ln P_k - \Delta \ln P/2) \) in order to determine \( f_{n_{\perp},j,k} = f(\theta_j, \ln P_k) \) from \( f_{n_{\perp},j-1,k} = f(\theta_j - \Delta \theta, \ln P_k), f_{n_{\perp},j,k-1} = f(\theta_j, \ln P_k - \Delta \ln P) \) and \( f_{n_{\perp},j-1,k-1} = f(\theta_j - \Delta \theta, \ln P_k - \Delta \ln P) \).

**Case 2: Ecliptic to Pole**

In the second case one must let the algorithm step from the ecliptic up to the pole, i.e., \( \theta = 90^\circ \) down to \( \theta = 0^\circ \). To accomplish this, one must multiply with \(-1\) because of the reversal in stepping direction, and make the substitutions \( i = j + 1 \) and \( j = k - 1 \). The resulting finite difference form of the matching condition is

\[
(A - B + V)f_{n_{\perp},j,k} = B^+ f_{n_{\perp}+1,j,k} + B^{++} f_{n_{\perp}+2,j,k} - A^- f_{n_{\perp}-1,j,k} - A^{--} f_{n_{\perp}-2,j,k} + W(f_{n_{\perp},j,k-1} - f_{n_{\perp},j-1,k}) + V f_{n_{\perp},j-1,k-1} + Q_0, \tag{4.80}
\]

with

\[
V = \frac{1}{2} \left( \frac{D_0}{\Delta \theta} - \frac{C_0}{\Delta \ln P} \right),
\]

and

\[
W = -\frac{1}{2} \left( \frac{D_0}{\Delta \theta} + \frac{C_0}{\Delta \ln P} \right).
\]

The computational molecule is:

**Combination of Cases 1 and 2**

The reason why these two algorithms must be used is that the \( \partial f/\partial \theta \)-term gives rise to a flux of particles in a particular direction. Therefore, to maintain the maximum stability, one must
choose the algorithm that steps in the direction of the resultant flux. In other words, the choice of the particular algorithm depends on the drift cycle. The two cases can be treated in combination by defining
\[ \zeta = \text{sgn}(qA). \]

Then (4.79) and (4.80) become
\[
(A - B + \zeta \nu) f_{n_s,j,k} = B^+ f_{n_s+1,j,k} + B^{++} f_{n_s+2,j,k} - A^- f_{n_s-1,j,k} - A^{--} f_{n_s-2,j,k} + W(f_{n_s,j,k-1} - f_{n_s,j+k,k}) + \zeta \nu f_{n_s,j+k,k} + Q_0
\]
(4.81)

with
\[ \nu = \frac{1}{2} \left( \frac{D_0}{\Delta \theta} - \zeta \frac{C_0}{\Delta \ln P} \right), \]
and
\[ W = -\zeta \frac{1}{2} \left( \frac{D_0}{\Delta \theta} + \zeta \frac{C_0}{\Delta \ln P} \right), \]

with the general computational molecule:

4.8.2 Solving the Radial Equation

To handle the discontinuity, equation (4.61) must now be solved on both sides of the shock while the matching condition (4.77) must be solved on the shock. This procedure needs four distinct boundary conditions.

On the inner boundary
\[ f_{s+\gamma_5,n_s,j,k} = v_1 f_{s+\gamma_5,n_s-1,j,k} + v_2 f_{s+\gamma_5,n_s+1,j,k} + v_3. \]
(4.82)

Approaching the shock from the upstream medium gives
\[ f_{s+\gamma_5,n_s-1,j,k} = v_4 f_{s+\gamma_5,n_s-1,j,k} + v_5 f_{s+\gamma_5,n_s+1,j,k} + v_6, \]
(4.83)

while ‘leaving’ the shock in the downstream medium requires
\[ f_{s+\gamma_5,n_s+1,j,k} = v_7 f_{s+\gamma_5,n_s-1,j,k} + v_8 f_{s+\gamma_5,n_s+1,j,k} + v_9. \]
(4.84)

Finally, at the outer boundary
\[ f_{s+\gamma_5,n_s+1,j,k} = v_{10} f_{s+\gamma_5,n_s+1,j,k} + v_{11} f_{s+\gamma_5,n_s-1,j,k} + v_{12}. \]
(4.85)

Special attention has to be paid to conditions (4.83) and (4.84). The assumption that \( f^+ = f^- \) yields for (4.83),
\[ v_4 = 0 \]
\[ v_5 = 0 \]
\[ v_6 = f^+_{s+\gamma_5,n_s,j,k}. \]
From (4.81) it can be seen that the discretization of (4.77) yields terms containing $f_{s+\gamma_3,ns-2,j,k}$, $f_{s+\gamma_3,ns-1,j,k}$, $f_{s+\gamma_3,ns+1,j,k}$, and $f_{s+\gamma_3,ns+1,j,k}$. This means that a valid boundary condition must contain all five of the above terms. However, (4.84) does not contain $f_{s+\gamma_3,ns-1,j,k}$ and $f_{s+\gamma_3,ns-2,j,k}$ and, therefore, these two terms must be eliminated. This is done by using (4.64) and (4.83), to write $f_{s+\gamma_3,ns-1,j,k}$ and $f_{s+\gamma_3,ns-2,j,k}$ in terms of $f_{s+\gamma_3,ns,j,k}$:

$$
\begin{align*}
    f_{s+\gamma_3,ns-1,j,k} &= Y_{ns-2} - X_{ns-2} f_{s+\gamma_3,ns-1,j,k} \\
    f_{s+\gamma_3,ns-2,j,k} &= Y_{ns-2} - X_{ns-2} f_{s+\gamma_3,ns-2,j,k}
\end{align*}
$$

where the $Y$'s and $X$'s have the same meanings as defined in Section 4.7.1. Now (4.81) becomes

$$
(A - A^- X_{ns-1} + A^- X_{ns-2} X_{ns-1} - B + \varsigma \psi) f_{s+\gamma_3,ns,j,k} =
$$

$$
B^+ f_{s+\gamma_3,ns+1,j,k} + B^{++} f_{s+\gamma_3,ns+2,j,k} + A^- X_{ns-2} Y_{ns-1} - A^- Y_{ns-1} - A^- Y_{ns-2} + W(f_{s+\gamma_3,ns,j,k-1} - f_{s+\gamma_3,ns,j+c,k}) + \varsigma V f_{s+\gamma_3,ns,j+c,k-1} + Q_0,
$$

and, therefore, we obtain for (4.84)

$$
\begin{align*}
    v_7 &= \frac{B^+}{D_{div}} \\
    v_8 &= \frac{B^{++}}{D_{div}} \\
    v_9 &= \frac{1}{D_{div}}(A^- X_{ns-2} Y_{ns-1} - A^- Y_{ns-1} - A^- Y_{ns-2} + W(f_{s+\gamma_3,ns,j,k-1} - f_{s+\gamma_3,ns,j+c,k}) + \varsigma V f_{s+\gamma_3,ns,j+c,k-1} + Q_0)
\end{align*}
$$

with

$$
D_{div} = A - A^- X_{ns-1} + A^- X_{ns-2} X_{ns-1} - B + \varsigma \psi.
$$

### 4.8.3 Solving the Polar Equation

To solve the polar equation, one must proceed as in Section 4.7.1 with one important difference: at the shock equation 4.81 must be used to calculate $f_{s+\gamma_3}$:

$$
(A - B + \varsigma \psi)f_{s+\gamma_3,ns,j,k} = B^+ f_{s+\gamma_3,ns+1,j,k} + B^{++} f_{s+\gamma_3,ns+2,j,k} - A^- f_{s+\gamma_3,ns-1,j,k} - A^- f_{s+\gamma_3,ns-2,j,k} + W(f_{s+\gamma_3,ns,j,k-1} - f_{s+\gamma_3,ns,j+c,k}) + \varsigma V f_{s+\gamma_3,ns,j+c,k-1} + Q_0.
$$

### 4.8.4 Solving the Energy Equation

As in the case of the polar equation (Section 4.7.2) the solution scheme is the same, but

$$
(A - B + \varsigma \psi)f_{s+1,ns,j,k} = B^+ f_{s+1,ns+1,j,k} + B^{++} f_{s+1,ns+2,j,k} - A^- f_{s+1,ns-1,j,k} - A^- f_{s+1,ns-2,j,k} + W(f_{s+1,ns,j,k-1} - f_{s+1,ns,j+c,k}) + \varsigma V f_{s+1,ns,j+c,k-1} + Q_0
$$

must be used at the shock to calculate $f_{s+1}$. 

76
4.9 Summary of Algorithm

Section 4.8 completes the numerical development of this thesis. The algorithm consists of a solution of the radial, polar and rigidity equations. In this final section we summarize the algorithms for these solutions as they were employed in a FORTRAN program. Steenberg (1995) subsequently rewrote this algorithm in C to enhance the program's portability.

4.9.1 Algorithm

Radial Equation

define
\[ X_{n-1} = X_n, X_{n-1}, \ldots, X_1, X_0, X_1, \ldots, X_{n-1}, X_n \]
\[ Y_0, Y_1, \ldots, Y_n \]
\[ X_{-i} = 0, i \geq 2, i \neq n + 1 \]
Set \( X_{-1} = -v_2, X_0 = -v_1 \) and \( Y_0 = v_3 \)
do for \( k = 1, \ldots, l - 1 \)
do for \( j = 1, \ldots, m - 1 \)
do for \( i = 1, \ldots, n_s - 1 \)
\[ Z_i = D_{r,1} - D_{r,1} X_{i-1} \]
\[ X_i = D_{r,1} - D_{r,1} X_{i-1} \]
\[ Y_i = E_{s,i,j,k} - D_{r,1} Y_{i-1} \]
done
\( X_{n_s} = -v_7 \)
\( X_{-(n_s+1)} = -v_8 \)
\( Y_{n_s} = v_9 \)
do for \( i = n_s + 1, \ldots, n - 1 \)
\[ Z_i = D_{r,1} - D_{r,1} X_{i-1} \]
\[ X_i = D_{r,1} - D_{r,1} X_{i-1} \]
\[ Y_i = E_{s,i,j,k} - D_{r,1} Y_{i-1} \]
done
\( f_{s+\gamma_3,i,1,j,k} = Y_{n_s} - X_{n_s-1} (v_{n_s-1} (v_{n_s-1} Y_2)) \)
do for \( i = n - 2, n - 3, \ldots, n_s + 1 \)
\[ f_{s+\gamma_3,i,j,k} = Y_i - X_{i} f_{s+\gamma_3,i+1,j,k} \]
done
\( f_{s+\gamma_3,n_s,j,k} = -X_{n_s} f_{s+\gamma_3,n_s+1,j,k} - X_{-(n_s+1)} f_{s+\gamma_3,n_s+2,j,k} + Y_{n_s} \)
do for \( i = n_s - 1, n_s - 2, \ldots, 1 \)
\[ f_{s+\gamma_3,i,j,k} = Y_i - X_i f_{s+\gamma_3,i+1,j,k} \]
done
done

Polar Equation

define
\[ X_{-m} = X_{-m}, X_{-m+1}, \ldots, X_{-1}, X_0, X_1, \ldots, X_{m-1}, X_m \]
\[ Y_{0-m} = Y_0, Y_1, \ldots, Y_m \]
\[ X_{-j} = 0, j \geq 2 \]

Set \( X_{-1} = -w_2, X_0 = -w_1 \) and \( Y_0 = w_3 \)

do for \( k = 1, \ldots, l - 1 \)
do for \( i = 1, \ldots, n - 1, i \neq n_s \)
do for \( j = 1, \ldots, m \)
\[ Z_j = D_{01} - D_{01} X_{j-1} \]
\[ X_j = \frac{D_{01} X_{j-1}}{Z_j} \]
\[ Y_j = \frac{E_{i+j,k} X_{j-1}}{Z_j} \]
done
\[ f_{s+\gamma, i,j,k} = \frac{1}{(A - B - \nu)} (B^+ f_{s+\gamma, n_s+1,j,k} + B^{++} f_{s+\gamma, n_s+2,j,k}) \]
\[ -A^{-} f_{s+\gamma, n_s-1,j,k} - A^{-} f_{s+\gamma, n_s-2,j,k} + W(f_{s+\gamma, n_s,j,k-1} - f_{s+\gamma, n_s,j-1,k}) \]
\[ -\nu f_{s+\gamma, n_s,j-1,k-1} + Q_0 \]
done

done

do for \( k = 1, \ldots, l - 1 \)
do for \( j = 1, \ldots, m \)
done

**Energy Equation**

do for \( i = 1, \ldots, n - 1, i \neq n_s \)
do for \( j = 1, \ldots, m \)
do for \( k = 1, \ldots, l - 1 \)
determine \( k' \)
\[ f_{s+1,i,j,k} = Q \Delta t + f_{s+\gamma, i,j,k'} + \frac{f_{s+\gamma, i,j,k'}}{\Delta \ln \bar{P}} (\ln \bar{P}_s - \ln \bar{P}_k') \]
done

done

do for \( k = 1, \ldots, l - 1 \)
do for \( j = 1, \ldots, m \)
done

\[ f_{s+1,n_s,j,k} = \frac{1}{(A - B - \nu)} (B^+ f_{s+1,n_s+1,j,k} + B^{++} f_{s+1,n_s+2,j,k}) \]
\[ -A^{-} f_{s+1,n_s-1,j,k} - A^{-} f_{s+1,n_s-2,j,k} + W(f_{s+1,n_s,j,k-1} - f_{s+1,n_s,j-1,k}) \]
\[ -\nu f_{s+1,n_s,j-1,k-1} + Q_0 \]
done

done
4.10 Discussion

Sections 4.7, 4.8 and 4.9 complete the numerical development of this thesis. The algorithm of Section 4.9 and the computer code developed around it are more versatile than might be expected. It is written so general that it can handle the following problems:

- an acceleration model where the SWTS is modelled as a true discontinuity according to equations (4.77) and (4.81);
- an acceleration model where the SWTS is modelled as a continuous transition of the solar wind from supersonic to subsonic speeds, since the model can calculate the solution even in areas where \( \nabla \cdot V < 0 \) with the aid of equation (4.76).
- This implies that an acceleration model with a cosmic-ray modified, or smoothed, SWTS can be modelled as a combination of a continuous decrease of the solar wind speed and a discontinuous transition with \( s < 4 \), constituting a weakened shock.
- In other words, models with an arbitrary plasma flow with compressions anywhere in the heliosphere, such as CIRs and MIRs, can be successfully negotiated.
- The code is also ‘downward compatible’ in the sense that a pure modulation model, where a galactic spectrum at the outer boundary is modulated without acceleration, can be obtained by setting the compression ratio equal to one \( (s = 1) \), and allowing a constant solar wind on both sides of the ‘shock’. This has the effect that the matching condition (4.77) reduces to

\[
\left( \frac{\partial f}{\partial r} \right)^{-} = \left( \frac{\partial f}{\partial r} \right)^{+}
\]

i.e., both the solution and its gradient are continuous across the shock.
- By reducing the number of grid points in the \( \theta \)-grid to 3, and changing the coefficients to a spherically symmetric form, both spherically symmetric acceleration and pure modulation solutions can be obtained with the same code.
- From the time-dependent nature of the model, it is obvious that steady-state (time-asymptotic) and time-dependent solutions can be obtained for every physical system modelled.

In the development of this code, it was attempted to write it in such a way that all of the above scenarios can be obtained by changing only a few (preferably one) numerical parameters. With relatively minor changes, this model can also be adapted to almost any diffusion-convection system in astrophysics.

In Chapter 5 this model will be applied to a two-dimensional heliosphere, with and without a SWTS, and the characteristics of the results will be discussed in detail.
Chapter 5

Basic Solutions

5.1 Introduction

The numerical solution of the Parker Transport Equation, as described in Chapter 4, forms the core of this thesis. In the forthcoming chapters the properties of this numerical solution will be explored. Its qualities, limitations and deficiencies will be discussed and finally, it will be applied to actual observations.

This subsequent development is an extensive task, because the effects of numerous physical and numerical parameters on the solutions must be explored in detail. To structure this development properly, Chapter 5 is devoted solely to a basic set of demonstration solutions.

These demonstration solutions are for a model heliosphere with a reasonable set of modulation parameters, and various forms of the solution are discussed in detail. This chapter is structured such that it gives by far the most comprehensive exposition of basic modulation and acceleration solutions of the transport equation yet. Almost all of the effects described here were discovered in the pioneering work of Jokipii and co-workers (Jokipii and Kopriva, 1979; Jokipii and Davila, 1981; Jokipii and Thomas, 1981; Jokipii, 1986; Jokipii, 1988; Jokipii, 1992; Kóta and Jokipii, 1994), and in the modulation solutions of Moraal and Gleeson (1975), Moraal et al. (1979), Potgieter and Moraal (1985), Moraal (1993), and Reinecke et al. (1993). At the outset, we note that we find no qualitative differences between our solutions and those presented previously. Therefore, this entirely independent numerical model confirms our agreement with all previous work. However, many of the properties of the previous solutions were never described as comprehensively as is done here.

In this chapter an optimal set of numerical variables, such as grid spacing, boundary conditions and time steps is chosen without discussing its motivation. Chapter 6 is devoted to discussing the influence these parameters have on the solutions.

The structure of this chapter is as follows:

- A brief description of the model heliosphere and its physical transport parameters is given in Section 5.2.
- Section 5.3 starts with a simple, standard, two-dimensional steady-state modulation solution of the TPE. This well-tested solution is included because it forms the benchmark to compare all subsequent developments with.
- In Section 5.4 this modulation solution is repeated, but with the time-dependent version developed in Chapter 4. It is shown that when the transport parameters are independent of time, this time-dependent solution converges asymptotically to the correct steady-state solution. This behavior demonstrates the basic soundness of the newly-developed solution technique.
Section 5.5 discusses the acceleration of protons by a termination shock of the solar wind. Two kinds of acceleration are involved here. Firstly, any galactic/interstellar cosmic ray component will be re-accelerated by the SWTS. Thus, in Section 5.5.1, it is shown how this re-acceleration alters the pure modulation solution of Section 5.4. Secondly, Section 5.5.2 describes how a source of low-energy protons, injected at the SWTS, is accelerated to cosmic-ray energies. In reality the acceleration/modulation process is a combination of these two effects. Thus, in Section 5.5.3, the previous galactic re-acceleration and the primary source acceleration solutions are combined. Finally, Section 5.5.4 demonstrates that, since the transport equation is linear, the solution of Section 5.5.3 is just the sum of the re-acceleration solution of Section 5.5.1 and the source acceleration solution of Section 5.5.2.

Up to this point drift effects in the large scale heliospheric magnetic field are ignored. This is done to demonstrate the underlying properties and principles in a simple fashion. The drifts are incorporated in Section 5.6, and it is found that they have significant effects, which complicates the simple underlying picture rather drastically.

Finally, in Section 5.7, it is demonstrated that the energy spectra of heavier species can basically be obtained by scaling the proton spectra with the mass-to-charge ratio, A/Z, of the heavier particles. Specific reference is made to the resulting spectra for Oxygen, the species with the best observed anomalous component.

5.2 The Model Heliosphere and its Transport Parameters

A standard set of solution parameters was chosen to demonstrate the capabilities of the model. The TPE was solved on a spatial domain (see Figure 5.1) that starts on the surface of the sun (r_{sb} = 0.005 AU) and extends up to a boundary at a heliospheric distance of r_b = 100 AU. Since symmetry around the ecliptic plane was assumed, the transport equation was solved from the pole (θ = 0) to the ecliptic plane (θ = π/2) and the solution was simply mirrored to obtain the solution in the southern parts of the heliosphere. In the cases where a solar wind termination shock was included in the modulation region, it was placed at a distance of r_s = 70 AU from the center of the Sun.

The Solar Wind Model: The solar wind (see Figure 5.2) is taken to be independent of polar angle and time. Inside the termination shock it has a constant radial velocity of 400 km/s. At the solar wind termination shock (if present) the flow velocity falls abruptly to 100 km/s. Beyond the shock the solar wind falls off proportional to 1/r^2, i.e., it is divergence free. This wind profile is described by the function

\[ V(r) = V_0 \left\{ 1 + H(r - r_s) \left[ \frac{1}{s} \left( \frac{r_s}{r} \right)^2 - 1 \right] \right\} \]

with \( V_0 = 400 \text{ km/s} \), \( s = 4 \) the compression ratio of the shock and \( H \) the Heaviside-function. In the cases where no termination shock is present \( V(r) = V_0 \) was used throughout the entire radial domain.

The Magnetic Field Model: The standard Parker spiral field model, described in Chapter 2 Sections 2.3.1 and 2.3.7, was used throughout this chapter.
The Diffusion Coefficients: Since the microphysics of particle transport in irregular magnetic fields is still poorly understood and in situ measurements of the turbulence in the IMF throughout the heliosphere are not yet possible, the spatial and rigidity dependence of the diffusion coefficients must be determined by purely empirical arguments: They must be chosen in such a way that they yield reasonable results in a modulation model. Following Jokipii and Davila (1981), the diffusion coefficients are assumed to be proportional to $1/B$.

Since the gyro-radius (and diffusion mean free path) rises with energy, the diffusion coefficients must also rise with increasing energy. Therefore an energy dependence was built into the diffusion

Figure 5.1: A schematic representation of the model heliosphere.

Figure 5.2: A schematic representation of the solar wind profile.
coefficients. In this chapter this energy dependence was assumed to be proportional to $\beta P$, i.e.,

$$\kappa_\parallel = (\kappa_\parallel)_0 \beta \frac{P}{P_0} \frac{B_0}{B}$$ (5.2)

and

$$\kappa_\perp = \frac{\kappa_\parallel}{100}$$ (5.3)

with $(\kappa_\parallel)_0 = 1.8 \times 10^{22}$ cm$^2$/s and $P_0 = 1$ GV. Reinecke et al. (1993) obtained considerable success with data fits using diffusion coefficients of this form.

