Chapter 1. Introduction

This chapter sets the stage for what follows through a brief presentation of the known characteristics of near-Earth space, with an emphasis on the observational basis of our knowledge. Data are presented which lead to a generalized model of the interaction of the solar wind with the Earth. This model forms the framework for later discussions.

a. Solar Wind and Magnetosphere

The driving force behind the solar-terrestrial interaction is the solar wind. This flow arises from the thermal imbalance of the hot (temperatures exceed 10^6 K) corona of the Sun. This imbalance drives an outflow of plasma from near-solar regions which passes Earth's orbit and has now been observed by the Voyager spacecraft to continue past the outermost known planets. Coronal heating is likely due to the presence of magnetic fields in the solar plasma [Priest, 1995], and these also lead to the possibility of interaction with near-Earth plasmas and magnetic fields as will be discussed below. Figure 1.1 [National Space Science Data Center, 1989] illustrates the fluid properties of the solar wind as seen near Earth for the period March 27 to April 22, 1986. CDAW events A and B, which will be discussed in Chapters 7 and 6, respectively, occur on April 1 and April 3. The solar wind bulk speed, as shown, varies between approximately 300 and 700 km/s and a typical value would be 450 km/s. The density of ions (primarily protons, H^+) varies between 3×10^6 and 3×10^7 m^{-3} (shown as 3 to 30 per cc on the figure) and a value of 10^7 m^{-3} could be taken as representative. Much higher values are possible, as suggested by occasional observations of 10^8 or more particles per m^3 (see Chapter 10). The temperature varies from 10^4 K to about 3×10^5 K, again typical values but considerably less than in the solar coronal source regions. A temperature of 10^5 K corresponds to a particle thermal energy of about 10 eV (one eV being 1.6×10^{-19} J): both this and the flow speed of order 0.001 c (c being 3×10^8 m/s) mean that the solar wind is a nonrelativistic flow. The kinetic energy associated with ion bulk flow is about 1 keV per particle, so dynamic pressure greatly exceeds thermal pressure in this flow. As might then be expected, the sound speed, $c_s = \sqrt{\frac{\gamma p}{\rho}}$, with $\gamma = \frac{5}{3}$ the ratio of specific heats at constant pressure, p the thermal pressure, and ρ the mass density, is less than the flow speed. Since the plasma must be approximately neutral and is primarily made of hydrogen, the number density of electrons is approximately equal to that of protons. Thus, the total number density is about twice that, n, of protons, and the pressure is $p=2nkT$, k is the Boltzmann constant of 1.38×10^{-23} $J K^{-1}$, and T the observed temperature. This results in a thermal pressure of 3×10^{-11} Pa. The mass density is simply the number density times mass for protons since protons and electrons are the main constituents, with the mass of electrons negligible. Thus the mass density is about 2×10^{-20} kg/m^3, and the sound speed may be calculated as 50 km/s. As expected this is considerably less than the bulk flow speed. The solar wind is thus a supersonic flow. The actual bulk flow direction (not obtainable from the

data shown) is approximately radial [Hundhausen, 1995]. The energy flux in the bulk flow is then $\frac{1}{2}\rho v^3$, which is readily calculated as 9×10^{-4} W/m^2, roughly 10^{-6} of the solar constant of 1 kW/m^2 which is mostly in the form of photons. Clearly the effects of the solar wind bulk flow will be minimal in areas able to interact with photons (such as those near the Earth's surface). This is not the case in ionized outer regions still primarily under the Earth's influence through its magnetic field: these regions are profoundly affected by the solar wind. However, the subtlety of solar wind effects as detected at the surface of the planet is primarily due to the small amount of energy carried in that flow compared to that of the photon flux.

Figure 1.1 Solar wind fluid properties observed with the IMP-8 satellite.

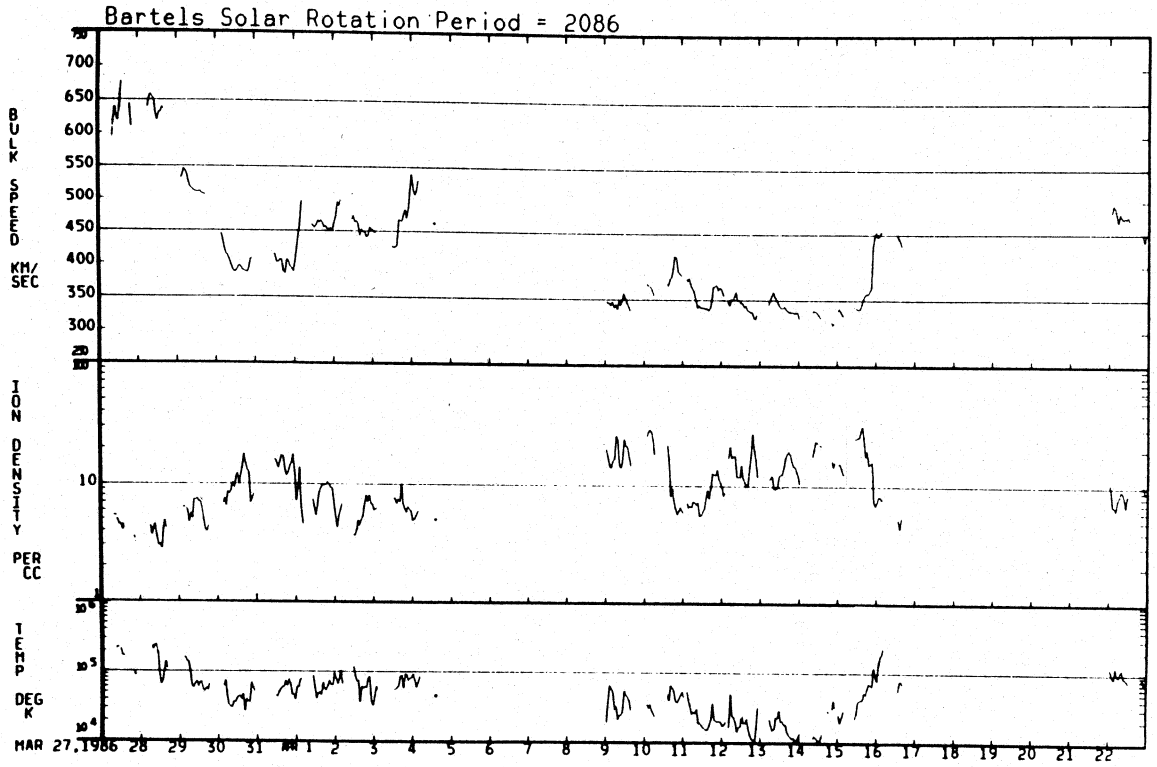