**General Boundary Conditions:** At the inner boundary the radial gradient was assumed to be zero ($\partial f/\partial r = 0$). Since the inner boundary was taken to be the solar surface, this condition is the same as stating that the radial streaming, the $r$-component of (3.37), is zero at the inner boundary. The boundary condition (4.82) represents the general form of the boundary condition at the inner radial boundary of the heliosphere:

$$\left. \frac{\partial f}{\partial r} \right|_{r=r_\odot} = 0.$$  

If the radial grid is chosen such that $r_1 - r_0 = r_2 - r_1 = \Delta r$, then the derivative with respect to $r$ can be approximated by a central difference as

$$\left. \frac{\partial f}{\partial r} \right|_{r} = \frac{f_{s+\gamma_3,2,j,k} - f_{s+\gamma_3,0,j,k}}{2\Delta r} = 0.$$  

This implies that

$$f_{s+\gamma_3,0,j,k} = f_{s+\gamma_3,2,j,k}$$  

or, in terms of 4.82

$$v_1 = 0$$

$$v_2 = 1$$

$$v_3 = 0.$$  

At the solar wind termination shock (if present) the distribution function and radial streaming were assumed to be related by $f^- = f^+$ and $S^+_r - S^-_r = Q_*$. As was shown in Section 4.8.2 the boundary conditions (4.83) and (4.84) then become

$$v_4 = 0$$

$$v_5 = 0$$

$$v_6 = f^+_{s+\gamma_3,n,s,j,k}$$  

and

$$v_7 = \frac{B^+}{D_{\text{div}}}$$

$$v_8 = \frac{B^{++}}{D_{\text{div}}}$$

$$v_9 = \frac{1}{D_{\text{div}}}(A^-X_{n_s-2}Y_{n_s-1} - A^-Y_{n_s-1} - A^-Y_{n_s-2} + \nu f_{s+\gamma_3,n_s,i,j,k} + \nu f_{s+\gamma_3,n_s,i+1,j,k} + Q_0)$$
with
\[ D_{\text{div}} = A - A^-X_{n-1} + A^-X_{n-2}X_{n-1} - B + \varsigma V. \]

At the outer boundary, the distribution function was put equal to some interstellar value, \( f_G \), which may be zero. Thus, the outer radial boundary condition, equation (4.85), can be derived from
\[ f|_{r=r_b} = f_G. \]

In other words
\[ f_{i=n} = f_{s+\gamma_3,n,i,k} = f_G \]

which yields
\[
\begin{align*}
v_{10} &= 0 \\
v_{11} &= 0 \\
v_{12} &= f_G.
\end{align*}
\]

The distribution function was assumed to be symmetrical about the poles and the ecliptic plane. This assumption implies that the polar angle gradient must be zero (\( \partial f / \partial \theta = 0 \)) at these positions in the heliosphere. From these assumptions, the specific forms of the boundary conditions (4.67) and (4.68) can be found:

\[
\frac{\partial f}{\partial \theta} \bigg|_{\theta=0,\pi/2} = 0.
\]

If the polar grid is chosen such that \( \theta_1 - \theta_0 = \theta_2 - \theta_1 = \Delta \theta \) and \( \theta_{m+1} - \theta_m = \theta_m - \theta_{m-1} = \Delta \theta \), then
\[
\frac{\partial f}{\partial \theta} \bigg|_{ij} = \frac{f_{s+\gamma_3,i,2,k} - f_{s+\gamma_3,i,0,k}}{2\Delta \theta} = 0
\]

or
\[ f_{s+\gamma_3,i,0,k} = f_{s+\gamma_3,i,2,k} \]
at the first boundary. The second boundary is obtained in the same way:
\[ f_{s+\gamma_3,i,m+1,k} = f_{s+\gamma_3,i,m-1,k}. \]

This implies that
\[
\begin{align*}
w_1 &= 0 \\
w_2 &= 1 \\
w_3 &= 0
\end{align*}
\]

and
\[
\begin{align*}
w_4 &= 0 \\
w_5 &= 1 \\
w_6 &= 0.
\end{align*}
\]

Except where specified differently, the solutions were obtained on a rigidity domain between \( P_{\text{min}} \) and \( P_{\text{max}} \). In the acceleration solutions, mono-energetic particles were injected at the shock at a rigidity, \( P_i \). The lower-rigidity boundary, \( P_{\text{min}} \), must be slightly lower than the injection rigidity, \( P_i \), in order to avoid specifying contradicting conditions on the lower-rigidity boundary on the shock, where the distribution function was put equal to zero.
Some Definitions: The model generates the omni-directional distribution function, \( f \), for cosmic rays in the heliosphere as function of two spatial coordinates, rigidity and time: \( f = f(r, \theta, P, t) \). To obtain quantities that are equivalent to those measured by cosmic ray detectors, this distribution function must be expressed as an intensity. The intensity with respect to kinetic energy, \( j_T = p^2 f \), is usually measured and therefore we express our results in terms of \( j_T \) in the units of particles/m²/s/sr/(MeV/nucleon).

The gradient in the intensity of cosmic rays in the heliosphere is defined as

\[
\mathbf{g} = \frac{1}{j_T} \nabla j_T = \frac{1}{j_T} \frac{\partial j_T}{\partial r} e_r + \frac{1}{r j_T} \frac{\partial j_T}{\partial \theta} e_\theta + \frac{1}{r \sin \theta j_T} \frac{\partial j_T}{\partial \phi} e_\phi
\]

\( = g_r e_r + g_\theta e_\theta + g_\phi e_\phi. \) (5.4)

The radial gradient is simply

\[
g_r = \frac{1}{j_T} \frac{\partial j_T}{\partial r} = \frac{\partial \ln j_T}{\partial r}, \tag{5.5}
\]

and the (usually measured) latitudinal gradient, \( g_\lambda \), is related to the polar gradient, \( g_\theta \), by

\[
g_\lambda = -r g_\theta = -\frac{1}{j_T} \frac{\partial j_T}{\partial \theta} = -\frac{\partial \ln j_T}{\partial \theta}. \tag{5.6}
\]

In practice, only the difference between two intensities, measured at positions 1 and 2 in the heliosphere, is known. From these measurements, the average radial and latitudinal gradients are defined as

\[
< g_r > = \frac{\Delta \ln j_T}{\Delta r} = \ln \left( \frac{j_{T2}/j_{T1}}{r_2 - r_1} \right) = \ln \left( \frac{j_{T2}/j_{T1}}{r_2 - r_1} \right)
\]

(expressed in % per AU) and

\[
< g_\lambda > = -\frac{\Delta \ln j_T}{\Delta \theta} = \frac{\ln(j_{T1}/j_{T2})}{\theta_2 - \theta_1} = \frac{\ln(j_{T1}/j_{T2})}{\theta_2 - \theta_1}
\]

(expressed in % per degree), where \((r_1, \theta_1)\) and \((r_2, \theta_2)\) are the two positions in polar coordinates.

With the time-dependent model, the distribution function is also a function of time. In this chapter we are primarily only interested in the time-asymptotic solution. In order to study the convergence of \( f \) to the time-asymptotic solution, we define the temporal gradient as

\[
g_t = \frac{1}{f} \frac{\partial f}{\partial t} = \frac{\partial \ln f}{\partial t}. \tag{5.9}
\]

This quantity can only be calculated between two subsequent time steps:

\[
< g_t > = \frac{\Delta \ln f}{\Delta t} = \frac{\ln(f_s/f_{s-1})}{\Delta t}, \tag{5.10}
\]

expressed in % per year. For purely practical reasons we usually plot the absolute value of this expression.

5.3 A Steady-State Modulation Solution for Protons

In Chapter 4 a numerical code was developed with which the time-dependent TPE can be solved in two spatial dimensions. Before this code can be used, it must first be thoroughly tested. Since
a well-tested numerical code, originally developed by H. Moraal (private communication, 1991), to solve the steady-state TPE

\[ \nabla \cdot [Vf - \mathcal{K} \cdot \nabla f] = \frac{1}{3p^2} \frac{\partial}{\partial p}(p^3 f) \]  

(5.11)

was readily available, it represented an ideal benchmark against which the newly developed time-dependent code can be tested and calibrated. This test is even more attractive in view of the fact that Moraal used the Peaceman-Rachford ADI scheme to solve equation (5.11), which is an entirely different scheme than the present LOD-based time-dependent algorithm.

This steady-state model cannot simulate a solar wind termination shock, and therefore a constant solar wind was used up to a modulation boundary, placed at \( r_b = 100 \) AU, where the distribution function was put equal to an approximation of the Local Interstellar Spectrum (LIS):

\[ f_G = (f_G)_0 \frac{\beta}{P^2} (T + 0.5E_0)^{-2.6} \]  

(5.12)

with \( \beta = v/c, P = pc/q \) the particle rigidity, \( T \) the kinetic energy and \( E_0 \) the proton rest mass energy.

The code was stepped downward in rigidity from \( P_{\text{max}} = 20 \) GV to \( P_{\text{min}} = 0.09 \) GV in 266 steps. The solution grid had 141 radial and 31 latitudinal nodes from \( \theta = 0 \) to \( \theta = \pi/2 \).

The diffusion coefficients used are given by equations (5.2) and (5.3). The effects of particle drifts due to gradients and curvatures in the large-scale magnetic field were not included in this run.

Figures 5.3 to 5.5 show the results of this steady-state modulation model. The format of this set of 16 figures will be used as a template to represent results throughout large parts of the rest of this chapter. Therefore, the features of this simple solution will be discussed in detail because they represent the underlying properties of the more involved solutions to follow.

Figures 5.3(a), (b) and (d) show the intensity spectra in the ecliptic plane, at a polar angle of 45° and over the poles respectively. The three solid lines in each frame show the intensities (from the top down) at 70, 30 and 1 AU. The dotted line shows the LIS (5.12). Figure 5.3(e) shows the inner and outer radial gradients in the ecliptic plane, calculated from (5.7) between 1 and 30 AU and between 30 and 60 AU respectively. Figure 5.3(f) shows the latitudinal gradients at 1, 30 and 60 AU, calculated from (5.8) between \( \theta = 90^\circ \) and \( \theta = 60^\circ \).

Figures 5.4(a) to 5.4(e) show the radial intensity profiles normalized to the interstellar intensity (5.12), for energies of 10, 30, 100, 300 and 1000 MeV respectively. The dotted line represents the radial profile over the poles, the short dashed line the profile at a polar angle of 30°, the medium dashed line at 60° and the solid line represents the profile in the ecliptic (90°). Figure 5.4(f) shows the equatorial intensity profiles of Figures 5.4(a)-(e).

Finally, Figures 5.5(a) to 5.5(e) show the polar intensity profiles relative to the interstellar intensity (5.12) for the above-mentioned five energies. The four lines in each frame show the polar angle profiles (from the top down) at 70 (solid line), 50 (long dashed line), 30 (medium dashed line) and 1 AU (short dashed line).

Some well-known characteristics can be seen from the spectral shapes: The modulation increases with decreasing energy. This is due to the rigidity dependence of the diffusion coefficients (5.2) and (5.3), which implies a cosmic-ray mean free path that increases with rigidity, decreasing the modulation effects with increasing rigidity. Therefore, a particle with a large mean free path,
Figure 5.3: Energy spectra, radial and latitudinal gradients obtained with a steady-state model. Except where otherwise specified, the intensities shown are measured in part/m²/s/sr/MeV. Figure (c) is an empty frame where time-dependent spectra will be shown in later sections.
Figure 5.4: Normalized intensity as function of radial distance for different energies as obtained with a steady-state model.
Figure 5.5: Normalized intensity as function of polar angle for different energies as obtained with a steady-state model. Figure (f) is an empty frame where a time-dependent quantity will be shown in later sections.
which is proportional to the gyro-radius, would barely be affected by the heliosphere. Hence the no-modulation condition at high energies.

The intensities also decrease with decreasing radial distance, as the convection effect of the solar wind keeps particles from freely diffusing into the inner heliosphere. The radial profiles of Figure 5.4 shows this in detail. Again, the high energy particles diffuse easier into the heliosphere than low energy particles. Hence the higher relative intensities of 1 GeV particles in the inner heliosphere than, for example, 10 MeV particles.

Notice the ‘skin-depth’ effect near the outer boundary in Figure 5.4. It is due to the underlying modulation integral

$$\exp \left( - \int r_b \frac{V}{\kappa} dr \right),$$

which is small at low energies, so that the diffusive length scale is only about 5 AU at 10 MeV. Deeper into the heliosphere the particles are in the adiabatic mode of propagation with the typically small radial gradients. The adiabatic limit is clearly absent at high energies, [see Figure 5.4(e)] because the diffusive length scale approaches the size of the heliosphere. This skin-depth barrier is equivalent to the concept of a modulation barrier in the outer heliosphere as described by Potgieter and Le Roux (1989) and Quenby et al. (1990).

From the three spectral shapes on Figure 5.3 (a, b, and c) one can clearly see that there is between 30 and 50 times more modulation at Earth (at 1 AU in the ecliptic) than at the equivalent position over the poles. This is also exemplified by the polar angle profiles in Figure 5.5 and the polar radial profiles in Figure 5.4.

The shape of the latitudinal gradients becomes clear if they are considered together with the polar angle profiles: At lower energies the latitudinal gradients are large and at higher energies they become small. Note that \( g_\lambda \) is strictly positive over the whole energy domain.

These effects may easily be explained by considering the geometry of the Parker spiral field and the radial diffusion coefficient,

$$\kappa_{rr} = \kappa_\parallel \cos^2 \psi + \kappa_\perp \sin^2 \psi. \quad (5.13)$$

In the ecliptic plane the spiral angle, \( \psi = \tan^{-1}[\Omega (r - r_0) \sin \theta / V] \), has an approximate value of 84° at 10 AU and continues to converge toward 90° beyond that. Since \( \cos^2 \psi \propto r^{-2} \), the second term dominates \( \kappa_{rr} \). In other words, small values of \( \kappa_{rr} \) prevail in the ecliptic plane and, therefore, strong modulation conditions exist there.

Over the poles, however, the spiral angle is zero and \( \kappa_{rr} = \kappa_\parallel \), which results in large values of \( \kappa_{rr} \) in this region. The larger \( \kappa_{rr} \) becomes, the more the situation approaches free propagation into the heliosphere. Hence the weak modulation observed in the model solutions.

Two effects in this solution deserve detailed discussion, namely the so-called adiabatic limit of modulation, and the bulging spectra in the outer heliosphere:

The Adiabatic Limit: The spectral shapes in Figure 5.3(a) show the well-known \( j_T \propto T \) limit at low energies. Simultaneously, the radial gradients in Figure 5.3(e) show maxima between 100 and 1000 MeV. Below these energies the radial gradients become small. This low-energy limit of \( j_T \propto T \) and small radial gradients is known as the adiabatic limit of modulation. Reinecke and Moraal (1992) made a detailed study of this limit.

In the adiabatic limit the adiabatic energy loss term dominates the transport equation (hence the name). Since the diffusion coefficients are proportional to \( \beta P \), it means that that all the diffusion terms in the Parker equation \([a_0(\partial^2 f/\partial r^2), b_0(\partial^2 f/\partial \theta^2), c_0(\partial f/\partial r) \) and \( d_0(\partial f/\partial \theta)\]
will decrease with decreasing energy. The energy loss term, \( e_0 \delta f / \partial \ln P \), however, only depends on the divergence of the solar wind and therefore starts to dominate the smaller diffusion terms. The Parker TPE (5.11) therefore reduces to

\[
V \cdot \nabla f = \frac{1}{3} (\nabla \cdot V) \frac{\partial f}{\partial \ln P}.
\]

In the case of spherical symmetry, with a solar wind independent of \( r \), and for a spectrum of the form \( f \propto p^{-\gamma} \), this equation reduces to the simple form

\[
g_r = \frac{2\gamma}{3r}.
\]

The only acceptable solution is \( \gamma = 0 \), otherwise there will exist unlimited large (positive or negative) radial gradients in the inner heliosphere (small \( r \)). Since \( j_T \propto p^2 f \), this produces the classical \( j_T \propto T \) spectra at non-relativistic energies.

Figure 5.3(d) shows the radial gradients as function of energy. The inner radial gradient in the figure was calculated between 1 AU and 30 AU in the ecliptic, using equation (5.7) and the outer radial gradient was calculated in the same way between 30 AU and 60 AU also in the ecliptic plane. The basic shape of both these gradients is the same: They start from some low value at the low-energy side and rise to a maximum and then fall again to a lower value towards higher energies. The basic shape may be understood by looking at the spectra in the ecliptic. Consider first the outer radial gradient: The maximum difference between the 30 AU spectrum and 70 AU spectrum is found around 200 MeV. The outer radial gradient thus reaches its maximum around that energy. Towards lower energies, in the adiabatic limit, the 30 and 70 AU spectra differ very little and the outer radial gradient falls towards zero, as expected.

Above the adiabatic limit (at high energies) the force field approximation (Gleeson and Axford, 1968b), which assumes zero streaming (\( S = 0 \)), becomes valid and one may write for the spherical symmetrical case

\[
CV f - \kappa_{rr} \frac{\partial f}{\partial r} = 0
\]

or

\[
g_r = \frac{CV}{\kappa_{rr}}.
\]

Since \( \kappa_{rr} \propto \beta P \), while \( C \) increases much less with rigidity, \( g_r \) must decrease with increasing rigidity if the force field approximation is valid, i.e., at high energies.

For the inner radial gradients the explanation is the same, with only minor differences distinguishing it from the outer radial gradients. The most noticeable of these differences is the shift in the maximum of the inner gradients toward higher energies (600 MeV) as can be seen from the difference between the spectra at 1 and 30 AU.

Bulging Spectra: The spectra in the outer heliosphere (70 AU) display a curious 'bulge' above the adiabatic limit. The peak of the \( r = 70 \) AU spectrum in the ecliptic plane is located at an energy of \( \sim 300 \) MeV. Below this peak the spectrum falls off steeply, before it recedes into the \( j_T \propto T \) adiabatic limit. Thus, the spectral shapes in the outer parts of the heliosphere are clearly different from those in the inner heliosphere. Reinecke and Moraal (1992) have shown that the origin of this effect is an inevitable side-effect of the adiabatic limit:

In the inner heliosphere the proton intensities are modulated very strongly and therefore the spectrum blends smoothly into the adiabatic limit at energies below the peak intensity at \( \sim 600 \) MeV. For the spectra in the outer heliosphere, like the \( r = 70 \) AU spectrum, the situation
is dramatically different. At this position, much nearer to the no-modulation boundary at 100 AU, the amount of modulation by the heliospheric processes is much less than in the inner heliosphere. Therefore, the resulting intensities should be much higher in these regions. Since the energy dependence of the TPE is assumed to be the same everywhere in the heliosphere, the spectra should go into the adiabatic limit at the same energies at which it happens in the inner heliosphere, i.e., at energies of about 30–40 MeV. Since the radial gradients become very small in the adiabatic limit, the spectrum at 70 AU is ‘forced’ downwards in intensity towards the levels in the inner heliosphere and the ‘bulge’ is created.

To summarize, this simple steady-state model yields realistic answers as can be seen, e.g., from the radial gradients in Figure 5.3(d): They are about 1 %/AU at 2000–3000 MeV, with maximum values of 2–3 %/AU at about 300–600 MeV. This is in general agreement with observations, e.g., McKibben (1989a). Much less information is available about the latitudinal gradients, but according to McKibben our calculated values are reasonable for a so-called positive drift cycle. These latitudinal gradients are naturally completely wrong for the negative drift cycle.

5.4 A Time-Asymptotic Modulation Solution for Protons

In this section the numerical solution of Sections 4.7 and 4.8 has been used to solve the time-dependent TPE

$$\frac{\partial f}{\partial t} = \nabla \cdot [\mathbf{K} \cdot \nabla f - \mathbf{V} f] + \frac{1}{3p^2} \frac{\partial}{\partial p}(p^2 f) + Q. \quad (5.14)$$

The aim of this procedure is to calibrate and compare the newly-developed solution to that of the older and thoroughly tested ADI solution with which the steady-state TPE (5.11), was solved in the previous section. For this reason, exactly the same physical model and boundary conditions were used:

The radial domain stretched from the surface of the sun to a no-modulation boundary at $r = r_b = 100$ AU. At $r = r_0$ the boundary condition $\partial f / \partial r = 0$ was used. At the outer boundary the distribution function was again put equal to $f_G$, given by equation (5.12), and $f$ was calculated at 141 distinct nodes in the radial domain. Symmetry around the ecliptic plane was again assumed and the boundary condition $\partial f / \partial \theta = 0$ was used at the pole and the ecliptic. The distribution function, $f$, was calculated at 31 nodes in the polar angle domain. The solution was mirrored about the ecliptic plane to obtain the solution in the southern hemisphere. The same diffusion coefficients, equations (5.2) and (5.3), and a constant solar wind were used.

Due to the inherent differences between a time-dependent and steady-state model, the rigidity domain and boundary conditions cannot be treated in the same way. The rigidity domain between $P_{\text{min}} = 0.09$ GV and $P_{\text{max}} = 20$ GV was divided into 88 nodes and the distribution function was put equal to $f_G$ at $P_{\text{max}}$.

Since time has replaced the natural logarithm of rigidity as the stepping parameter, an initial condition in time is required to get the code running. For this initial condition at $t = t_0$ the whole heliosphere was assumed to be filled with protons with an energy spectrum given by equation (5.12), the LIS. This situation can be viewed as switching the modulation effect of the sun on at $t = t_0$.

The code was allowed to run until sufficient convergence was obtained after about 868 days or 2.4 years model time (4000 time steps). In this way, the time-dependent equivalent of the previously discussed steady-state model was obtained. This case will simply be called the pure modulation case. The main differences between this case and the steady-state solutions will be discussed.

92
Figure 5.6: Energy spectra, radial and latitudinal gradients obtained with a time-asymptotic model.
Figure 5.7: Normalized intensity as function of radial distance for different energies as obtained with a time-asymptotic model.
Figure 5.8: Normalized intensity as function of polar angle for different energies as obtained with a time-asymptotic model.
Figures 5.6 to 5.8 present the results of this run in the same format that was used in the previous section. There are two additional figures, namely 5.6(c) and 5.8(f). The other figures in this set correspond exactly to those in the same positions in the previous set.

Figure 5.6(c) shows the same spectra as Figure 5.6(a) but the intermediate spectra at 174, 347, 521 and 694 days model time are also shown in the dashed lines. This figure gives a good intuitive measure of the convergence of the distribution function with increasing time. To describe this convergence more analytically, Figure 5.8(f) shows the previously defined temporal gradient, equation (5.10), calculated between the last and penultimate time step, as function of rigidity at two positions in the ecliptic plane. The solid line represents the results at 70 AU and the dashed line the results at 30 AU. The change in $f$ with time at 70 AU is at the most $\sim 4 \times 10^{-5} \%$/year at energies around 60–70 MeV. At lower energies the change is more or less independent of energy. Beyond 100 MeV the change in $f$ becomes insignificant. The changes in $f$ at 30 AU are even less than at 70 AU.

The reader is invited to use the loose figures in the back of this thesis as overlays to confirm that any differences between the

- time-asymptotic LOD solutions of Figures 5.6–5.8, and
- steady-state ADI solutions of Figures 5.3–5.5

are entirely inconsequential. This validates the numerical development in Chapter 4 of this thesis, and the use of the new solution in more complicated models, such as those containing a solar wind termination shock.

In fact, the solution shown here is not the simple LOD of Section 4.7, but the modified LOD of Section 4.8. Thus, it contains a 'shock' at 70 AU and matches the solutions inside and outside this 'shock' according to equation (4.77). The compression ratio of this 'shock' was put equal to 1 and the solar wind was kept constant beyond the ‘shock’ so that it has no physical effect, and the matching condition, equation (4.77), simply reduces to

$$\left(\frac{\partial f}{\partial r}\right)^+ = \left(\frac{\partial f}{\partial r}\right)^-.$$  

Since the steady-state ADI and time-asymptotic LOD solutions are for all practical purposes identical, three important conclusions can be drawn from this:

Firstly, the steady-state solution of the TPE can be obtained from the time-dependent TPE by using a time-independent solar wind model and diffusion coefficients and calculating the time-asymptotic solution. Secondly, the modified LOD method developed in Chapter 4 is working well and produces physically correct answers. Thirdly, this result has been tested with a pure Modified LOD method as developed in Section 4.7, as well as the model containing a 'phantom shock' that was incorporated into the Modified LOD numerical code with the techniques developed in Section 4.8. Therefore, this confirms the validity of the approach with which the matching condition on the shock was incorporated into the numerical scheme.

5.5 The Acceleration of Cosmic-Ray Protons by the SWTS

In this section we show solutions of the newly-developed numerical code to demonstrate its capabilities to model the acceleration of particles at the solar wind termination shock. Therefore, to obtain a two-dimensional steady-state solution in a heliosphere with an accelerating termination shock, we must again obtain a time-asymptotic solution as in the previous section.
This solution technique will be used to address the problem of anomalous cosmic rays (ACR), that are thought to be singly-charged ions that originate inside the heliosphere and are accelerated to cosmic-ray energies at the SWTS. Although the anomalous component is observed most strongly in Oxygen (O), Helium (He), Argon (Ar) and Neon (Ne), we will start by considering the acceleration of protons at the SWTS.

Both the acceleration and re-acceleration of protons will be studied in order to demonstrate a number of general characteristics. The solar wind termination shock was modelled with a discontinuous drop in the solar wind and the matching condition, equation (4.77). As in the previous section, the effects of particle drifts were ignored to avoid complicating the solutions unnecessarily.

In all cases the code was started with some initial condition imposed on the distribution function, $f$, and the solution was updated in steps of 2.6 hours (model time) until the process was stopped after sufficient convergence was obtained.

### 5.5.1 The Re-Acceleration of Galactic Cosmic-Ray Protons by the SWTS

In this section the acceleration capabilities of the numerical code will be demonstrated for the first time. The SWTS is simulated by a discontinuous drop in the solar wind velocity at a distance of 70 AU from the sun. Equation (5.1) was used to model the solar wind. Equations (5.2) and (5.3) were again used to represent the diffusion coefficients.

Except for the existence of the accelerating shock, the same boundary conditions that were used in Section 5.4 have again been used in this section. This re-acceleration approach is probably the most correct way to describe the modulation effect of the sun on galactic cosmic rays. Again, as in the previous case, the code was stopped at about 2.4 years model time because sufficient convergence was obtained.

The only major difference in the boundary conditions of the problem is the addition of an $f = 0$ condition at $P_{min}$, on the shock only. This cannot be avoided, since the Wendroff Implicit method used to solve the matching condition on the shock [see equation (4.81), Section 4.8] requires an initial condition in rigidity to operate.

If the results of this section, Figures 5.9 to 5.11, are compared with Figures 5.6 to 5.8 from the pure modulation model, it is apparent that re-acceleration causes significant and interesting changes from a pure modulation solution. These are:

- The spectra on the shock (Figures 5.9(a), 5.9(b), 5.9(c)) rise above the LIS due to the acceleration effect. This same effect is visible on Figure 5.10(e) for 1 GeV particles as function of radial distance. In general, all intensity spectra are somewhat higher than the pure modulation spectra of Figure 5.6 because of the re-acceleration.

- There is a sharp drop in intensity on the shock spectrum in Figure 5.9(a) at the lowest energies. This is a boundary effect, caused by the $f = 0$ initial condition at $P_{min}$ in the Wendroff Implicit scheme, with which the matching condition was solved. Since this boundary condition is only specified at the shock and not elsewhere, it is of no big concern, since the time-asymptotic solution away from the shock is not affected by it.

- The radial profiles in Figures 5.10(a)–(e) are quite different from those produced by the pure modulation model: At low energies, between the inner boundary and the shock, the profiles are very nearly the same as that of the pure modulation case, but with increasing energy the differences become quite large. At 1 GeV the shock intensity exceeds the LIS.
Figure 5.9: Energy spectra, radial and latitudinal gradients obtained with a re-acceleration model.
Figure 5.10: Normalized intensity as function of radial distance for different energies as obtained with a re-acceleration model.
Figure 5.11: Normalized intensity as function of polar angle for different energies as obtained with a re-acceleration model.
Jokipii et al. (1993) have found a similar effect. Beyond the shock the radial profiles are very interesting. At low energies the intensities rise with distance, following a concave structure. This structure is due to the effect of the outer boundary that is fairly near to the shock. The effect of the outer boundary on the solutions will be explored in Chapter 6. This sharp rise beyond the shock is quite different in profile from the skin-depth effect in the pure modulation case of Figure 5.7. However, it is of the same order of magnitude. A spacecraft will also have difficulty to identify a shock on this property alone. A sharp increase in intensity might be due to either the skin-depth effect associated with a modulation boundary, or a shock crossing.

- The inner radial gradient in Figure 5.9(e) is about 1.5 times larger than that produced by the pure modulation model. The shape of the curve (and, therefore, the explanation), however, stays much the same.

- The outer gradient differs by approximately the same amount, but it has a somewhat wider and flatter peak at about the same energy than in the pure modulation model. This effect can also be seen in the radial profiles where the gradient between 30 and 60 AU becomes large compared to the pure modulation case at lower energies and stays large at higher energies.

- Figures 5.11(a) to 5.11(e) show the polar profiles obtained from the re-acceleration model. The results are very similar to the pure modulation case. The largest difference is that the profile on the shock (70 AU) begins to bulge upwards about the ecliptic. This is due to the fact that acceleration in the solar wind termination shock is more effective about the ecliptic, where the magnetic field is more or less perpendicular to the normal on the shock face and less effective towards the poles where the shock becomes essentially parallel. This effect shows up much more strongly in the pure acceleration case (Section 5.5.2).

- At high energies, the latitudinal gradients become slightly negative ($\sim -0.2\%$/deg.) at energies $> 600$ MeV, due to the slight bulging effect in the polar intensity profiles at these energies.

Figure 5.11(f) shows that there are major differences in the convergence characteristics of the pure modulation and re-acceleration cases. The most noticeable difference is the 'kink' between 1000 and 2000 MeV. This is easily explained with the aid of the definition in equation 5.10: If $f_s > f_{s-1}$ then $\ln(f_s/f_{s-1}) > 0$ and if $f_s < f_{s-1}$ then $\ln(f_s/f_{s-1}) < 0$. In the pure modulation case the distribution function always decreases as it converges from the initial condition towards the time-asymptotic value and therefore $\ln(f_s/f_{s-1}) < 0$. The absolute value thereof could then be plotted on the logarithmic scale in Figure 5.8(f). In the re-acceleration case the situation is more complex. Below 1000–2000 MeV, the solution decreases with time as in the pure modulation case, but at the dip in Figure 5.11(f) this trend changed around. The acceleration process accelerates particles to higher energies, effectively increasing the solution values with time and, therefore, $\ln(f_s/f_{s-1}) > 0$ at these high energies.

The other major difference with the pure modulation case is the magnitude of $< \dot{\gamma}_t >$. The acceleration process continuously accelerates particles to higher energies. Since the high energy particles are continuously cooled at the same time, this causes the solution to change significantly over much longer time scales and would probably do so indefinitely. Therefore, it seems that the convergence of time-asymptotic solutions with an accelerating shock influencing the system will be much slower.

Apart from several interesting features, the solution shows that the re-acceleration of galactic cosmic rays by the SWTS does not have dramatic effects on the modulation, especially in the inner half of the heliosphere.
5.5.2 Acceleration of a Low-Energy Proton Source by the SWTS

In this section, only particles that are accelerated from very low (quasi-thermal) energies toward higher energies, uncontaminated by the galactic background, are studied. Therefore, the LIS on the outer boundary is put to zero \( f(r = r_b) = 0 \) and particles are injected at the shock at some low energy, \( P_i \).

In Chapter 6 it will be shown that the accelerated spectra are independent of the shape and the rigidity of the peak intensity of the source, as long as the source is steeper than \( P^{-4} \). Therefore, to save computer time, a mono-energetic source of the form

\[
Q = Q_* \delta(P - P_i) = (Q_*)_0 \delta(r - r_s) \delta(P - P_i)
\]

was injected on the shock at rigidity \( P_i = 0.1 \) GV. The magnitude was arbitrarily chosen as \((Q_*)_0 = 10^7 \) particles/GV/(AU)^2/S.

Furthermore, the distribution function was put equal to zero at both \( P_{\text{min}} = 0.09 \) GV (\( T = 5 \) MeV) and at \( P_{\text{max}} = 20 \) GV. The injection rigidity, \( P_i \), was chosen to be larger than \( P_{\text{min}} \) to avoid a boundary condition \( (f = 0) \) that conflicted with a finite source at the same rigidity.

The code was started with an empty heliosphere and was allowed to run until 3 years (1094 days) model time had elapsed. This longer time-span was necessary because convergence is generally slower in the acceleration case than in the pure modulation case.

Consider the spectra in Figures 5.12(a) to 5.12(d). The three lines in each frame are (from the bottom up) the spectra at 1 AU, 30 AU and the upper one is the spectrum established on the shock at 70 AU. The intensity in the lower-energy part of the shock spectra is approximately proportional to \( T^{-1} \gamma \) with \( 0.7 < \gamma < 0.8 \). According to the theory of diffusive shock acceleration, this should be \( j_T \propto T^{-1} \), since for non-relativistic particles this corresponds with spectra proportional to \( p^{-2} \) (see Section 3.3). This discrepancy is caused by the loss of accuracy in the Wendroff Implicit method operating on a relatively coarse rigidity grid. In Chapter 6 it will be demonstrated that the solution converges towards the desired \( T^{-1} \) energy dependence if the rigidity grid is refined.

Depending on heliolatitude, the shock spectra start to fall off rapidly beyond about 300–1000 MeV. This is due to the well-known curvature effect in the acceleration, when \( V_T/\kappa_{tr} \) becomes < 1 (Moraal and Axford, 1983; Drury, 1983). For the chosen parameters this occurs at 1660 MeV in the ecliptic plane and at 1070 MeV at 45°. At the poles \( V_T/\kappa_{tr} < 1 \) for all energies used in the simulation. Therefore, little acceleration takes place over the poles and the shock spectrum in Figure 5.12(d) was not formed there, but was established there from the neighboring latitudes through the process of diffusion. An additional effect can also be seen on the shock spectrum at 0°. Below the cutoff if clearly has a concave profile instead of \( T^{-1} \). This is a real effect, related to the polar boundary, and it is discussed in detail in Section 5.6.2.

The next set of figures (Figures 5.13(a) to 5.13(f)) shows the radial profiles at five energies. At low energies (10 and 30 MeV) the profiles in the inner heliosphere are very flat, resulting in relatively small radial gradients. Nearer to the shock the intensities rise sharply, giving rise to very large radial gradients. Beyond the shock the intensities stay constant because particles at these low energies are not affected by the outer boundary because their diffusion mean free paths are quite small.

With rising energy, the situation changes dramatically. Inside the shock, the flat part of the profile in the inner heliosphere disappears completely and the radial gradient becomes a weak function of radial distance. The outer boundary, where \( f = 0 \), starts to make its effect known. The profiles start to curve downwards towards the boundary.
Figure 5.12: Energy spectra, radial and latitudinal gradients obtained with a source acceleration model.
Figure 5.13: Normalized intensity as function of radial distance for different energies as obtained with a source acceleration model.
Figure 5.14: Normalized intensity as function of polar angle for different energies as obtained with a source acceleration model.
Both the inner and the outer radial gradients are much larger (about a factor 3) than in the preceding cases. This is expected since the radial gradients at high energies are approximately given by \( g_r = CV/\kappa_{rr} \) and since the Compton-Getting factor,

\[
C = -\frac{p}{3f} \frac{\partial f}{\partial p},
\]

is approximately three times larger than in the pure modulation case. The onset of the adiabatic limit is clearly seen in both the spectra at 30 AU in the ecliptic and in the radial gradients, where the outer radial gradient falls off very sharply towards lower energy.

The latitudinal gradients in the outer heliosphere become negative at \( T \gtrsim 200 \text{ MeV} \). Figures 5.14(d) and 5.14(e) show that this is due to the higher acceleration efficiency on the shock.

In the previous time-dependent runs, the temporal gradient decreased dramatically with increasing energy, signifying that the solution converges faster towards the time-asymptotic limit at higher energies. In this case, however, the temporal gradient rises with increasing energy as can be seen from Figure 5.14(f). This is not unexpected, since it is in accord with the acceleration process where a low-energy proton source is accelerated towards higher energies. In the low-energy part of the spectrum, the solution has already converged towards the \( J_T \propto T^{-\gamma} \) power law at early times, while the high energy tail of the spectrum is still changing dramatically due to the continuing acceleration process [see Figure 5.12(c)]. The decision on when to stop the code is therefore not so clear-cut as in the pure modulation case.

The preceding cases of pure modulation, re-acceleration and pure acceleration, are simple in the sense that Galactic cosmic rays and cosmic rays accelerated at the SWTS are considered separately. For more realism, all these effects should be combined into one model in order to get the whole picture. This brings us to the acceleration/re-acceleration case.

### 5.5.3 Combined Source Acceleration and Re-Acceleration of Protons by the SWTS

The SWTS will simultaneously accelerate particles from a low-energy source and re-accelerate any pre-existing interstellar cosmic-ray spectrum. Thus, in this section the acceleration solution of Section 5.5.2 and the re-acceleration solution of Section 5.5.1 are combined. This means that exactly the same parameters were used as in Section 5.5.2, except that the spectrum on the outer boundary at \( r_b = 100 \text{ AU} \) was changed from \( f = 0 \) to the standard interstellar spectrum (5.12).

The spectra produced by this model are shown in Figures 5.15(a) to 5.15(d). Comparing these spectra at 70 AU with those generated by the pure acceleration and re-acceleration models, one notices the following characteristics: On the low energy side of the spectrum the accelerated particles dominate the re-accelerated particles. At higher energies the effect of the LIS becomes larger, causing the spectrum to become harder and to start converging to the spectrum of the LIS with increasing energy. The energy region over which the effect of shock acceleration dominates depends on the magnitude of the source function, \( Q \). If \( Q \) is increased, shock acceleration will dominate up to very high energies, while if it is decreased, the solution will converge toward the re-acceleration case.

The spectra in the inner and middle heliosphere are more reminiscent of those produced by the re-acceleration model than those produced by the source acceleration model. A careful comparison reveals that below the intensity maxima, the spectra in the inner heliosphere follow the same energy dependence as the spectra of the source acceleration model, while they converge toward those produced by the re-acceleration model above the maxima. In Chapter 7 more attention
Figure 5.15: Energy spectra, radial and latitudinal gradients obtained with a combined source acceleration and re-acceleration model.
Figure 5.16: Normalized intensity as function of radial distance for different energies as obtained with a combined source acceleration and re-acceleration model.
Figure 5.17: Normalized intensity as function of polar angle for different energies as obtained with a combined source acceleration and re-acceleration model.
will be given to the conditions under which the effects of source acceleration are small or large in the proton spectra measured in the inner to middle heliosphere.

The radial profiles in Figures 5.16(a) to 5.16(f) clearly show that, at most energies the profiles are very similar to that of the source acceleration case. The only difference is the magnitude of the intensities, which are higher than in the source acceleration case because of different normalization points. In the present case, the intensities were normalized at the boundary as in the pure modulation and re-acceleration cases, while the source acceleration profiles were normalized on the shock at 90°.

Beyond the shock, the intensities are almost the same as those produced by the source acceleration model, except near the boundary where the solution falls to the LIS value and not to zero as in the source acceleration case. At the highest energies (1000 MeV) the profiles begin to assume some of the re-acceleration characteristics.

The polar profiles in Figures 5.17(a) to 5.17(c) are very similar to those produced by the source acceleration model. The only major difference is the magnitude of the intensities due to the differences in normalization.

The rate of change in the distribution function [Figure 5.17(f)] is now a combination of the re-acceleration and source acceleration case. In the low energy parts the effect of source acceleration dominates, but toward higher energies the effect of the LIS presented at the boundary dominates. The result is that \( \frac{\partial f}{\partial t} \) rises slightly with energy until about 1000 MeV, whereafter it falls off sharply.

5.5.4 A Linear Combination of Proton Source Acceleration and Re-Acceleration Solutions

A comparison of the results of the acceleration/re-acceleration case with those of the source acceleration and re-acceleration cases suggests that the acceleration/re-acceleration case is just a simple linear superposition of the re-acceleration and source acceleration cases.

This should, of course, be the case because the TPE is a linear PDE. Thus, if \( f_1 \) and \( f_2 \) are both solutions, then any linear combination,

\[
   f = af_1 + bf_2,
\]

is also a solution of the TPE.

The boundary conditions (and their numerical employment), however, are slightly different in the two cases. Thus, an important question is whether the source acceleration and re-acceleration solutions of Sections 5.5.1 and 5.5.2 may be simply added together to obtain the compound solution of Section 5.5.3. Such a linear combination was done in Figures 5.18(a) to 5.18(d).

Comparing these figures to Figures 5.15(a), (b) and (d) shows that, except for the differences at the high energy side, where the intensities are a little higher than in the acceleration/re-acceleration case, the spectra obtained in this way are almost identical to those obtained in the acceleration/re-acceleration case.

The differences at the high-energy side are due to the different boundary conditions imposed there. In the acceleration/re-acceleration case, the entire solution was pinned down on the high-energy boundary at the LIS value. In the combination, however, only the re-acceleration solution is pinned down at the LIS value and the small contribution of the source acceleration solution must be added to this.

This simple demonstration is not altogether trivial, because this linear combination must be used to study the modulation of all species, except protons. These other cosmic-ray species
Figure 5.18: Proton energy spectra obtained by a linear combination of a source acceleration and re-acceleration solution.

consist of fully-stripped nuclei with $A/Z = 2$, as well as singly-ionized atoms with $A/Z > 2$. Thus, except for Hydrogen, these two components are different species with different rigidities for the same kinetic energy per nucleon. Consequently, the modulation of such a species can only be studied by the linear combination of the re-acceleration of the fully stripped galactic component, and the source acceleration of the singly-ionized component.

5.6 Drift Solutions/Effects

In the previous sections, only the effect of $E \times B$-drift has been included in the solutions since this type of drift is implicit in the energy change term of the TPE,

$$\frac{1}{3} (\nabla \cdot V) \frac{\partial f}{\partial \ln p},$$

as was shown by Forman and Gleeson (1975). Curvature and gradient drift have, thus far, been ignored. These effects will now be included into the model solutions.
These drift effects, contained in the anti-symmetric part of the diffusion tensor, manifest themselves as drift velocities of particle guiding centers given by expression (3.8),

\[ \mathbf{v}_d = \nabla \times \kappa \mathbf{e}_B. \]

The radial and poleward components of this velocity are given by expressions (3.38) and (3.39) respectively. The concept of neutral sheet drift in this field was discussed in Section 3.8, and the neutral sheet drift velocity is given by (3.64).

For demonstration purposes, a flat neutral sheet, with the tilt angle \( \alpha = 0^\circ \), was assumed throughout this section. This results in the strongest possible drift conditions. To enhance the numerical stability of the model, the drift field was distributed over an angle \( \Delta \theta_{ns} = 4^\circ \) about the neutral sheet, as described in Section 3.8.

A full 22-year cycle consists of two 11-year magnetic states. In the so-called positive drift state the field in the Northern (Southern) hemisphere points away from (toward) the sun, with the field directions reversed in the alternate negative drift state.

During the positive drift cycle, drift effects generally augment diffusion since the drift and diffusive fluxes are in the same general direction in the largest part of the heliosphere. In the negative drift cycle, however, drift and diffusion tend to work in opposite directions, resulting in different cosmic-ray intensities in each of the two cycles.

5.6.1 The Effects of Drift on Modulation Solutions

This subsection presents the standard drift solution to a pure modulation problem, as was first done by Jokipii and Kopriva (1979a,b) and Moraal et al. (1979) and repeated in numerous solutions since then. The present solution contains nothing new or unique; the drift effects to be discussed are standard knowledge, and the solution is merely included for the sake of completeness.

Exactly the same model, with the same parameters as Section 5.4 was used, apart from the fact that drifts were included. Thus, the results must be evaluated in comparison with the equivalent no-drift results from Section 5.3 or 5.4.

The first set of results, shown in Figures 5.19 to 5.21, represents the positive drift cycle. The most striking, and well-known, features are:

- The general level of modulation is much smaller and, therefore, the modulated intensities are higher than in the equivalent no-drift case.
- Because of this small modulation, Figure 5.19(e) shows that the radial gradients generally are also much smaller.
- As expected from the small radial gradients, the latitudinal gradients calculated at 30 and 70 AU are very similar. Only the latitudinal gradient calculated at 1 AU is somewhat different from the other two.
- The radial and polar profiles are very similar to the no-drift profiles, except for the smaller gradients.
- From Figures 5.21(f) and 5.19(c) it can be seen that the solution converges much faster toward the time-asymptotic solution than in the no-drift case of Section 5.4.

The negative drift case is represented by the set of Figures 5.22 to 5.24, with the following noticeable features:
Figure 5.19: Energy spectra, radial and latitudinal gradients obtained with a time-asymptotic model for the positive drift case.
Figure 5.20: Normalized intensity as function of radial distance for different energies as obtained with a time-asymptotic model for the positive drift case.
Figure 5.21: Normalized intensity as function of polar angle for different energies as obtained with a time-asymptotic model for the positive drift case.
Figure 5.22: Energy spectra, radial and latitudinal gradients obtained with a time-asymptotic model for the negative drift case.
Figure 5.23: Normalized intensity as function of radial distance for different energies as obtained with a time-asymptotic model for the negative drift case.
Figure 5.24: Normalized intensity as function of polar angle for different energies as obtained with a time-asymptotic model for the negative drift case.
From Figure 5.22(e) it can be seen that the radial gradients are much larger than in the positive drift cycle. This is also apparent from the radial profiles in Figure 5.23 and the energy spectra in Figures 5.22(a) to 5.22(d).

Large latitudinal gradients, with values up to $-3.6\%$/deg., can be seen at low energies in Figure 5.22(f). The reason for this is immediately apparent from the polar angle intensity profiles in Figure 5.24, which show that the cosmic-ray intensities decrease away from the ecliptic plane.

At low energies, the effects of drift on the polar angle distributions are more pronounced. Figure 5.24(a) shows that the intensities fall off from the poles towards a minimum in intensity around 30°, whereafter they start to rise again towards a local maximum at the ecliptic, forming a sharp cusp-like structure there.

From Figures 5.24(f) and 5.22(c) it is seen that in this drift cycle the solution converges much slower to the time-asymptotic solution than in the no-drift case.

The qualitative explanation of these drift results is standard. In the positive drift cycle particles generally drift toward the ecliptic plane, thus augmenting the diffusive flux from pole to ecliptic. This leads to smaller latitudinal gradients. Because this equatorially directed flux effectively fills the region around the ecliptic plane with particles, this leads to smaller radial gradients in the ecliptic plane. In this positive cycle, the outward-directed current sheet flux about the ecliptic plane serves as an effective sink of particles in the system by transporting them out of the heliosphere. This causes the slightly sharper minimum intensity profile around the ecliptic than in the equivalent no-drift case.

In the negative drift cycle particles generally drift towards the poles, opposite to the diffusive flux. Thus, those particles that diffuse into the heliosphere in the polar regions, are inhibited to reach the ecliptic regions. Since diffusion and drift oppose each other, the convective flux starts to dominate at some intermediate latitude and particles will be transported out of the heliosphere at this latitude (Jokipii, private communication, 1992).

In the negative drift cycle the neutral sheet flux is directed radially inward. These particles tend to escape from the sheet by pitch angle scattering and then drift towards the poles. This inward flux of particles along the ecliptic current sheet causes a local maximum intensity about the ecliptic plane, leading to the negative latitudinal gradients. The outward, convection-dominated flux at the intermediate latitudes ($\sim 30°$) causes a minimum in the intensities in this region, giving the polar angle profiles their characteristic shapes.

Therefore, the essence of this section is the well-established phenomenon that drift aids the diffusive transport in the positive drift cycle, and opposes it in the negative drift cycle. This picture also explains why the positive drift solutions converge faster, and the negative drift solutions slower than in the no-drift case.

5.6.2 The Effects of Drift on Acceleration Solutions

This subsection repeats the source acceleration solution of Section 5.5.2, this time with the drift effects included. The only minor difference with the no-drift solution is that the upper boundary of the rigidity domain, $P_{\text{max}}$, was shifted down from 20 GV to 4 GV. This was needed to enhance the stability of the numerical solutions.

Figures 5.25 to 5.27 and Figures 5.28 to 5.30 show the results of the positive and negative drift cycles, respectively.
Figure 5.25: Energy spectra, radial and latitudinal gradients obtained with a source acceleration model for the positive drift case.
Figure 5.26: Normalized intensity as function of radial distance for different energies as obtained with a source acceleration model for the positive drift case.
Figure 5.27: Normalized intensity as function of polar angle for different energies as obtained with a source acceleration model for the positive drift case.
Figure 5.28: Energy spectra, radial and latitudinal gradients obtained with a source acceleration model for the negative drift case.
Figure 5.29: Normalized intensity as function of radial distance for different energies as obtained with a source acceleration model for the negative drift case.
Figure 5.30: Normalized intensity as function of polar angle for different energies as obtained with a source acceleration model for the negative drift case.
An amazing feature of these solutions is that the, by now familiar $j_T \propto T^{-1}$ spectral shape on the shock below the curvature cutoff, has disappeared. Instead, for the positive drift cycle, the intensity near the ecliptic plane drops more rapidly than $j_T \propto T^{-1}$; it then reaches an inflection point before it falls off into the high-energy cutoff region. At the poles this situation is reversed in the sense that the spectrum below the cutoff is almost independent of kinetic energy. Figures 5.28(a) and 5.28(d) show that in the negative drift cycle these spectral features are reversed. In this case the shock spectrum in the ecliptic is very flat, while the spectrum at the poles has the more complicated form. The intensities at 60° show that in both cases there is a gradual transition from one type of shock spectrum to the other.

These differences in spectral shape automatically lead to differences in intensities on the shock for the two drift cycles. In the negative drift cycle the intensity at the ecliptic regions is higher than for the positive cycle, while the situation at the poles is reversed. In short, drift effects have a major influence on the acceleration efficiency of the shock.

Figure 5.31: The shock spectra presented by Jokipii (1988).

Jokipii (1988) observed the same tendency in his earlier drift solutions. His calculated shock spectra are repeated in Figure 5.31. There are quantitative differences between the detailed spectral shapes of the two models, but these are due to different choices of parameters. As far as we can ascertain, Jokipii and co-workers did not present an explanation for this phenomenon.

To understand these shock spectra, one has to consider the polar angle distributions, shown in Figures 5.27(a)–(e) and 5.30(a)–(e). The profiles on the shock clearly show a sharp 'dip' ('peak') in the intensity around the ecliptic plane in the positive (negative) drift cycle. This feature, due to the neutral sheet drift, reaches a maximum height and width at intermediate energies of the order of 100 MeV. Thus, for the positive drift cycle, say, the 30 MeV and 100 MeV intensities near the ecliptic plane are depleted relative to the 10 MeV intensity, leading to the steep spectrum. In the negative drift cycle this same effect occurs at the poles. The flat spectra observed at the poles (ecliptic) in the positive (negative) drift cycle are due to exactly the opposite reason. Here, the inward drift flux along the polar line (ecliptic neutral sheet) tends to keep the intensity at intermediate energies high with respect to the intensity at lower energies causing the flat, hard spectra.

In terms of drift fluxes, these spectral effects are interpreted as follows. In the positive drift
cycle the neutral sheet transports particles radially out of the heliosphere, while shock drift transports them towards the poles. Since both these processes are proportional to $\beta P$, the more energetic particles are transported away, and are not available for substantial re-acceleration in the ecliptic regions of the shock. This leads to lower accelerated intensities with a softer spectrum. Since the particles that drift along the shock repeatedly cross it on their way to the poles, they will acquire additional energy through the process of shock-drift acceleration (Jokipii, 1987, 1992). These effects combine to produce the hard, flat shock spectra at the poles in the positive drift cycle. In the negative drift cycle the argument for both the ecliptic and polar regions is, evidently, reversed.

Köta and Jokipii (1994) recently gave a similar explanation for these effects. They solved the transport equation across a plane shock which intersects two current sheets a distance $d$ apart. Particles were allowed to drift along the shock face as well as along the current sheets, such that the drift along the current sheets is in opposite directions. They found that the resultant spectra were enhanced above the expected $j_T \propto T^{-1}$ near the current sheet that transports particles towards the shock in the downstream region. Near the current sheet that transports particles away from the shock in the downstream region, they found that the spectrum forms a 'dip'.

This situation of Köta and Jokipii is qualitatively applicable to that of the heliosphere when their upper neutral sheet is replaced by a polar 'neutral line' as described by Moraal (1993). In fact, it is not only drifts along this polar line that cause the effect. Referring back to Figure 5.12 of Section 5.5.2 it shows that in the no-drift situation the same effect occurs at the poles, and there it is due to rapid diffusion because of the Parker field geometry. Within about 1° from the poles the effective radial diffusion rapidly increases from kappa perp. to kappa parallel, thus transporting the particles rapidly away from the shock before substantial acceleration can occur. This leads to the softer spectra at the poles.

We note in passing that there may be a sign error in the Köta and Jokipii paper. This can be rectified when the first word, 'electrons' in the caption of their Figure 2, is replaced by 'positrons'.

Turning to the spatial intensity distributions, it is noticeable that they differ only quantitatively from the equivalent no-drift distributions, and these differences are best quantified by the radial and latitudinal gradients. The inner radial gradient in the positive drift cycle [Figure 5.25(e)] is similar to the no-drift gradient but smaller, and at the high energy end it does not drop off as fast with increasing energy. The outer gradient is almost independent of kinetic energy, with the low-energy drop-off towards the adiabatic limit absent. This is apparently due to the large number of low-energy particles in the soft ecliptic spectrum as discussed above.

In the negative cycle the radial gradients are considerably larger than in the positive cycle, which is a well-known drift effect, and their energy dependence is nearer to that of the no-drift case. Figure 5.25(f) shows that the latitudinal gradients between 90° and 60° at 1, 30 and 60 AU in the positive drift cycle are all strictly positive over the entire energy interval. This is an obvious consequence of the polar angle intensity profiles between 90° and 60°. For the negative drift cycle Figure 5.28(f) shows that the latitudinal gradients are all strictly negative over the entire energy domain. The magnitudes of these negative gradients are also quite large, especially the one at 60 AU.

Finally, as was the case in the pure modulation-with-drift solution, Figures 5.27(f) and 5.30(f) indicate that the positive drift cycle aids convergence of the numerical solution to its asymptotic values, while the negative drift cycle hampers it.
5.7 The Acceleration of Cosmic-Ray Oxygen by the SWTS

As stated in Section 5.5, we wish to use the numerical code developed in Chapter 4 primarily to address the problem of the origin of ACR. Up to now only acceleration solutions of protons, with $A/Z = 1$, have been produced with this model. In this section the re-acceleration and acceleration of heavier elements will be demonstrated. The only distinguishing factor between different elements is in the mass-to-charge ratio, $A/Z$. For protons $A = 1$ and $Z = 1$ and, therefore, $A/Z = 1$. For all other fully stripped nuclei $A/Z \approx 2$, with small variations possible due to isotopic composition (especially for the heavier elements). Thus, the TPE will yield the same solution for all nuclei heavier than Hydrogen if the same source and/or input functions are used.

In the case of ACR all the particles are singly ionized ($Z = 1$) and thus the ratio $A/Z = A$ for all species. Since singly-charged Oxygen, with $A/Z = 16$, has the best observed anomalous component, specific reference will be made towards it.

5.7.1 Solutions for Heavier Species

In Figure 5.32 the results of a re-acceleration run with $A/Z = 2$, a LIS given by equation (5.12) and the same parameters as in Section 5.5.1 are shown. Since the radial and polar angle distributions contain no new information, they were omitted and only the spectra, radial and latitudinal gradients are shown.

The major difference between the re-acceleration case for protons (Section 5.5.1) and this case for $A/Z = 2$ (fully-ionized Helium, Oxygen, etc.) is that features have shifted towards lower energies. For instance, it can be seen from Figures 5.32(a) and 5.9(a) that the cross-over between the LIS and the spectrum on the shock has shifted from the region between 400-500 MeV to 300-400 MeV.

According to Cummings et al. (1984) it was Fisk who first suggested that if the input spectra at the modulation boundary are similar, then features (such as spectral peaks) observed in proton spectra should occur at an energy for each species for which the diffusion coefficient, $\kappa$, is the same. Cummings et al. used this characteristic in the observed spectra of anomalous He, C, N, O, Ne and Ar to argue that anomalous cosmic rays are singly-ionized ions. Figure 5.33 (from Cummings and Stone, 1990) shows the spectra of different anomalous species, scaled in energy and intensity to the Oxygen spectrum with a $\chi^2$ hypothesis testing technique.

In Figure 5.34 the spectral results of an acceleration solution with $A/Z = 16$ are shown. The particles were injected at $P_i = 0.4$ GV since singly-ionized Oxygen has the same $\beta P$ at this rigidity than protons at $P_i = 0.1$ GV. Using the same considerations, the energy domain was changed to $P_{\text{min}} = 0.3$ GV and $P_{\text{max}} = 16$ GV. All the other parameters were the same as in Section 5.5.2. From these spectra two major features can be observed: Firstly, features shift in energy by an approximate factor of 16 and secondly, the intensities at corresponding features in energy are about $256 = 16^2$ higher than in the case of the solutions in Section 5.5.2.

Together with the results from Cummings and Stone (1990), this suggests that it may be possible to approximately obtain a model solution for singly-ionized Oxygen by taking the proton solution, scaling the energies down by a factor of 16 and multiplying the intensities by a factor of $16^2$.

In this chapter the diffusion coefficients, given by equations (5.2) and (5.3), are proportional to $\beta P$ and, therefore, the spectral features should occur at energies where $\beta P$ is the same. More generally, the diffusion coefficients may be considered to be proportional to $\beta P^\gamma$. Working with
Figure 5.32: Energy spectra, radial and latitudinal gradients obtained with a re-acceleration model for fully ionized Oxygen ($A/Z = 2$). The unit of the intensity is part/m$^2$/s/sr/(MeV/nuc).
Figure 5.33: Voyager 2 energy spectra of ACR He, C, N, Ne, and Ar scaled to the ACR O spectrum for the period 1987/105–313 (Cummings and Stone, 1990).
Figure 5.34: Energy spectra, radial and latitudinal gradients obtained with a source acceleration model for singly-ionized Oxygen ($A/Z = 16$). The unit of the intensity is part/m$^2$/s/sr/(MeV/nuc).
this energy dependence, we can write $\beta$, $P$ and $\beta P^\gamma$ as function of kinetic energy, $T$. Assume that $E_0$ is the rest mass energy of a proton, then the particle velocity and rigidity become

$$\beta = \frac{v}{c} = \frac{Am_0vc}{m_0c^2} = \frac{\sqrt{T(T + 2E_0)}}{T + E_0},$$

(5.15)

and

$$P = \frac{pc}{Z} = \frac{A}{Z}\sqrt{T(T + 2E_0)},$$

(5.16)

where we have used $A^2E^2 = A^2(T + E_0)^2 = p^2c^2 + A^2m_0^2c^4$ with $m_0$ the rest mass of a proton. With this, the energy dependence of the diffusion coefficient becomes

$$\beta P^\gamma = \frac{(A/Z)^\gamma [T(T + 2E_0)]^{(\gamma+1)/2}}{T + E_0} = \frac{\alpha^\gamma[T(T + 2E_0)]^{(\gamma+1)/2}}{T + E_0}$$

(5.17)

with $\alpha = A/Z$. In the non-relativistic limit ($T \ll E_0$) this becomes

$$\beta P^\gamma \approx \alpha^\gamma E_0^{(\gamma-1)/2}(2T)^{(\gamma+1)/2}$$

while in the relativistic limit

$$\beta P^\gamma \approx \alpha^\gamma T^\gamma.$$  

According to Fisk's statement, the spectral peaks of two different species (denoted by subscripts 1 & 2) will occur at energies where the diffusion coefficients are the same: $\kappa_1 = \kappa_2$ or $(\beta P^\gamma)_1 = (\beta P^\gamma)_2$. Using equation (5.17) this becomes

$$\frac{T_2}{T_1} = \left(\frac{T_1 + 2E_0}{T_2 + 2E_0}\right) \left(\frac{T_2 + E_0}{T_1 + E_0}\right)^{2\gamma\gamma+1} \left(\frac{\alpha_1}{\alpha_2}\right)^{2\gamma\gamma+1}. \quad (5.18)$$

In the non-relativistic limit (5.18) reduces to

$$\frac{T_2}{T_1} \approx \left(\frac{\alpha_1}{\alpha_2}\right)^{2\gamma\gamma+1}, \quad (5.19)$$

while in the relativistic limit (5.18) reduces to

$$\frac{T_2}{T_1} \approx \frac{\alpha_1}{\alpha_2}. \quad (5.20)$$

The ratio $T_2/T_1$ can be interpreted as a scaling factor with which the energy scale of species 1 must be multiplied to obtain the spectra of species 2.

An analysis with the general scaling factor (5.18) is in principle quite difficult, since it is transcendental in $T_2/T_1$. Moreover, the energy dependence of the diffusion coefficients is quite complicated since we have a transition from a simple non-relativistic limit to another simple relativistic limit (see Figure 5.35). However, for $\gamma = 1$ the scaling factor $T_2/T_1$ has the same value in the two limits. This can be seen from Figure 5.36 which shows the ratios $T_H/T_O$, $T_H/T_{He}$, and $T_O/T_{He}$ as function of proton kinetic energy.

In their analysis, Cummings and Stone (Cummings et al., 1984; Cummings and Stone, 1987, 1990) used data measured by the Voyager spacecraft during the 1977 solar minimum period. Figure 5.37 shows that the energy scaling factors $(T/T_O)$ obtained by Cummings and Stone in scaling the spectra of He, C, N and Ne to that of O in Figure 5.33 are well correlated with $A/Z$. 

132
The extrapolation to Ar was done by Cummings and Stone and we added the extrapolation to Hydrogen. The least squares fit that we employed to fit the data was of the form

\[ \frac{T}{T_0} = a \alpha^b = a(A/Z)^b \]

with coefficients \(a = 33.63\) and \(b = -1.24\). If compared with (5.19), this fit suggests that \(\gamma = 1.63\). If this is so, then the modulation process is strongly \(P\)-dependent. Either \(\kappa_{||}\) or \(\kappa_{\perp}\) or both seem to have a stronger rigidity dependence than \(\beta P\). The possibility that this strong energy dependence can be a drift effect, where \(\kappa_T\) has a stronger rigidity dependence, is also not excluded. The conclusion that \(\gamma \sim 1.6\) is also in accord with results from quasi-linear theory which suggests an energy dependence for \(\kappa_{||}\) as high as \(\beta P^{1.5}\) (e.g., Bieber et al., 1994).

### 5.7.2 Application to Model Results

For our diffusion coefficients, the parameter \(\gamma = 1 \ (\kappa \propto \beta P)\) and the energy scaling factor (5.19) is \(T_2/T_1 \approx \alpha_1/\alpha_2\). Figure 5.38 shows the ecliptic spectra for protons \((\alpha = 1)\) and anomalous Helium \((\alpha = 4)\) scaled to anomalous Oxygen energies by using a scaling factor of \(\alpha^{-1}\) in energy and a scaling factor \(\alpha^2\) in intensity. The three spectra agree remarkably well. The small differences are mainly due to the finite number of rigidity nodes in the numerical code, which makes it difficult to run the code at exactly correctly spaced values of \(\beta P\). Referring back to Figure 5.36 shows that the deviations at the highest energies occur because one goes out of the non-relativistic limit.

To understand this \(\alpha^2\) scaling in intensity, consider the TPE for the omni-directional distribution function,

\[ \frac{\partial f}{\partial t} + \cdots = Q(p). \]
Figure 5.36: Scaling factor, against energy for scaling from H\(^+\) to He\(^+\) and O\(^+\) as well as scaling from He\(^+\) to O\(^+\).

As was shown in Section 3.2 this is the same as

\[
\frac{\partial U_p}{\partial t} + \cdots = 4\pi p^2 Q(p),
\]

in terms of the differential number density, \(U_p \propto 4\pi p^2 f\). To obtain the number density, \(N\), this equation must be integrated over all momenta to obtain

\[
\frac{\partial N}{\partial t} + \cdots = \int_0^\infty 4\pi p^2 Q(p) \, dp.
\]

The integrand on the right-hand side can be written in terms of rigidity instead of momentum:

\[
4\pi p^2 Q(p) = 4\pi \left(\frac{Z}{c}\right)^2 P^2 Q(P) \frac{dP}{dp} = 4\pi \frac{Z}{c} P^2 Q(P) = 4\pi \frac{Z}{c} P^2 Q_0 \delta(P - P_1).
\]

Integrated over rigidity, this gives

\[
\int_0^\infty 4\pi P^2 Q_0 \delta(P - P_1) \, dp = 4\pi P_1^2 Q_0.
\]

Since protons and singly-ionized Oxygen are injected at different rigidities, it is apparent from (5.22) that \(Q_0\) should also be changed to ensure that the same number of particles is injected into the system:

\[
Q_0 P_{i1}^2 = Q_2 P_{i2}^2.
\]
Figure 5.37: Energy scaling factor, $T/T_0$, versus $A/Z$, (or $\alpha$). The solid line is a least squares fit of the non-relativistic approximation to the He, C, N, O and Ne data. The dashed lines are extrapolations to Ar and H. The dotted line is the relativistic approximation.

or

$$\frac{Q_{02}}{Q_{01}} = \left( \frac{P_{i_1}}{P_{i_2}} \right)^2$$  \hspace{1cm} (5.23)

where the subscript ‘1’ denotes Hydrogen and ‘2’ denotes Oxygen.

The injection rigidity for singly-ionized Oxygen, $P_{i_2}$ was chosen such that the diffusion coefficients for Hydrogen and Oxygen have the same value (the same $\beta P$), i.e.,

$$\beta_{i_1} P_{i_1} = \beta_{i_2} P_{i_2}.$$  

This can be written as

$$\frac{P_{i_1}^2}{\sqrt{P_{i_1}^2 + \alpha_1^2 E_0^2}} = \frac{P_{i_2}^2}{\sqrt{P_{i_2}^2 + \alpha_2^2 E_0^2}}$$

or

$$\left( \frac{P_{i_2}}{P_{i_1}} \right)^2 = \frac{P_{i_1}^2 + \sqrt{P_{i_1}^2 + 4[P_{i_1}^2 + \alpha_1^2 E_0^2]E_0^2}}{2[P_{i_1}^2 + \alpha_1^2 E_0^2]}.$$  \hspace{1cm} (5.24)

At the injection rigidity both species are non-relativistic and therefore $P_{i_1} \ll \alpha_1 E_0$ and $P_{i_1} \ll \alpha_2 E_0$ and (5.24) becomes

$$\left( \frac{P_{i_1}}{P_{i_2}} \right)^2 = \frac{\alpha_1}{\alpha_2}.$$  

The ratio, (5.23), can now be written as

$$\frac{Q_{02}}{Q_{01}} = \frac{\alpha_1}{\alpha_2}.$$  \hspace{1cm} (5.25)
This means that if $Q_{01} = Q_{02}$, and particles are injected such that $\beta P$ is the same for both Oxygen and protons, a factor of $A$ (16) times more Oxygen ions than protons will be injected into the system.

Furthermore, the intensities are plotted in terms of number/(MeV/nucleon) = (number/MeV)nucleon = (number/MeV)×$A$ and not number/MeV. This results in another factor of $A$ (16) difference between the proton and singly-ionized Oxygen cases. Together with the source normalization difference this causes a total difference of $A^2 = 16^2 = 256$ in intensity.

**5.7.3 A Linear Combination**

In Section 5.5.4 it was shown that a simple linear combination of a source acceleration and a re-acceleration run for protons forms a valid approximation for the physical problem where source acceleration and re-acceleration occur simultaneously. This demonstration can only be done for protons, because both the galactic and anomalous species have $A/Z = 1$. Figure 5.39 shows a linear combination of an acceleration run with singly-ionized Oxygen, $A/Z = 16$, and a re-acceleration run with Oxygen nuclei, $A/Z = 2$ (galactic oxygen), to obtain composite spectra in the ecliptic plane, over the polar regions and at a polar angle of 45°.

At energies above 300 MeV this composite solution gives the spectrum of galactic Oxygen as calculated with a re-acceleration model. At 70 AU, on the shock, the intensities are higher than the LIS for most part of this energy range. This is due to the re-acceleration of the galactic Oxygen nuclei at the solar wind termination shock. At the high-energy boundary, the intensity converges towards the LIS value due to the boundary condition imposed there. At the lower part of this energy range ($\sim 200$–$300$ MeV) the intensity curves upward into the soft spectrum that is typical of a source acceleration solution, well into the curvature cutoff region. Below this
Figure 5.39: Oxygen energy spectra obtained by a linear combination of a source acceleration (for $A/Z = 16$) and re-acceleration (for $A/Z = 2$) solution. The unit of the intensity is part/m$^2$/s/sr/(MeV/nuc).

point the solution is dominated by singly-charged heliospheric Oxygen.

Towards the inner heliosphere this spectrum on the shock gets modulated by the well-known processes discussed in Sections 5.3, 5.4 and 5.5. Despite the huge amount of modulation at low energies, the intensity at Earth abruptly starts to rise between 200-300 MeV to levels far above that of the LIS, forming what is known as the anomalous component of cosmic rays.
Chapter 6

Limitations and Properties of the Solution

6.1 Introduction

In Chapter 5 the basic properties of the newly-developed numerical solution of the cosmic-ray transport equation were described. Before this model can be applied widely to cosmic-ray transport in the heliosphere, there are several properties and limitations of the solution that must be discussed.

Before any model can be applied to a specific field of research, the limiting assumptions under which it was developed must be taken into consideration to avoid applying the model to systems in which it is not valid. Applied to the TPE, one must understand that this equation applies only to the transport of energetic particles (cosmic rays) in the diffusive limit which is based on the assumptions that there exists sufficient scattering by some process to keep the distribution function nearly isotropic, and the individual particles have random speeds substantially larger than any background fluid motion.

The TPE, given by \( (3.12) \), for example, cannot be applied to systems where the background plasma flow velocity is of the same order as the mean particle velocity. An example where this is the case is inside the termination shock radius of a pulsar wind where the upstream plasma flow velocity is relativistic \( (V \leq c) \).

In conjunction with such limitations, another set of limitations is inherited from the numerical method used to solve the TPE. Moreover, the limitations of computer accuracy, memory and speed are also inherently present in the solutions produced by the model. This set of numerical limitations will be discussed in the first part of this chapter, in Section 6.2, by probing the upper and lower limits of the numerical code.

Section 6.3 is devoted to parameter variations. A standard set of heliospheric parameters was used for demonstration purposes in Chapter 5. Physical parameters like the diffusion tensor, the extent of the heliosphere and the position of the shock will be varied in this chapter to gain insight into the properties of the model.

6.2 Limitations of the Model

In this section the limitations of the numerical model, arising from domain parameters such as grid spacing and domain boundaries, will be discussed. This discussion will be presented in three parts: The conditions under which numerical instabilities arise will be discussed first. This will be followed by a discussion of the parameters affecting the accuracy of the model and, finally, a section will be presented in which the mesh considerations, arising from the preceding two sections, will be discussed.
6.2.1 Theoretical Stability

Because of the compound nature of the acceleration model, as described in Sections 4.7 and 4.8, a straightforward stability analysis, such as the von Neumann or Matrix analysis, on the composite technique is difficult, if not impossible. A von Neumann stability analysis, for instance, depends on the ability to write an operator splitting technique (ADI, LOD, etc.) as two or more discretized equations into which a solution in the form

\[ e^{\gamma t} e^{i\beta z} \]  

(6.1)

must be substituted as an initial condition. In terms of the variables defined in Section 4.7, \( t \) is the time coordinate, \( x \) either the radial (r), polar (\( \theta \)) or rigidity (ln P) coordinate, and \( i \) denotes the complex number \( \sqrt{-1} \). Expression (6.1) contains an error in the form of an oscillation at the \( t \)-level given by \( e^{i\beta z} \). Generally, \( \gamma = \gamma(\beta) \) is complex and \( \beta \) real.

The condition under which this initial error will be damped as \( t \) increases is given by

\[ |e^{\gamma \Delta t}| = |\xi| \leq 1 \quad \forall \gamma \]

with \( \xi \) the so-called amplification factor. In this form, the stability condition is known as the von Neumann stability condition (see Section 4.4.4). This inequality, applied to the result of the initial substitution into the three component discretized equations, yields an amplification factor for each of these equations. The product of these amplification factors gives the amplification factor for the whole scheme (Lapidus and Pinder, 1982).

For some ADI methods, like the Peaceman-Rachford method, the component equations are conditionally stable, but due to a symmetry between the amplification factors the combined two-dimensional scheme is unconditionally stable. An expansion to three dimensions breaks this symmetry, resulting in a conditionally stable Peaceman-Rachford technique in three dimensions.

In the case of a LOD split, the component equations, solved with the unconditionally stable Crank-Nicolson implicit method of Section 4.4.1, are always unconditionally stable. Therefore, any product of the component amplification factors will yield an amplification factor that implies unconditional stability for the LOD scheme, even with expansion to higher dimensions.

Applied to our problem, a solution of the form (6.1) must be substituted into the discretized form of all three the component equations of the TPE to determine under which conditions this error will be damped. The radial and polar equations, (4.57) and (4.58), have been discretized with the Crank-Nicolson implicit method of Section 4.4.1, yielding the discretized forms (4.61) and (4.66). Since the Crank-Nicolson method is unconditionally stable (DuChateau and Zachmann, 1986:127; Lapidus and Pinder, 1982:160), assume, for illustration's sake, that the radial equation, (4.61), has an amplification factor \( |\xi_r| \leq 1 \), and the polar equation, (4.66), has an amplification factor \( |\xi_\theta| \leq 1 \).

The energy equation (4.59), however, is solved by quasi-analytical means as described in Section 4.7.3. The expression (6.1) substituted into (4.76) yields an amplification factor \( |\xi_F| = 1 \), indicating neither growth nor damping of the introduced oscillation. Since (4.76) can not be regarded as a discretized form of (4.59), this result is somewhat suspect. However, there is no reason to expect instabilities arising from (4.76) during the calculations in this part of the code. To conform with the other two components, we should define a hypothetical amplification factor, \( |\xi_F| = 1 \). Any errors arising in this part of the solution technique will be damped with an increasing time coordinate by the Crank-Nicolson methods used in the other parts. Therefore, the combined amplification factor, \( |\xi| = |\xi_r\xi_\theta\xi_F| \leq 1 \), and the method should be stable. On the other hand, there is no proof of this.
With the addition of the Wendroff method to solve the continuity condition on the shock, the situation gets more complicated. The Wendroff method has an amplification factor that implies unconditional stability (DuChateau and Zachmann, 1986:147), i.e., an amplification factor \(|\xi_W| \leq 1\). Even if one makes an assumption about \(\xi_P\), it is unclear how to combine the product \(\xi_{c_{\alpha}}\xi_{c_{\beta}}\xi_P\) and \(\xi_W\) to obtain an overall amplification factor. If it is a simple multiplication like \(\xi = \xi_{c_{\alpha}}\xi_{c_{\beta}}\xi_P\xi_W\), the whole method may be stable, but in the case of a linear combination, \(\xi = \alpha\xi_{c_{\alpha}}\xi_P + b\xi_W\), the overall amplification factor may not be smaller than or equal to one.

Therefore, such theoretical reasoning about the overall stability is suspect. Thus, we turn to empirical methods to obtain a ‘feeling’ of the method’s stability. Before doing so from Section 6.2.3 onwards, we also discuss the aspect of the solution’s accuracy.

### 6.2.2 Theoretical Accuracy

Finite difference methods are based on truncations of the Taylor series as shown in Section 4.3.2. The concept of accuracy is directly related to the truncation errors of these Taylor series. In this section we refer to the accuracy of methods as first or second order, where these terms have the meanings as explained in Section 4.3.2.

The Crank-Nicolson implicit method, applied to a model PDE of the form

\[ u_t = u_{xx}, \]

is accurate to second order in both \(t\) and \(x\). This, however, holds only for this specific model equation (Lapidus and Pinder, 1982:160). Therefore, with the addition of a first-order term and position-dependent coefficients, the resulting accuracy may actually be less than second order in \(x\). A Crank-Nicolson based LOD will inherit all the characteristics of the Crank-Nicolson method, and will also be accurate to second order when applied to a simple multi-dimensional model equation with coefficients one (Lapidus and Pinder, 1982:157,273).

In the present LOD implementation, the radial and polar equations are both solved with the Crank-Nicolson method and are, therefore, accurate to second order in \(r\) and \(\theta\). The energy equation, however, is solved by quasi-analytical means. This involves the backtracing of an analytically calculated characteristic. The accuracy of this process is limited only by the machine-precision parameter in the computer. Unfortunately, the method loses accuracy in the interpolation process, in the last stage of the correction for energy as described Section 4.7.3. As it is, this linear interpolation technique requires most of the energy equation’s calculation time and, therefore, a second-order interpolation technique would be too costly. Hence, its accuracy is at least of order one, but it will probably not have second-order accuracy as the other two equations.

The Wendroff Implicit method, used to solve the continuity equation on the shock (3.44), is accurate to second order in both \(\ln P\) and \(\theta\) (DuChateau and Zachmann, 1986:147). Coupled with the modified LOD scheme, the general accuracy of the solution (obtained by intuitive means, since none other is available) is as follows: The solution is accurate to second order in both \(r\) and \(\theta\). The accuracy in time is between first and second order, due to the deviation from the model parabolic equation with coefficients one, and the accuracy in \(\ln P\) is of second-order accuracy on the shock and of larger than or equal to order one \((\geq O(\Delta \ln P))\) elsewhere.

This is a vague conclusion, and from a theoretical point of view, one cannot decide what the overall accuracy of the solution is. Thus, since both the stability and accuracy cannot be fully calculated, the subsequent part of this section tests the solution by purely empirical means.
6.2.3 The Control Solution

A pure acceleration no-drift solution for protons is used as a reference solution, against which all subsequent solutions are compared. Figures 6.1 to 6.3 show a complete representation of this so-called control set of solutions. The format is the same as was used in Chapter 5. A detailed explanation of this format is found in Sections 5.3 and 5.4, and will, therefore, not be repeated here. The set of parameters described in Section 5.5.2 was used to obtain this set of solutions. However, coarser rigidity and time grids were used to minimize the computer runtime needed for this rather extensive search for instabilities and inaccuracies.

The rigidity domain was discretized with a grid containing \( l = 88 \) nodes, which implies \( \Delta \ln P = 0.062 \). The time step was put equal to 5.2 hours and the code was started and allowed to run for 5040 time steps, i.e., a total model time of about 3 years. All other aspects of the grid are discussed in the subsequent sections.

Henceforth, the results shown in Figures 6.1 to 6.3 will be referred to as the 'Control'.

6.2.4 The Radial Grid and Domain

The radial grid in the present model is not a simple linear grid but a transformed grid with uneven grid spacing. If the position in the grid is determined by the index, \( i = 1, 2, \ldots, n \), the grid is given by

\[
 r_i = \frac{r_{ib} - r_s}{2} + \frac{r_s - r_{ib}}{2} \frac{\tanh \left( b_1 \left( i - \frac{n+1}{2} \right) \right)}{\tanh \left( b_1 \frac{n-1}{2} \right)} + r_s + \frac{e^{b_2(i-n_s)} - 1}{e^{b_2(n-n_s)} - 1} \tag{6.2}
\]

where \( r_{ib} = r_1 \) gives the inner boundary of the heliosphere, \( r_b = r_n \) the outer boundary, \( r_s = r_{ns} \) the position of the shock, and \( i = n_s \) the grid index of the shock. This grid transformation results in a fine grid in the inner heliosphere and on both sides of the shock. This is needed because mathematical limits are approximated at the shock with second-order difference formulae. The transformation gives large radial steps in the middle heliosphere, and far beyond the shock, where the grid steps increase exponentially. To specify a grid in a specific model heliosphere with a predetermined inner and outer boundary, and a termination shock in between, one needs to specify both the parameters \( n \) and \( n_s \). The coefficients \( b_1 \) and \( b_2 \) were chosen such that both \( r_i \) and its derivative with respect to \( i \) are continuous at the shock.

The Control, as well as all the solutions of Chapter 5, uses 141 radial grid points, with the shock placed at grid point 101. If a linear grid is used, this gives intervals \( \Delta r = 0.7 \) AU in a heliosphere with boundary radius \( r_b = 100 \) AU. This is too coarse to handle the strong radial gradients, as well as their rapid change across the shock. In the Control, \( b_1 = 0.054 \) and \( b_2 = 0.334 \), such that the two intervals on both sides of the shock are \( r_{ns} - r_{ns-1} = 0.035 \) AU and \( r_{ns+1} - r_{ns} = 0.031 \) AU. This successfully approximates the limits

\[
 \lim_{r \to r_{ns}} \frac{\partial f}{\partial r} = \left( \frac{\partial f}{\partial r} \right)^- \\
 \lim_{r \to r_{ns}} \frac{\partial f}{\partial r} = \left( \frac{\partial f}{\partial r} \right)^+
\]

in the matching condition (3.44). To achieve this with a linear grid would require \( \sim 3000 \) grid points, increasing the runtime by a factor of 22, and the computer RAM usage 44 times. Runs with a linear grid spacing, up to the limit of the storage capacity, produce no intelligible results.
Figure 6.1: Energy spectra, radial and latitudinal gradients for protons obtained with a source acceleration model used as a Control. The dotted line in Figure (a) indicates a $T^{-1}$ slope, while the vertical dashed line marks the curvature cutoff point for acceleration, where $V_{r}/\kappa_{rr} = 1$. Except where stated differently, the intensity ($J_T$) is in the units of particles/m$^2$/s/sr/MeV.
Figure 6.2: Normalized intensity as function of radial distance for different energies as obtained with a source acceleration model used as a Control.
Figure 6.3: Normalized intensity as function of polar angle for different energies as obtained with a source acceleration model used as a Control.
Our experience is that a radial grid with a ratio $n:n_s$ given by 141:101 (101 grid points from the inner boundary of the heliosphere to the termination shock and 40 grid points from the shock to the boundary) produces reliable solutions, as can be seen from the Control case and the results in Chapter 5. Figure 6.4 shows the results of a run with a coarser radial grid with $n:n_s = 71:51$. This model suffers no loss of stability, but the solutions are of considerably less accuracy. The spectra and radial profiles in the ecliptic reveal negative radial gradients in the inner heliosphere at all energies up to at least 1 GeV. This effect, however, is most pronounced at low energies where the negative radial gradients persist even in the outer heliosphere, causing the ecliptic spectrum at 30 AU to fall below the ecliptic spectrum at 1 AU, as is shown in Figure 6.4(a). From the radial profiles in Figures 6.4(b) and (c), it is apparent that these negative radial gradients exist in the solutions even as far as 30° away from the ecliptic plane at the lower (< 100 MeV) energies.

Figure 6.4: The effects of a coarse radial grid with $n = 71$ and $n_s = 51$ on the no-drift source acceleration solutions, to be compared with the Control of Figures 6.1 and 6.2, which has $n:n_s = 141:101$. 

145
6.2.5 The Polar Grid and Domain

Like the radial grid, the polar grid can also be transformed. We employed a transformation such that the grid spacing is fine about the poles and the ecliptic:

\[ \theta_j = \begin{cases} \frac{\pi}{4} \left[ 1 + \frac{\tanh(a(j-m-3))}{\tanh(a(m-1))} \right] & \text{if } j \leq \frac{m+1}{2} \\ \frac{\pi}{4} \left[ 3 + \frac{\tanh(a(4j-3m-1))}{\tanh(a(m-1))} \right] & \text{if } j > \frac{m+1}{2} \end{cases} \]  

(6.3)

with \( j = 1, 2, \ldots, m \). The second part of this transformation was not used explicitly, because the solution was limited to \( 0 \leq \theta \leq \pi/2 \), with a symmetric boundary condition \( (\theta f/\partial \theta = 0) \) at \( \theta = \pi/2 \), given by (4.68) with \( w_4, w_5 \) and \( w_6 \) assuming the values described in Section 5.2. The solution was then mirrored around \( \theta = \pi/2 \). Nevertheless, the ability to solve the TPE from \( \theta = 0 \) to \( \theta = \pi \) was built into the model in the form of a switch which enables the second half of (6.3).

Such a transformed grid has fine spacing at the poles and in the ecliptic. At the poles this fine spacing is potentially useful because of the rapid change in diffusion coefficient with \( \theta \). At \( r = 60 \) AU, for example, the effective radial diffusion coefficient, \( \kappa_{rr} \), given by (5.13), changes from essentially parallel diffusion to perpendicular within one degree off the poles. At the ecliptic, fine grid spacing is potentially useful to handle the large neutral sheet drift, together with its strong \( \theta \)-dependence, given by expression (3.64).

This transformation was found to be less important than the radial transformation, and the Control, as well as all the solutions of Chapter 5 have \( a = 0 \) to produce a linear grid. The value of \( m \) in all these solutions was \( m = 61 \) (or 31 grid points in a quarter heliosphere run from \( \theta = 0 \) to \( \theta = 90^\circ \)) to produce a \( \theta \)-grid with steps of \( \Delta \theta = 3^\circ \).

The left column of Figure 6.5 shows a no-drift run on a coarser grid, with \( m = 41; \Delta \theta = 4.5^\circ \), while for the right-hand column an even coarser grid with \( m = 21; \Delta \theta = 9^\circ \) was used. Instabilities and/or inaccuracies clearly start to appear, especially on the shock. This shows that the Wendroff matching condition is the most susceptible to this spacing.

Although not shown, it is somewhat surprising that when drifts are added to the model, this does not place stronger constraints on the \( \theta \)-grid.

6.2.6 The Rigidity Grid and Domain

The rigidity grid was taken to be linear in \( \ln P \) between a lower \( (P_{\text{min}}) \) and upper boundary \( (P_{\text{max}}) \):

\[ \ln \left( \frac{P_k}{P_{\text{min}}} \right) = (k-1) \Delta \ln P, \quad k = 1, 2, \ldots, \ell \]  

(6.4)

with \( \Delta \ln P = (\ln P_{\text{max}} - \ln P_{\text{min}})/(\ell - 1) \).

In the Control and the solutions of Chapter 5, \( \ell \) was taken as 88, i.e., \( \Delta \ln P = 0.062 \). This yielded relatively good results in the sense that a finer grid did not alter the solutions appreciably.

The rigidity grid and domain have three important effects on the solutions.

Firstly, the number of grid points affects the power law form of the spectrum below the curvature cutoff. The Control, as well as the no-drift solutions of Chapter 5, have power law spectra on the shock \( \propto T^{-0.8} \), instead of the expected \( T^{-1} \) (\( f \propto p^{-1} \)) form, predicted by (3.29) and (3.30). Figure 6.6(a) shows that halving the number of rigidity steps to \( \ell = 44 \) deteriorates this power law to \( \propto T^{-0.6} \), while Figure 6.6(b) shows that increasing the number of rigidity steps fourfold to \( \ell = 264 \), lets the shock spectrum to converge to the expected \( T^{-1} \) form. This effect places
Figure 6.5: The effects of coarse polar grids on the no-drift source acceleration solutions. Figures (a), (c) and (e) were produced with $m = 41 \ (\Delta \theta = 4.5^\circ)$ and Figures (b), (d) and (f) with $m = 21 \ (\Delta \theta = 9^\circ)$. 
Figure 6.6: Figure (a) shows the effect of a coarse ($\ell = 44$) rigidity grid, while (b) shows a fine ($\ell = 264$) rigidity grid. Figures (c) and (d) show a solution where particles have been injected at 450 keV. This run was done with $\ell = 88$ over a much wider rigidity domain, resulting in a coarser grid than the Control. The different energy scale in Figure (c) is included to demonstrate the behavior of the solution between the injection energy and the minimum energy (where $j_T = 0$).

rather severe limitations on accurate model solutions and data fitting. Our present computer system has only sufficient storage capacity for $\ell \leq 264$.

Secondly, the rigidity domain of the solution, from $P_{\text{min}}$ to $P_{\text{max}}$ is, fortunately, less important. In the Control and the solutions of Chapter 5 $P_{\text{min}} = 0.09$ GV (4.5 MeV for protons), particles are injected on the shock at $P_i = 0.1$ GV (5 MeV), and $P_{\text{max}} = 20$ GV (20 GeV). All the previous solutions show the spectra from the injection energy (5 MeV for protons) upwards. Notice that $P_{\text{min}}$ must necessarily be less than $P_i$, because the specification of a source as well as a lower-boundary spectrum conflicts. The chosen value $P_i = 0.1$ GV is far too high. The pick-up ions in the solar wind should have energies up to four times that of the wind flow energy (Section 3.4), which, at 400 km/s is about 1 keV. Thus, an ideal injection energy is $\sim 1$ keV, or $P_i = 0.0015$ GV. The code cannot handle these low rigidities because (a) it takes too long and requires too much storage, and (b) the diffusion coefficients, which always decrease with decreasing rigidity, become so small that the code cannot handle them. This is because all the terms containing the diffusion coefficients, or any spatial derivative thereof, will effectively
vanish with respect to the convection and energy change terms. The transport equation then effectively becomes
\[ \frac{\partial f}{\partial t} = -V \frac{\partial f}{\partial r} + \frac{1}{3} \left( \frac{2V}{r} - \frac{\partial V}{\partial r} \right) \frac{\partial f}{\partial \ln P} + Q \] (6.5)
which is no longer parabolic. Thus, the numerical code, written for a parabolic equation, breaks down.

The effects of a reduced \( P_{\text{min}} \) and \( P_1 \) are shown in Figure 6.6(c), where \( P_1 \) is decreased to 0.03 GV (450 keV for protons) and \( P_{\text{min}} \) to 0.02 GV (250 keV). This figure also demonstrates the behavior of the solution between \( P_{\text{min}} \) (where \( f = 0 \)) and \( P_1 \). Figure 6.6(d) shows the same result, but on the same scales as all the other spectra. In comparison with Figure 6.1(a) this shows that even if the grid is coarser than the Control case, the high-energy tail is not appreciably different from that of the Control. This insensitivity to the lower momentum boundary is rather obvious. Below the curvature cutoff, the shock spectrum simply continues in the power law fashion, while the modulated spectra are all determined, through adiabatic losses, by the high energy part of the shock spectrum, typically at energies above the curvature cutoff. This result is also highly fortunate because it demonstrates that it is not necessary to run the solution down to realistic injection energies to obtain reliable results.

Thirdly, with the addition of drifts, the code produces instabilities near the upper boundary in energy. In Figures 6.7(a), (b) and (c) we show three drift solutions in the positive drift state with a flat neutral sheet on a greatly expanded intensity scale of 17 orders of magnitude.
In Figure 6.7(a) \( P_{\text{max}} = 20 \, \text{GV} \), while in 6.7(b) \( P_{\text{max}} = 5 \, \text{GV} \). Figure 6.7(c) is identical to Figure 6.7(b), except that the number of rigidity steps is halved to \( \ell = 44 \). These figures show that the instabilities are localized at high energies, i.e., their existence has very little influence on the lower-energy solutions. If the smaller \( P_{\text{max}} \) of Figure 6.7(b) is chosen, they diminish, while the time evolution of Figure 6.7(b) shows that they are also damped in time. In general, being aware of their existence and properties, these instabilities are no great cause for concern.

### 6.2.7 The Time Grid and Domain

The time grid is a simple, linear relation which is generated by incrementing the time with an amount \( \Delta t \) every time step. The model is stopped after about 3 years model time. The solutions in the Control were calculated over 3040 time steps of 5.2 hours each. Figure 6.8 shows the same solutions, calculated with 630 time steps of 42 hours each. The total model time was 3 years, as in the Control.

The spectra in the ecliptic (and at \( 45^\circ \); not shown) are very nearly the same as the Control on the shock and in the middle heliosphere. In the inner heliosphere, however, the spectra are similar only at low energies, but quite different at higher energies, displaying much lower intensities there. This is easily understood by considering the method of characteristics: Figure 6.9 shows that in regions of positive divergence of the solar wind, the characteristic in \((\ln P, t)\)-space has a negative slope, i.e., particles migrate along this characteristic from higher to lower energies. In the acceleration case, no boundary condition is specified on the upper boundary because the spectrum on the shock develops in time from low to high energies. In the outer heliosphere, the divergence of the solar wind is small and, therefore, the characteristic lines have small negative slopes. In the inner heliosphere, the divergence of the solar wind is large and the characteristics have large negative slopes. The distribution function value at the end of the time interval (right side) is the same as the function value at a corresponding point (at a higher energy) at the beginning of the time interval (left side). Near the upper boundary in rigidity, however, a typical distribution function value at a specific point in rigidity on the right side of the time interval can only be traced back along its characteristic to the upper boundary and not to the left side of the interval as required.

Therefore, near the upper boundary in rigidity, the method of characteristics obviously migrates faulty values. The only way to limit this effect is to choose a fine time step. This is, unfortunately, expensive because the runtime of the code is almost exactly proportional to the number of time steps.

Considering the spectra over the polar regions in Figure 6.8(b), one notices huge differences with the Control, which may be attributed to inaccuracies. Another indication of these inaccuracies is clearly visible in the radial profiles on the poles which show a ‘dip’ in the middle heliosphere. This dip is most pronounced in the 100 MeV profile which exactly corresponds with the dip in this energy region in the spectrum at 30 AU over the poles. It has been verified that a further increase in the time step to 83 hours causes this dip to become an instability, producing negative intensities over the poles.

The cause of the trouble in this case seems to be the Crank-Nicolson schemes that have some difficulty in integrating the spatial part of the TPE correctly near boundaries where coefficients and the distribution function change rapidly.

Figure 6.10 shows radial profiles calculated with the same parameters as the Control case. The only difference is that the solution is shown at 1 year model time, instead of the usual 3 years.
Figure 6.8: Inaccuracies arising from a coarse time grid with $n_t = 630$, i.e., $\Delta t = 42$ hours.
model time. The radial profiles show clear negative radial gradients, especially at the lower
\[ \Delta t_2 > \Delta t_1 \]

![Diagram](image)

Figure 6.9: Boundary effects in the method of Characteristics. The two figures on the left represents a small divergence of the solar wind and the two on the right a large divergence of the solar wind.

energies. This is not a numerical effect, but rather a physical effect generated by insufficient convergence. This effect is real in the sense that it is the correct solution for a heliosphere that is not yet in equilibrium. Particles are being accelerated to high energies at the termination shock from a source function injected there. These particles diffuse into the heliosphere where they are being modulated intensely by the solar wind and IMF. In this process they suffer adiabatic energy losses, causing a downward migration of particles in momentum space. This adiabatic cooling process is the most efficient in the inner heliosphere, causing massive energy losses there. Normally, the radial profiles at lower energies are very flat as can be seen from Figure 6.2 in the Control, with small gradients, while those at high energies are steep with large positive gradients. The rise in intensity towards the inner boundary at lower energies is, therefore, due to severely cooled high-energy particles superimposed upon the normally flat radial profiles.

Thus, negative radial gradients may be observed in species accelerated at the termination shock at early times, but at later times this effect is effectively erased by the modulation process: The low-energy particles causing this effect are cooled further and convected into the middle and outer heliosphere, flattening the radial profiles.

This brings us to a potential limitation that depends on the domain of the time grid: If time-asymptotic solutions are desired, care must be taken to ensure sufficient convergence to avoid erroneous results and conclusions. In true time-dependent solutions, however, negative radial gradients can be expected, in agreement with Le Roux (1990).

### 6.2.8 Grid Ratios

Intuitively, one might assume that a simple refinement of a particular grid will improve the quality of the solution. Unfortunately, this is not the case. There is a relationship between the
radial grid and the time grid. This is similar to the well-known stability criteria, imposed on explicit and some implicit methods, all of which place restrictions on the ratio $\Delta t/\Delta x$, where $\Delta x$ is a spatial grid spacing. Although the Crank-Nicolson method is supposed to be unconditionally stable, we find, in agreement with Le Roux (1990:44) that the stability condition

$$ \left| \frac{(V + v_{dr}) \Delta t}{\Delta r} \right| \leq 1. \quad (6.6) $$

must be met. This means that if the radial grid is refined by a certain amount, the time grid must also be refined by an equal amount, or just enough to keep the above condition below unity. The time grid, however, can be refined independently of the radial grid.

No equivalent dependence between the time grid and the $\theta$-grid or rigidity grid was found in our empirical studies. A refinement of these two grids just enhances the quality of the solution.

### 6.2.9 Summary of Mesh and Domain Considerations

During the development of these solutions, experience has shown over and over that the critical parameter in domain and, especially, mesh considerations is the time grid and its domain. A refinement of the time grid can only improve the quality of the solutions. Many problems and instabilities arising during the utilization of the model, such as large gradients or locally large divergences, may be mended by decreasing the time step. Unfortunately, the runtime penalty is frequently quite large, almost unacceptably so. One may choose diffusion coefficients and other parameters to be 'numerically friendly' to enable the usage of a coarser time grid in order to minimize computer time, but such actions will severely limit the applicability of the model to observations.

To avoid erroneous time-convergent solutions, one must also ensure that the total amount of model time is sufficient to facilitate good convergence. In the case of the modulation and re-acceleration of a finite LIS, sufficient convergence is attained after about $2r_{b}/V$ (1.66 years for $r_{b} = 70$ AU). For pure acceleration cases this convergence time rises dramatically to more than 3 years model time. How much more is uncertain, for true convergence is never quite attained.
at the highest energies. One is usually forced to impose a time limit due to the practical reason of computer time.

From the inaccuracies and instabilities arising from a radial grid divided into 71 nodes (51 up to the shock) it is clear that the grid should generally be chosen to be finer than this. Unfortunately, a refinement in the radial grid must also be accompanied by a refinement in the time grid, especially if the model parameters are chosen such that the numerical code is barely stable.

The rigidity grid does not affect the stability of the overall model in the no-drift case. It primarily affects the accuracy of the Wendroff method, used to calculate the spectrum on the shock, in the sense that it does not yield the theoretically expected $T^{-1}$ dependence at low energies. Therefore, if physically correct solutions are needed for data fitting and other quantitatively important applications, one has no choice but to pay the runtime penalty of decreasing $\Delta \ln P$ to at least 0.015. In the case of simple demonstrations, however, $\Delta \ln P = 0.062$ will be sufficient since none of the basic physics is affected dramatically.

In the full drift case, the magnitude of $\Delta \ln P$ does indeed affect the stability. A lower limit for stability lies between $\Delta \ln P = 0.126$ and $\Delta \ln P = 0.062$, provided that the rigidity domain is limited at a relatively low upper rigidity, e.g., $P_{\text{max}} = 5$ GV.

Finally, the $\theta$-grid is relatively unimportant. One may refine it without worrying too much about some relationship between the $\theta$-grid and the other parameters. The number of $\theta$-grid points has little effect on the model's stability and the primary consideration lies in the quality of the solutions. However, a grid finer than $\Delta \theta = 4.5^\circ$ is recommended.

In this section, the limitations of the model were only considered for an acceleration case. Typically, modulation runs, with no termination shock, remain sufficiently stable and accurate with half the rigidity steps ($\ell = 44$) and a time step that is twice as large. In addition, the number of radial steps may be halved. Thus, a pure modulation run can be done at least eight times faster with much less storage capacity used.

### 6.3 Parameter Variations

In Section 6.2 we varied the numerical discretization parameters to demonstrate the stability, accuracy and overall quality of the model. In this section, the physical parameters are varied to gain some insight into the effects of the magnitude of the diffusion coefficients, the position of the shock and the outer boundary, drift and diffusion effects beyond the shock and, finally, the effect of the shock structure itself. This last effect is tested by replacing the discontinuous structure of the shock with a continuous transition from the supersonic to the subsonic state.

Once again, the tests will be done on the acceleration runs only, because most of the effects on a pure modulation model are well-known.

#### 6.3.1 Diffusion Coefficients

In Figure 6.11 the magnitude of both $\kappa_{||}$ and $\kappa_{\perp}$ is doubled relative to the Control. In comparison with the Control, it can be seen that (a) the spectrum on the shock cuts off at a lower energy, (b) that there is less modulation, and (c) that the radial gradients, especially just inside the shock, are smaller.

These effects are well understood. In Section 3.3 it was mentioned that the curvature cutoff for acceleration should occur at the energy where $V_{r_s}/\kappa_{rr} \approx 1$. With the parameters of Figure 6.11
this occurs at 730 MeV, marked with a dashed line. For the Control this value is at 1800 MeV. At sufficiently high energies, the radial gradient is well approximated by the force field expression

\[ g_r = \frac{CV}{\kappa_{rr}} \]

(e.g., Reinecke et al., 1993 and references therein). Since the spectral shape (and, therefore, \( C \)) does not differ that much between the two cases, the gradients in Figure 6.11 are almost half of the values in the Control. These smaller gradients then cause smaller overall modulation. The 10 MeV intensity at 1 AU in Figure 6.11(b), for instance, is 4 times the value of the Control.

### 6.3.2 The Outer-Boundary Radius

In Section 2.2.4 the result was referenced that the stagnation point of the solar wind beyond the shock should at most be at twice the distance of the SWTS. This should also roughly indicate the boundary of the modulation region. In our demonstration runs, with the shock at \( r_s = 70 \) AU, the outer boundary was at \( r_b = 100 \) AU, which is relatively small.

The left column of Figure 6.12 shows results with identical parameters as the Control, except that the extent of the downstream region was halved to 15 AU by placing the outer boundary at \( r_b = 85 \) AU. In the right column of Figure 6.12 the downstream region was doubled to 60 AU by placing the boundary at \( r_b = 130 \) AU.

This factor of 4 increase in the thickness of the downstream region has a moderate effect on the intensities on the shock and in the upstream region. The cutoff on the shock (if defined as that point where the power law ends) doubles from approximately 200 MeV to 400 MeV, while the intensities in the inner heliosphere are about 2.5 times higher. This effect is due to a reduced probability of escape downstream.

---

1 Doubling \( \kappa_{rr} \) does not necessarily halve this energy from 1800 to 900 MeV because of the non-linear relationship (5.17) between \( \beta P \) and \( T \).
Figure 6.12: Variation of the boundary, \( r_b \). In the left column \( r_b = 85 \text{ AU} \) and in the right column \( r_b = 130 \text{ AU} \).
6.3.3 The Position of the Shock and the Total Domain of Modulation

In Figure 6.13 the diffusion coefficients are doubled relative to the Control, as in Section 6.3.1, and together with this all dimensions of the heliosphere are doubled by increasing the boundary from 100 AU to 200 AU, and the shock is moved from 70 AU to 140 AU. The spectra that were drawn at 1, 30 and 70 AU in the Control are now drawn at 2, 60 and 140 AU. Thus, all the dimensions of the heliosphere are doubled.

This solution must be compared to Figure 6.11 because they have the same diffusion coefficients (twice the values of the Control). The spectra and gradients are basically similar. This is not surprising. With the diffusion coefficients of the form (3.65) and (3.66), the effective radial diffusion coefficient, \( \kappa_{rr} \), in the outer heliosphere is \( \propto r \). Thus, both in Figures 6.11 and 6.13, the parameter \( V_{rs}/\kappa_{rr} = 1 \) at 730 MeV, producing similar accelerated spectra and almost similar modulated spectra.

Diffusion coefficients \( \propto r \) in the ecliptic plane are very much demanded by Pioneer/Voyager observations in the outer heliosphere (e.g., Reinecke et al., 1993). Thus, under these conditions, it is quite difficult to determine the scale size of the heliosphere. This result is basically similar to that of pure modulation runs (e.g., Potgieter and Moraal, 1985; Reinecke et al., 1993) that increasing the modulation boundary in proportion to the diffusion coefficients leaves modulation virtually unchanged.

6.3.4 Drift Beyond the SWTS

In this section we investigate the effects of particle drifts in the outer heliosphere beyond the solar wind termination shock. In Section 3.6.2 it was shown that, in a Parker spiral field, these drifts are quite small. In two runs, for positive and negative drift states, all the drift effects beyond the shock were put to zero. These results are not shown since the small changes inside the shock, caused by this effect, do not merit a figure. The main difference between runs with and without drift beyond the shock may be found at the poles. Here, the energy spectra on the shock in the positive drift case are slightly softer than the case with full drift beyond the shock. Similar effects are seen over the poles in the negative drift case where the shock spectrum...
actually falls below the spectrum at 30 AU between 20 and 100 MeV.

However, there are no significant differences in the upstream solutions between the full drift case and this one. Thus, if a Parker spiral is assumed in the downstream region, drift beyond the shock has no real effect on model solutions in the upstream regions. It was pointed out in Section 2.3.7, however, that a Parker spiral downstream of the SWTS cannot be realistic. Thus, further investigations with different downstream field properties are needed.

6.3.5 The Diffusion Coefficient Beyond the Shock

Considering (2.46), it can be seen that if \( \tan \psi^- \gg 1 \), the magnetic field increases with a factor \( s = 4 \) across the shock. This implies that, since the diffusion coefficients are \( \propto B^{-1} \), they will decrease with a factor \( s = 4 \). However, this is only true in the region around the ecliptic since \( \tan \psi^- \propto \sin \theta \). Over the poles \( \tan \psi^- = 0 \) and the magnetic field (and, therefore, the diffusion coefficients) will not change across the shock. Furthermore, the Parker spiral predicts a magnetic field that rises \( \propto r \) beyond the shock if the solar wind falls off \( \propto 1/r^2 \). This results in diffusion coefficients which decrease \( \propto r \).

The validity of this is debatable since it is unlikely that the magnetic field will have the Parker geometry over the poles for reasons discussed in Section 2.3.1. In addition to this, one would expect the diffusion coefficients to increase towards interstellar space. At this point it should be noted, however, that it may be physically correct that the diffusion coefficients will decrease immediately beyond the shock since MIRs and GMIRs, which are regions of low diffusion coefficients, migrate outwards at the solar wind speed. Beyond the shock where the solar wind speed decreases with \( r \), these structures will tend to accumulate to effectively decrease the diffusion coefficients there. Therefore, in absence of any guiding evidence, we investigated what the solutions would be like if an \( s = 4 \) drop was forced for all polar angles and the diffusion coefficients were allowed to rise \( \propto r \) beyond the shock.

The results, shown in Figure 6.14, indicate that relative to the Control these different conditions in the downstream region have relatively small effects in the upstream region. Among these minor effects is the evidence for slightly more modulation found in this case than in the Control, as can be seen from comparing Figures 6.14(b) and 6.2(a).

In the downstream region the solutions are different as expected. These differences, however, are located only at lower energies where the ecliptic radial profiles fall off exponentially (a straight line on the log-lin scales), whereas they stayed more or less constant in the Control. At high energies, the radial profiles are very similar to that of the Control.

At first glance, there seems to be a discontinuity in the distribution function over the poles, which is visible at high energies in the radial profiles. This seems to conflict with the \( f^- = f^+ \) condition imposed on the shock. Closer inspection of these radial profiles, however, revealed that this is not a true discontinuity but a very fast transition [over several grid points of the fine mesh given by (6.2)]. This transition is very small at low energies, but increases significantly towards higher energies. This is probably due to the high diffusive fluxes over the poles.

6.3.6 Variation in the Compression Ratio

From (3.30) it can be seen that if a weak shock with a compression ratio \( s = 2 \) is employed to accelerate particles, the distribution function of these particles should have a momentum distribution function spectrum \( \propto p^{-6} \), or \( j_T \propto T^{-2} \). In Figure 6.15 the results of a source acceleration run with such a weak, \( s = 2 \), shock are shown. Since we are aware of the deterioration of the
$j_T \propto T^{-1}$ spectrum with a course rigidity grid, as was demonstrated in Section 6.2.6, this run in Figure 6.15 was done with $\ell = 264$ to obtain the best possible results.

As expected, the spectrum on the shock is almost perfectly $\propto T^{-2}$, as can be seen from Figure 6.15(a). This very soft spectrum results in a large Compton-Getting factor, $C$. Thus, for sufficiently large energies the force field expression

$$g_r = \frac{CV}{r_{rr}}$$

explains why there are very large radial gradients and, therefore, large modulation of the accelerated spectrum, as is evident from Figures 6.15(a) and (b), showing $\sim 250$ times the modulation of the control case. Due to this huge modulation, it is difficult to see that the typical observed ACR spectra can be produced by a shock with a compression ratio considerably less than 4.

### 6.3.7 The Structure of the SWTS

Up to this point, the SWTS has been modelled as a true discontinuity, matching the intensities (3.18) and streaming (3.44) on both sides of the shock with the technique developed in
Section 4.8.

Such a shock transition will, however, not always be a true discontinuity. Due to the effects of cosmic-ray pressure (of the particles accelerated on the shock) and the creation of waves due to streaming across the shock, the change in plasma flow may become extended in space. Jones and Ellison (1991) reviewed this topic of shock smoothing. In general, part of the transition may be continuous, leading to a so-called shock precursor, together with a discontinuous part of the transition, called a subshock.

It may well be that the SWTS is modified in this sense. We, therefore, investigate whether such a modified shock structure has an influence on the accelerated spectra. In principle, such a model should have no significant effect on the efficiency of the acceleration because the acceleration is determined by the total amount of compression and not by its spatial distribution.

To test this hypothesis, a solar wind 'shock' transition from a constant upstream value to a $1/r^2$ downstream value was chosen, with

$$V(r) = \begin{cases} 
1 & \text{if } r < r_s - \delta_s \\
1/a \cos(br + c) + d & \text{if } r_s - \delta_s \leq r \leq r_s + \delta_s \\
\frac{a}{s} \left(\frac{r_s + \delta_s}{r}\right)^2 & \text{if } r > r_s + \delta_s 
\end{cases} \quad (6.7)$$

and its derivative with

$$\frac{\partial V(r)}{\partial r} = \begin{cases} 
0 & \text{if } r < r_s - \delta_s \\
-ab \sin(br + c) & \text{if } r_s - \delta_s \leq r \leq r_s + \delta_s \\
\frac{2a}{r_s} \left(\frac{r_s + \delta_s}{r}\right)^2 & \text{if } r > r_s + \delta_s 
\end{cases} \quad (6.8)$$

The compression ratio, $s$, was chosen as 4, while the parameter $\delta_s$ was chosen such that the upstream to downstream transition occurred over a distance of 2 AU. With this, the parameters $a$, $b$, $c$ and $d$ were calculated such that this transition was continuous in both $V$ and $\partial V/\partial r$. For different choices of $s$ and $\delta_s$, these four parameters have to be recalculated in each case. To facilitate the easy changing of $s$ and $\delta_s$, a routine using the bisect algorithm was built into the code to solve the resulting transcendental equation in calculating the parameters $a$, $b$, $c$ and $d$. Because of the fineness of the transformed radial grid (6.2) across the shock, this 2 AU transition occurred over 42 grid points.
Particles were injected isotropically at \( r = r_s \), with a Gaussian in rigidity and a spiked profile in radial distance:

\[
Q(r, P) = Q_0 \frac{1}{\pi} \frac{\alpha_r}{\alpha_r^2 + (r - r_s)^2} e^{-\frac{1}{2} \left( \frac{P - P_i}{\alpha_P} \right)^2}.
\]  

(6.9)

Since this model does not contain a discontinuous shock, the Wendroff matching condition does not play a role in the solution. Instead, the matching conditions (3.18) and (3.44) across the 'shock' simply reduce to \( f^- = f^+ \) and \( (\partial f/\partial r)^- = (\partial f/\partial r)^+ \). Thus, the acceleration is effected in the region of adiabatic compression in the region where \( \nabla \cdot \mathbf{V} < 0 \), exactly analogous to the adiabatic cooling in the remainder of the heliosphere where \( \nabla \cdot \mathbf{V} > 0 \).

The parameters chosen were \( P_i = 0.1 \) GV, \( \alpha_P = 0.005 \) GV, \( r_s = 70 \) AU and \( \alpha_r = 0.1 \) AU. Both the parameters \( \alpha_r \) and \( \alpha_P \) may be interpreted as the width at half maximum of the spiked function and Gaussian, respectively. The value of \( Q_0 \) was adjusted to give the same intensity as the discontinuous runs at an energy of \( T = 100 \) MeV. All other parameters were kept identical with the Control. To enhance the stability of the continuous transition, the time step was taken as half that of the Control and, therefore, the Control was rerun with the finer time step to obtain a reference with which the continuous transition runs can be compared. The results of the discontinuous run are shown in the left column of Figure 6.16 and those of the continuous transition are shown in the right column.

Comparing the results of the continuous transition in the right column of Figure 6.16 with those of the discontinuous transition in the left column, shows that the resulting spectra are (a) quite similar as expected, but (b) that there are instabilities at both the low- and high-energy ends. The instabilities at the high-energy side are easily understood in terms of diffusion mean free paths that rise with energy. At low energies the particles 'see' the continuous transition as such, but at higher energies the thickness of the transition region becomes smaller than the diffusion mean free path. This causes the solution to become unstable. This high-energy instability can be removed if the spatial extent of the transition is increased dramatically. This, however, is undesirable since this affects the modulated spectra in the region around the shock.

This unstable nature of the continuous transition model was found for a wide range of parameters and at low energies it is due to a combination of the source function, spread in energy, and the small diffusion coefficients; in all the runs, both \( \kappa_{||} \) and \( \kappa_{\perp} \) are \( \propto \beta P \). Thus, at low energies, this destroys the parabolic nature of the TPE as was discussed with equation (6.5). The comparison between a continuous and discontinuous transition was, therefore, repeated, but this time with diffusion coefficients such that at \( P < 0.66 \) GV \( (T = 210 \) MeV for protons), \( \kappa_{||} \) and \( \kappa_{\perp} \) are \( \propto \beta \). Above this 'kink' point they have the same \( \beta P \) dependence as before. The results, shown in Figure 6.17, show that (a) the instabilities at low energies are less, (b) the shock spectrum is affected very little by this change, and (c) because the diffusion coefficients are much higher at low energies, the modulated intensities are much higher than those in Figure 6.16.

Thus, we conclude that (a) the discontinuous transition model is numerically much superior to the continuous one, and (b) that for the same total compression, both methods yield approximately the same results. If it is found, therefore, that the SWTS is strongly smoothed in its structure; this complication can be handled well by our model.
Figure 6.16: Energy spectra, radial and latitudinal gradients obtained for a smooth energy dependence of the diffusion coefficients for discontinuous (left) and continuous (right) transitions.
Figure 6.17: Energy spectra, radial and latitudinal gradients obtained for a 'kinked' energy dependence of the diffusion coefficients for discontinuous (left) and continuous (right) transitions. Below $T = 210$ MeV the diffusion coefficients are $\propto \beta$ and above they are $\propto \beta \rho$. 

163
Chapter 7

Applications of the Solution

7.1 Introduction

The purpose of this thesis was to develop a numerical model with which the theory of anomalous cosmic-ray acceleration at the solar wind termination shock can be tested and explored. The numerical model, the solutions it produces and its properties and limitations were described in Chapters 4, 5 and 6, respectively.

There are many potential applications of this model, of which only a few will be explored cursorily in this chapter. The first application is the acceleration and modulation of anomalous cosmic rays. The study of these anomalous cosmic rays are quite important since the steeper spectra produced by the acceleration process result in a large Compton-Getting factor and, therefore, in typically large radial gradients since \( g_r = CV/\kappa_{rr} \), according to the simple force field solution. Secondly, as was seen in Section 5.5.1, the solar wind termination shock does not affect galactic cosmic rays very strongly, but since anomalous cosmic rays are probably accelerated there, it represents a much better tool with which the conditions at the SWTS and the outer heliosphere can be studied.

A second possible application of this model is to study temporal effects in the modulation. As early as 1974 hysteresis effects, that develop with time between the intensities at two different energies, were observed in neutron monitor data. These hysteresis effects were attributed to energy-dependent time lags (O’Gallagher, 1975; Moraal, 1976; Popielawska, 1992). Lopate and Simpson (1991) have investigated these observed time lag effects and obtained theoretical time lags. We are, however, not aware of hysteresis and time lag studies with an acceleration model.

Therefore, in the first part of this chapter we shall attempt to simulate the acceleration and modulation of anomalous cosmic rays with specific reference to singly-ionized Helium (He\(^+\)) and Oxygen (O\(^+\)), as well as the problem of anomalous Hydrogen. The second part of this chapter contains a preliminary study of energy-dependent time lag effects and the resultant hysteresis effects. This will be followed by an introductory study of time dependent acceleration effects.

7.2 Acceleration of the Anomalous Component

Following the lead of Fisk et al. (1974), Pesses et al. (1981) and Jokipii (1986), we assume that singly-ionized ions are accelerated at the SWTS and modulated by the usual processes to form the anomalous component of cosmic rays. Therefore, our newly-developed numerical model can be applied to the problem of the acceleration and modulation of anomalous cosmic rays (ACR). This model is used to produce solutions that approximate some of the experimental data gathered by IMP8 and the Pioneer and Voyager spacecraft.
7.2.1 Acceleration and Modulation of Anomalous Helium and Oxygen in 1977 and 1987

In this section we simulate the observations of He$^+$ and O$^+$, recorded during the 1977/78 and 1987 solar minima. These observations, shown in Figure 7.1, were obtained with experiments on the IMPS satellite at 1 AU, and experiments on the Pioneer 10 and 11, and Voyager 1 and 2 spacecraft at radial distances of 1, 2, 6, 15, 24, and 42 AU in the ecliptic plane. Figure 7.1(a) shows the spectra of O$^+$ observed during the 1977/78 solar minimum. From the bottom up they were measured at 2, 6 and 15 AU. Figure 7.1(b) shows the O$^+$ spectra measured at the 1987 solar minimum and from the bottom up they are the spectra at 1, 24 and 42 AU. Figure 7.1(c) shows the 1977/78 He$^+$ spectra at 2, 6 and 15 AU and Figure 7.1(d) shows the 1987 He$^+$ spectra at 24 and 42 AU. These data were taken from Reinecke et al. (1993).

A preliminary data fit was done by Steenkamp and Moraal (1993a). In that paper a no-drift version of the continuous transition model, described in Section 6.3.7, was used. Here the data fits are repeated with the discontinuous transition model, also excluding drifts. The reason for this is the same as was described in Moraal (1993): When drifts are added to model the observed positive and negative latitudinal gradients, the quality of the radial gradients in the ecliptic is destroyed. This phenomenon makes it exceedingly difficult to find data fits with drift models. The reason for this probably lies in our limited understanding of the large-scale field configurations in the heliosphere.

The continuous spectra shown in Figure 7.1 were calculated with a no-drift version of the numerical solution. The modulation region has a spherical shock at $r_s = 120$ AU. The radial solar wind, $V$, has a uniform speed of 400 km/s inside the shock, while outside the shock it falls off divergence free, $\propto 1/r^2$, until an outer boundary at $r_b = 200$ AU, where we put $f = 0$. Particles were injected isotropically along the shock face (at $r = r_s$) at an injection rigidity of $P_i = 0.2$ GV (1.25 MeV/nuc) for He$^+$ and $P_i = 0.4$ GV (0.3 MeV/nuc) for O$^+$.

The diffusion coefficients used were of the form given by (3.65) and (3.66), with $P_0 = 1$ GV, while a standard Parker spiral field (2.29) was used. The parameters $a_{\parallel}$ and $a_{\perp}$ are included to allow some variation of the spatial dependence about the nominal $B_{\parallel}$ form. The parameters $b_{\parallel}$ and $b_{\perp}$ allow variations on the $B P^\parallel$ rigidity dependence. The fits in Figure 7.1 were obtained by using

$$
\begin{align*}
(k_{\parallel})_0 &= 1.2 \times 10^{22} \text{ cm}^2/\text{s}, & a_{\parallel} &= 0.75 & b_{\parallel} &= 1.0 \\
(k_{\perp})_0 &= 1.5 \times 10^{21} \text{ cm}^2/\text{s}, & a_{\perp} &= 0.97 & b_{\perp} &= 1.0 
\end{align*} \quad (1977/78)

(7.1)

and

$$
\begin{align*}
(k_{\parallel})_0 &= 3.1 \times 10^{22} \text{ cm}^2/\text{s}, & a_{\parallel} &= 0.625 & b_{\parallel} &= 1.0 \\
(k_{\perp})_0 &= 9.0 \times 10^{20} \text{ cm}^2/\text{s}, & a_{\perp} &= 1.0 & b_{\perp} &= 1.0 
\end{align*} \quad (1987)

(7.2)

The numerical code was started with an empty heliosphere, and the source was switched on at $t = 0$. The solution was updated in time steps of 5.2 hours, until sufficient saturation was reached after 3 years.

Generally, this model provides a reasonable fit to the observations. In greater detail, the 1977/78 O$^+$ spectra in Figure 7.1(a) are harder than those predicted by the model. The radial gradients are also not correct, with the model predictions at 6 AU slightly lower than the data, and slightly higher at 2 AU. The 1987 O$^+$ spectra in Figure 7.1(b) are also harder than the model predictions. The radial gradients in the outer heliosphere (24 and 42 AU) are quite acceptable, but in the inner heliosphere they are too small with the calculated 1 AU intensities higher than the observed ones.

The He$^+$ fits are generally of a better quality than the O$^+$ fits with acceptable radial gradients during both solar minima. The observed spectra are generally slightly softer at higher energies.
Figure 7.1: Observed and calculated spectra of He$^+$ and O$^+$. The dotted lines in each frame represent the spectrum on the shock at $r_s = 120$ AU.
than the model results, but the overall spectral shape is in quite good agreement with the observations.

Therefore, our present no-drift acceleration model can reproduce the He$^+$ and O$^+$ spectra observed during the solar minima of 1977/78 and 1987 with reasonable accuracy. This is yet another confirmation that the present theory on the origin of anomalous cosmic rays is consistent with the observed spectra of these species. Detailed discrepancies between predictions and observations remain. As mentioned previously, an important next step in this modelling of the acceleration and modulation of anomalous cosmic rays is the addition of guiding center drift and drift acceleration.

7.2.2 Anomalous Hydrogen

Steenberg (1995) made the first extensive use of the present model by starting a detailed investigation on the detectability of anomalous Hydrogen in the heliosphere. Christian et al. (1988) and Christian (1989) deduced that they detected anomalous Hydrogen at 24 AU. However, the methods that they used were not based on actual solutions of the cosmic ray TPE and, therefore, Reinecke and Moraal (1992) showed that the 'anomalous' bulges in the Hydrogen spectra can also be explained by pure modulation effects. Lopate and McKibben (1991) also studied this problem.

To date, neither Christian et al.'s nor Reinecke and Moraal's argument can yield conclusive answers on the existence and abundance of anomalous Hydrogen. Therefore, our newly developed numerical model was used by Steenberg (1995) to investigate the matter further. The problem with anomalous Hydrogen is that it is impossible to distinguish it from galactic Hydrogen, since both are protons. Thus, the problem must be handled in a fundamentally different way as was done for anomalous He$^+$ and O$^+$ in the previous section: a galactic boundary spectrum, as well as a source of protons on the shock must be specified simultaneously.

The existence of anomalous Hydrogen in the inner heliosphere is purely dependent upon the strength of the source of protons arriving at the SWTS. If a large source exists there, the anomalous component will dominate the galactic component and will be easily recognized throughout the heliosphere. On the other hand, if the strength of the source is very small, the anomalous component will be overshadowed by the galactic component and will be undetectable throughout large parts of the heliosphere.

The answer to this problem lies in the detailed form of the observed spectra. In attempting to fit the hydrogen spectrum, one must choose the strength of the source function and modulation parameters such that one can obtain reasonable fits for the available data. After such a fit is obtained, one can re-run the solution, but with the source on the shock switched off. As was shown in Section 5.5.4, the difference between the two solutions gives a good idea of the contribution of the anomalous component to the intensity.

This process was done in great detail by Steenberg (1995) and he obtained the fits shown in Figure 7.2 with the following parameters: The boundary of the heliosphere was chosen to be at $r_b = 120$ AU and the shock was placed at $r_s = 90$ AU with a compression ratio of $s = 4$. The rigidity domain was taken to be from $P_{\text{min}} = 0.09$ GV (4.5 MeV) to $P_{\text{max}} = 20$ GV. Particles were injected isotropically along the shock face at an injection rigidity $P_i = 0.1$ GV (5 MeV). The source function had a strength of $10^5$ particles/GV/(AU)$^2$/S. The solution was updated every 2.8 hours and was allowed to converge towards saturation until a model time of 1.3 years had passed to obtain a nearly time-asymptotic solution.
Figure 7.2: Observed and calculated spectra of protons, from Steenberg (1995). Filled data points are from Reinecke et al. (1993); open data points are from Moraal (1990).
The diffusion coefficients had the general form of (3.65) and (3.66) with

\[
\begin{align*}
(\kappa_\parallel)_0 &= 1.32 \times 10^{23} \text{ cm}^2/\text{s}, & a_\parallel &= 1.0 & b_\parallel &= 2.0 \\
(\kappa_\perp)_0 &= 4.05 \times 10^{19} \text{ cm}^2/\text{s}, & a_\perp &= 1.5 & b_\perp &= 0.5,
\end{align*}
\]  

(7.3)

with \(\kappa_\parallel\) and \(\kappa_\perp\), and the effective radial \(\kappa_r\) shown in the bottom right panel of Figure 7.2.

The boundary spectrum and the spectrum on the shock, in the other panels at 1, 24 and 42 AU, are self-evident. The calculated modulated spectra are best explained with the 42 AU panel. The full line is the data fit, produced by the combined acceleration/re-acceleration model, in which a source of low-energy particles is accelerated on the shock, and the pre-existing boundary spectrum is re-accelerated there. The dash-dot spectrum is the accelerated (or anomalous) spectrum only, i.e. with the boundary spectrum switched off. Conversely, the dashed spectrum is the re-accelerated boundary spectrum, with the source on the shock switched off.

Each of the panels also shows a \(j_T \propto T\) line, which is the expected form of modulated spectra in the adiabatic limit of modulation.

From these data fits, Steenberg (1995) made the following conclusions: The 42 AU spectrum clearly contains an anomalous contribution because the combined spectrum lies significantly above the re-accelerated spectrum. At 24 AU, and especially at 1 AU, this contribution is, however, negligible. In the outer heliosphere the spectra clearly bulge above the \(j_T \propto T\) line, but this is not necessarily due to an anomalous component because the re-acceleration spectrum also shows this bulge. Reinecke and Moraal (1992) showed that this bulge is even present in a pure modulation model.

These conclusions can, however, for several reasons not be unique:

- As mentioned previously, drift effects are excluded.
- The time-asymptotic solution may not be valid. In a continuation of his study, Steenberg found, in fact, that temporal effects in the model change the diffusion coefficients needed for a proper fit quite drastically. However, these effects do not drastically affect the relative contributions of the anomalous component.
- The diffusion coefficients used here differ quite drastically from the ones in Section 7.2.1 used for He\(^+\) and O\(^+\). This may not be too surprising, because Steenberg chose different dimensions for the heliosphere than were used in the He\(^+\) and O\(^+\) fits. It is somewhat surprising, however, that Steenberg finds it difficult to fit H\(^+\) spectra with identical parameters than He\(^+\) and O\(^+\) spectra.

These difficulties are encouraging in the sense that they demonstrate that anomalous species are quite sensitive to modulation parameters, and that they may, generally, provide a better means to study the details of modulation and acceleration than pure galactic species. For this reason, detailed studies of anomalous cosmic rays, especially in the outer heliosphere, should be vigorously continued with this model.

### 7.3 Time-Dependent Effects

All the solutions up to this point were time-asymptotic solutions, where all the parameters were taken to be independent of time and the model time was allowed to advance until the solution does not change appreciably with time. However, with time-dependent parameters, the numerical solution can also be used to study time-dependent effects in the modulation (as has, of
course, been done for several existing solutions mentioned in the introduction). In general, such
time-dependent solutions can be quite difficult to interpret in detail. Therefore, in the final two
sections we demonstrate the nature of time-dependent solutions by varying diffusion coefficients
in a simple harmonic fashion over the solar cycle. These solutions should be the starting point
for detailed studies of time lag effects in the modulation. One effect that is usually attributed
to such time lags is the existence of hysteresis loops between high and low energy cosmic rays
over the 11-year modulation cycle.

Potgieter and Le Roux (1992a, 1992b) and Le Roux and Potgieter (1992a, 1992b) did intensive
studies of the cosmic-ray response to variations in the neutral sheet tilt angle by using drift
models of the modulation. Here, the drift effects will be ignored to point out that, especially
for anomalous cosmic rays, there are significant phase lags due to simple convection-diffusion
effects.

7.3.1 Hysteresis Effects

We first investigate time lag and hysteresis effects in a pure modulation model, which was
published by Steenkamp and Moraal (1993b). Numerous examples of such hysteresis effects
have been observed, and these effects are often regarded as being due to the temporal, $\partial f/\partial t$,
term in the transport equation: cosmic rays need time to respond to temporal changes of the
transport parameters. O'Gallagher (1975), Hatton and Bowe (1981), Forman et al. (1986), Perko
(1987), Lopate and Simpson (1991), and Popielawska (1992) have, amongst others, studied these
time-delay phenomena in the modulation.

O'Gallhager (1975) observed phase lags of 100 to 200 days of low-energy particles relative to
neutron monitor intensities. Significant hysteresis loops were observed in the data comparing
high vs. low cutoff rigidity neutron monitors. The early observations were summarized in Moraal
(1976). It was also observed that the peak intensities of cosmic rays do not occur at the same
time as solar minimum. Lopate and Simpson (1991) have observed time lags of approximately
90 days between 13 GV (Huancayo neutron monitor) and 0.4 GV (IMP8 satellite) intensities,
and time lags of 20 to 30 days between the 13 GV (Huancayo) and 5 GV (Climax neutron
monitor) intensities.

Many theoretical studies of these time lags have been made. Among these were Hatton and Bowe
(1981), who designed a delayed-response function model to simulate an 11-year modulation cycle.
Forman et al. (1986) studied time lags via the force field approach, from which they showed
that the phase speed of cosmic-ray variation may travel twice as fast as the modulation agent.
Perko and Fisk (1983) and Perko (1987) modelled an 11-year solar cycle as a series of short-
term pulses in the solution of a one-dimensional cosmic-ray TPE. Lopate and Simpson (1991)
used a two-dimensional model based on a statistical test-particle approach to study these time
lags with and without drifts. In these calculations they studied the effect of a travelling wavy
neutral sheet on these time lags. Lopate and Simpson (1991) calculated actual time lags due
to the propagation of particles and not phase lags in the modulation. Potgieter and Le Roux
published four papers (Potgieter and Le Roux, 1992a, b; Le Roux and Potgieter, 1992a, b)
where they used a two-dimensional time-dependent TPE including drifts, in which the influence
of such a travelling wavy neutral sheet on the temporal intensity profiles were studied.

From these studies it is, however, not yet clear precisely what the common cause of these lags
and hysteresis patterns is. For instance, for realistic gradients and diffusion coefficients it can
be readily shown that each of the terms on the right-hand side of the TPE (3.12) has a time
scale of the order of months, whereas the temporal term on the left hand has an observed time
scale of $\approx 5$ years (half a solar cycle). Because this time scale is at least 20 times as long as
those of the individual terms, it can be argued that the cosmic-ray heliosphere is always in near equilibrium, and temporal effects may therefore be ignored. Therefore, on simple order of magnitude arguments it is tempting to say that temporal effects are not important, and that the modulation cycle is just a sequence of quasi-steady states. The accuracy of this statement will be discussed shortly.

Another way of stating this is in terms of the wavelength of an 11-year harmonic variation propagating at 400 km/s, which gives \( \lambda \approx 900 \) AU. A heliosphere of the order of 100 AU thus contains only a fraction of this wave at any given time, and it should therefore oscillate with almost the same phase everywhere (like an LCR circuit as opposed to the phase effects which are important in propagation along a transmission line).

With these order-of-magnitude arguments in mind, Steenkamp and Moraal (1993b) solved the time-dependent TPE for protons in a model heliosphere with a passive outer boundary (no termination shock) at 120 AU, using the diffusion parameters of Section 7.2.1. As in Section 7.2.1, drifts are excluded to avoid complicating the underlying physics too much. As modulation agent, \( \kappa_{||}, \kappa_{\perp} \) and, therefore, \( \kappa_{rr} = \kappa_{||} \cos^2 \psi + \kappa_{\perp} \sin^2 \psi \), are varied by multiplying them with the harmonic function

\[
\exp[i(\omega t - kr + \delta)],
\]

which has a period \( T = 2\pi/\omega = 11 \) years, propagating with a speed \( V = 400 \) km/s, implying a wavelength \( \lambda = 2\pi/k = 926 \) AU. Figure 7.3 shows a series of snapshots of the effective \( \kappa_{rr} \) in the ecliptic plane (full lines), spaced 1/4 of a full period apart. The dashed lines show the harmonic function (7.4) with which the background diffusion coefficients \( (\propto 1/B) \) were multiplied. The most important point to note is that the magnitude and spatial dependence of \( \kappa_{rr} \) are drastically different at \( t = T/4 \) and \( 3T/4 \) i.e., in the ascending and descending phases of modulation.

Figure 7.3: The effective radial \( \kappa_{rr} \) and the cyclic function as function of radial distance and time. The radial diffusion coefficient were normalized such that its magnitude is equal to the value of the cyclical variation at the boundary, which varied between 1 and 3 with an amplitude of one.
The initial runs are for a spherically symmetric heliosphere, advancing in time steps of 1.56 days, to produce the well-known hysteresis loop in the intensity at 60 AU, shown by the solid line of Figure 7.4(a). This loop is of the characteristic form observed in cosmic-ray counting rates. The diffusion coefficient used for this run is the ecliptic value of a $\kappa_{rr}$, calculated from the average of the two sets of parameters used in Section 7.2.1, multiplied with (7.4).

It is not immediately clear how such a loop is established. To investigate this, one keeps in mind that a series of steady-state solutions corresponds to setting $\partial f / \partial t = 0$ and $k = 0$ in (7.4), and these solutions will obviously trace a single line up and down through the middle of this loop. Thus, one might say that the width of the loop is a measure of the importance of time-dependent effects. This is, however, an oversimplification. To determine the effect of the $\partial f / \partial t$ term, one must first ask what will happen in a series of steady-state solutions which have the correct spatial dependence of $\kappa_{rr}$ at each time, shown in Figure 7.3, built into them. We call these quasi-steady (QS) states, and 33 of these states were calculated over the 11-year cycle. These 33 states are shown by the connected beads in Figure 7.4(a). The QS loop is approximately of the same width as the time-dependent (TD) loop but, somewhat surprising, it goes the other way round. Figure 7.4(b) shows the same results for 1 GeV vs. 5 GeV intensities, producing a much thinner loop.

To understand the origin of the loops of Figures 7.4(a) and 7.4(b), we plot, in Figure 7.4(c), the 11-year time history of values of $\kappa_{rr}$ (arbitrarily normalized), together with TD and QS intensities, in full and dashed lines, respectively. The dot on each curve denotes its minimum value. Notice that at 60 AU $\kappa_{rr}$ reaches its minimum 260 days (8.5 months) after the ‘solar maximum time’ at 5.5 years, because of the propagation delay at 400 km/s.

The 5 GeV and 1 GeV intensities reach their minimum $\approx$ 4 months after the local minimum in $\kappa_{rr}$. This happens because, although $\kappa_{rr}$ has reached its minimum at 60 AU, the particles respond to conditions beyond 60 AU, where, according to Figure 7.3, the strongest modulation conditions have not yet been reached. This phenomenon was discussed by Forman et al. (1986). The time lag between the TD and QS solutions at 1 GeV is small ($\approx$ one month), and at 5 GeV it is almost entirely absent. This means that in the neutron monitor range (> 1 GeV), the modulation is almost a succession of quasi-steady states to which the particles can easily respond, and time-dependent effects are not important.

The 10 MeV situation is dramatically different. The QS minimum is reached $\approx$ 2 months ahead of that in $\kappa_{rr}$, while the TD minimum lags the QS one with almost a year. The QS minimum leads the local minimum in $\kappa_{rr}$ because low-energy particles do not directly propagate in from outside: They are severely cooled products of high-energy particles, and were cooled most effectively at radial distances < 60 AU, where the minimum in $\kappa_{rr}$ occurred before its minimum at 60 AU. The TD intensity at 10 MeV lags the QS intensity by almost a year because these low-energy particles do not have sufficient time to adjust to the changes in modulation agent. Thus, TD effects at low energies are very important; in Figure 7.4(a) they are so important as to flip the QS hysteresis loop around in the other direction.

The heliosphere is certainly not spherically symmetric, and Figure 7.4(d) shows the equivalent of Figure 7.4(a) in a two-dimensional heliosphere, using the same diffusion parameters. The QS loop is now wider, while the TD loop is narrower and gets a double structure. It is difficult to understand or describe the origin of this complicated loop, apart from noting that the difference between Figures 7.4(a) and 7.4(c) is only latitudinal diffusion (through $\kappa_L$). What is clear, however, is that before one can interpret such a loop, one has to make prior reference to the QS
Figure 7.4: Hysteresis loops and temporal variation in intensity for an 11-year variation.
state, as well as the spherically symmetric results of Figures 7.4(a). Otherwise it is easy to draw the wrong conclusion that the narrow TD loop of Figure 7.4(a) signifies that temporal effects are unimportant.

Figure 7.4(e) is the counterpart of Figure 7.4(a), but this time calculated at 10 AU. The narrower QS loop signifies that the 1 GeV and 10 MeV particles respond to modulation parameters in regions of space which are, effectively, much nearer to one another (both in the inner heliosphere). From the 10 AU equivalent of Figure 7.4(c) it follows that the propagation speed of the 1 GeV minimum intensity, from 10 to 60 AU, is 450 km/s. This phase speed exceeds \( V \) because, according to Forman et al. (1986), the particles at 60 AU respond to average modulation conditions between 60 and 120 AU, while at 10 AU they respond to changes between 10 and 120 AU. These two "average modulation distances" are weighted according to the radial dependence of \( \kappa_{rr} \), but their difference is always less than the 50 AU separation between the two points. The 10 MeV TD minimum propagates at only 360 km/s. Here one can hardly think of an equivalent explanation because this effective speed is a combination of two effects: the average modulation distance combined with large time-dependent effects.

This (grossly oversimplified) solution clearly shows that time-delay and/or hysteresis effects in the modulation can, in principle, be understood, but are extremely complicated to relate to observed time lags. The best example is that the almost non-existent TD loop in Figure 7.4(d) does definitely not imply that time lag effects are not important.

In all likelihood the modulation is not the result of a smooth 11-year cycle of the modulation agent, but due to a build-up of short-term pulses of enhanced magnetic turbulence. The frequency of repetition of such pulses then determines the 11-year envelope. This approach was followed by authors like Perko and Fisk (1983), Le Roux and Potgieter (1993) and Potgieter and Le Roux (1994) to model the 11-year modulation cycle. As a matter of interest, the propagation of a single pulse was studied to see how the lag effects differ from those in a smooth cycle. Figure 7.5(a) shows the effect of such a pulse in \( \kappa_{rr} \) with a duration of 1.5 years, released on the sun at \( t = 0 \), reaching 60 AU 3/4 of a year later. The intensity minima occur at the same relative times to the minimum in \( \kappa_{rr} \), than for the full wave in Figure 7.4(c), and the small difference between the 1 GeV QS and TD pulses confirms that time-dependent effects above 1 GeV remain small. The huge asymmetric wake left by the 10 MeV intensity, however, produces an entirely
different hysteresis loop (shown in Figure 7.5(b)) than the 11-year wave of Figure 7.4(a).

Steenberg (1995) continued this investigation by adding acceleration to this problem to study the temporal behavior of the anomalous component.

### 7.3.2 Temporal Effects in Acceleration Models

In this section time-dependent runs, done by Steenberg (1995) over a time span of 33 years model time, are shown to show that the temporal profiles for accelerating models are dramatically different. As in the previous section, the diffusion coefficients were multiplied with a waveform given by (7.4) to obtain a harmonic variation of the diffusion coefficients. Since Steenberg (1995) did this to specifically address the problem of the anomalous proton component, all the parameters of Section 7.2.2 were used, instead of those used in the previous section. As in Section 7.3.1 the particle species is protons \((A/Z = 1)\). Appropriate energy scalings for other species can be done with the aid of the results of Section 5.7.1. The solution was updated in steps of 18 hours model time, until a time span of 33 years model time had elapsed. The first run was done for a pure modulation case with the modulation boundary at \(R_b = 90\) AU where the boundary spectrum was assumed to be of the form (5.12). This run, with the passive modulation boundary, is essentially a repetition of the run in Figure 7.4(c), but with different diffusion parameters. The 10, 100 and 1000 MeV temporal profiles at 1, 23, 42 and 90 AU are shown in Figure 7.6. From Figure 7.4(c) it can be seen that the minima produced by the one-dimensional TD model occur, on average, 6 months after the minimum in the diffusion coefficient. With the large differences in diffusion parameters in mind, it is interesting to note that in this (two-dimensional) case, on average, the maxima (minima) also occur about 6 months after the maxima (minima) in the diffusion coefficients for all radial distances.

Figure 7.7 shows exactly the same calculations for a pure shock acceleration case, i.e., for an anomalous cosmic-ray species. That is, instead of a modulation boundary at 90 AU, there exists a shock which accelerates particles with the intensity zero on the outer boundary at 120 AU.

The lags between the maxima in the intensities and the maxima in the diffusion coefficients are now dramatically larger, up to 3.5 years for the 1000 MeV intensity at 1 AU. In addition, it is clear that the 1000 MeV intensity maxima occur at approximately the same time at all radial distances (although, of course, they occur later relative to the maxima in the diffusion coefficients at 1 AU than in the outer heliosphere).

At low energies, e.g., in the 10 MeV case, the intensity maxima occur later in the inner heliosphere than in the outer heliosphere. This looks like an inward-propagating intensity profile. This effect is quite remarkable because it demonstrates that, in principle, it will be hard to relate outward-propagating disturbances with features in the intensity for anomalous species, as is often done for galactic cosmic rays.

The explanation for these inward-propagating (and long) lags is that the maximum in \(\kappa_T\), must first propagate outwards towards the shock before it can influence the acceleration process. The resultant accelerated particles then diffuse inwards to be further modulated by the outward propagating wave in \(\kappa_T\). In other words, the large values of \(\kappa_T\) at the maxima result in less effective acceleration at the shock, but the same maxima also cause less modulation. Conversely, small values of \(\kappa_T\) result in more efficient acceleration and stronger modulation. These two mechanisms, working against each other, produce the huge lags seen in Figure 7.7.

Figure 7.8 shows the results of similar calculations with the re-acceleration of the boundary spectrum, used in Figure 7.6, included. No source of low-energy particles was present on the shock. From this it is evident that the effects seen in the case of Figure 7.7 are also present.
Figure 7.6: The 10, 100 and 1000 MeV modulated intensities at 1, 23, 42 and 90 AU for an 11-year variation in the diffusion coefficients for a time-dependent pure proton modulation model. The vertical lines indicate the maxima of the cyclical function and the dots that of the modulated intensities.
Figure 7.7: The 10, 100 and 1000 MeV modulated intensities at 1, 23, 42 and 90 AU for an 11-year variation in the diffusion coefficients for a time-dependent source acceleration model for protons. The vertical lines indicate the maxima of the cyclical function and the dots that of the modulated (accelerated) intensities.
Figure 7.8: The 10, 100 and 1000 MeV modulated intensities at 1, 23, 42 and 90 AU for an 11-year variation in the diffusion coefficients for a time-dependent re-acceleration model for protons. The vertical lines indicate the maxima of the cyclical function and the dots that of the modulated intensities.
Figure 7.9: The 10, 100 and 1000 MeV modulated intensities at 1, 23, 42 and 90 AU for an 11-year variation in the diffusion coefficients for a time-dependent source acceleration/re-acceleration model for protons. The vertical lines indicate the maxima of the cyclical function and the dots that of the modulated intensities.
in this case, but only in a moderated form. This is not unexpected since in Section 5.5.1 it was seen that such a re-acceleration case contains only moderate acceleration effects. In this moderated form the lags, while not as huge as in the case of Figure 7.7, are still significantly larger than those observed in the case of Figure 7.6.

For interest's sake, Figure 7.9 shows the resulting temporal profiles due to a run where both a source on the shock, as well as a boundary spectrum was used, i.e., a combination between the results shown in Figures 7.8 and 7.9. The results are more reminiscent of the re-acceleration case of Figure 7.8 than the source acceleration case of Figure 7.7. If the source function is increased dramatically, the results will approach those shown in Figure 7.7.

To summarize, it is clear that if acceleration is added to the time lag problem, the resultant lags will increase due to the delayed effect of the shock acceleration on the solutions in the inner heliosphere. This provides interesting opportunities for continued modulation studies such as: (a) the correlation of times of maximum intensity of galactic cosmic rays and anomalous cosmic rays after solar minimum periods; (b) specific attention to this problem for Hydrogen. The delay times at various radial distances may be a tool to infer the strength of the anomalous contribution to the total intensity. On the modelling side it is clear that these insights must be developed extensively with the inclusion of drifts.

7.4 Summary

As was shown in Sections 7.2.1 and 7.2.2, acceleration of pick-up ions at the SWTS can quantitatively account for the observed spectra and radial gradients of the anomalous component of cosmic rays. From the fits in Section 7.2.2 and the temporal variations in Figure 7.9, it seems unlikely that anomalous Hydrogen can be detected in the inner heliosphere. In the outer heliosphere, however, one should almost certainly be able to detect the anomalous Hydrogen component. Steenberg (1995) motivated this conclusion in greater detail.

The preliminary study of the temporal effects in the modulation in Sections 7.3.1 and 7.3.2 showed that time lag effects and hysteresis effects can be produced with our time-dependent model. In Section 7.3.2 it was shown that shock acceleration has a profound effect on these lags, so that they can become very large. In addition to this, the results suggest that it may be difficult to correlate the outward propagation of disturbances with features in anomalous cosmic-ray intensities.

In the future, studies of these time lag effects must become more detailed with emphasis on the origin and nature of these lags in time.
Chapter 8
Summary and Conclusions

The primary objective of this thesis was the development of a numerical model with which the time-dependent TPE can be solved. The ability to solve the TPE across an accelerating SWTS, was the basic requirement for this numerical model. Therefore, as such, there was no hypothesis to test in this thesis. Instead, this thesis is dedicated to the development of this numerical model and to discuss its characteristics in detail.

Chapters 1 to 3 functioned as an introduction where the necessary physics of cosmic-ray transport in the heliosphere was discussed. The development work for the numerical solution was done in Chapter 4. A modified LOD scheme was used to solve the TPE. The treatment of the spatial part of the TPE was quite standard, with Crank-Nicolson schemes used to solve the two component equations given by (4.57) and (4.58). The treatment of the rigidity part of the equation, given by (4.59), was not standard. A quasi-analytical treatment of this part of the TPE, based on the method of characteristics, was blended into the comprehensive LOD scheme. This treatment of the energy changes in the TPE accounts for much of the model's flexibility, since no restrictions are placed on $\nabla \cdot \mathbf{V}$.

The incorporation of the SWTS that accelerates particles requires the solution of the matching condition (3.44), which was numerically solved by the (relatively standard) Wendroff implicit solution scheme. The main technical contribution of this thesis was the incorporation of this numerical treatment of the shock with the above-mentioned two Crank-Nicolson schemes and the method of characteristics into the overall LOD-based scheme.

In Chapters 5 and 6 the characteristics of the model and the solutions it produces were discussed. Chapter 5 contained a discussion of the effects of current sheet and shock drift on the acceleration solutions produced with our model. Chapter 5 also contained a theoretical continuation of the data-based species scaling studies of Cummings et al. (1984), and Cummings and Stone (1987, 1990) with the aid of our solution of the TPE. Chapter 7 showed a few preliminary applications of the model. Simple data fits on He$^+$, O$^+$ and H$^+$ were shown to illustrate the model's capability to produce self-consistent solutions that agree quantitatively with observed data. It was also shown that the model can be successfully applied to study time-dependent effects in the modulation.

This time-dependent, two-dimensional acceleration model is the second of its kind. The first was developed by Jokipii (1986). The present model was developed to ensure that we can stay competitive in the cosmic-ray physics community. The developed model will function as the starting point of several research projects, the first of which has already reached completion with the M.Sc. thesis of Steenberg (1995), simultaneously with this work. Steenberg applied the model to the problem of anomalous Hydrogen that was inferred at 24 AU by Christian et al. (1988). He also did the first time-dependent studies on the emergence of the anomalous component with the aid of a numerical solution of the time-dependent TPE.

For the future, the newly-developed model should be applied to the following problems:
The study of Steenberg (1995) on anomalous Hydrogen is not conclusive. According to his work, the model should be applied to the problem with a much greater emphasis on the time-dependent nature of the origin of anomalous Hydrogen during the full solar cycle.

Probably the greatest drawback of almost all our model studies up to now is the absence of gradient and curvature drift in the applications. The reason for this is given in Moraal (1993). In the future, much attention should be given to the drift problem to obtain a single set of parameters with which the known observations of all galactic and anomalous species can be explained, irrespective of the solar magnetic state.

The work on time-dependent modulation by several authors, e.g. Perko and Fisk (1983), Le Roux (1990), Potgieter and Le Roux (Potgieter and Le Roux, 1992a, b; Le Roux and Potgieter, 1992a, b), should be continued with a systematic approach by first studying the effects of simple time dependencies of all possible parameters, one by one, until the underlying physics is well understood. Thereafter, one can attempt to model the 11-year cycle with multiple time-dependent parameters. Special attention should be given to dynamic effects on the termination shock. The effects of a time-variable shock position and compression ratio on the resultant cosmic-ray distributions and spectra are largely unknown.

Little is known of the conditions outside the SWTS. The newly-developed model can be a valuable tool with which this region, and the effects thereof on cosmic-ray transport, can be probed.

A large amount of experience is needed to apply this model to actual data fits, especially in view of the approaching 1996/97 solar minimum. With four functioning spacecraft in the outer heliosphere, numerous near-Earth observations and two previous cosmic-ray cycles to compare the observations with, this solar minimum period should produce a wealth of new results.

The present solution is based on a pure test-particle treatment of the shock. Future developments must include the non-linearities due to the back-reaction of the accelerated cosmic rays on the shock structure.

The problem of travelling interplanetary shocks can be addressed by a derivative of this model. Such travelling interplanetary shocks have leading and trailing shocks where particles can undergo a moderate amount of acceleration. Such effects, built into a fully time-dependent acceleration model, may provide invaluable insight into the problem of the effects and importance of CIRs, MIRs and GMIRs in establishing the 11- and 22-year modulation cycles.

The present model can, with minor modifications, be applied to supernova remnants. The most important of these modifications should be a radial grid expanding with time. Jokipii and Ko (1987) constructed a first model of this kind. Such a model should be able to successfully address the problem of the production and origin of cosmic rays.

Cosmic-ray transport in the galaxy can be modelled with the aid of dynamical halo models as has already been demonstrated by Jokipii and Morfill (1987). In these models a galactic wind, similar to the solar wind, is proposed that, on the average, blows away from the galactic plane. Therefore, this is also a convection-diffusion problem and solutions of the TPE, like our model, can be successfully applied to this problem.
Hess, V. 1911, *Physikal. Zeits.*, 12, 998
Hyde, V. 1911, *Physikal. Zeits.*, 12, 998
———. 1988, *Cospar (Helsinki)*


———. 1965, Plan. Space Sci., 13, 9


Potgieter, M.S. 1984, Ph.D. thesis, Potchefstroom University of CHE, South Africa


185
Webb, G.M. 1975, Ph.D. thesis, University of Tasmania, Australia
Zirken, J.B. 1977, Coronal Holes and High Speed Wind Streams, Colorado Associated University Press, Boulder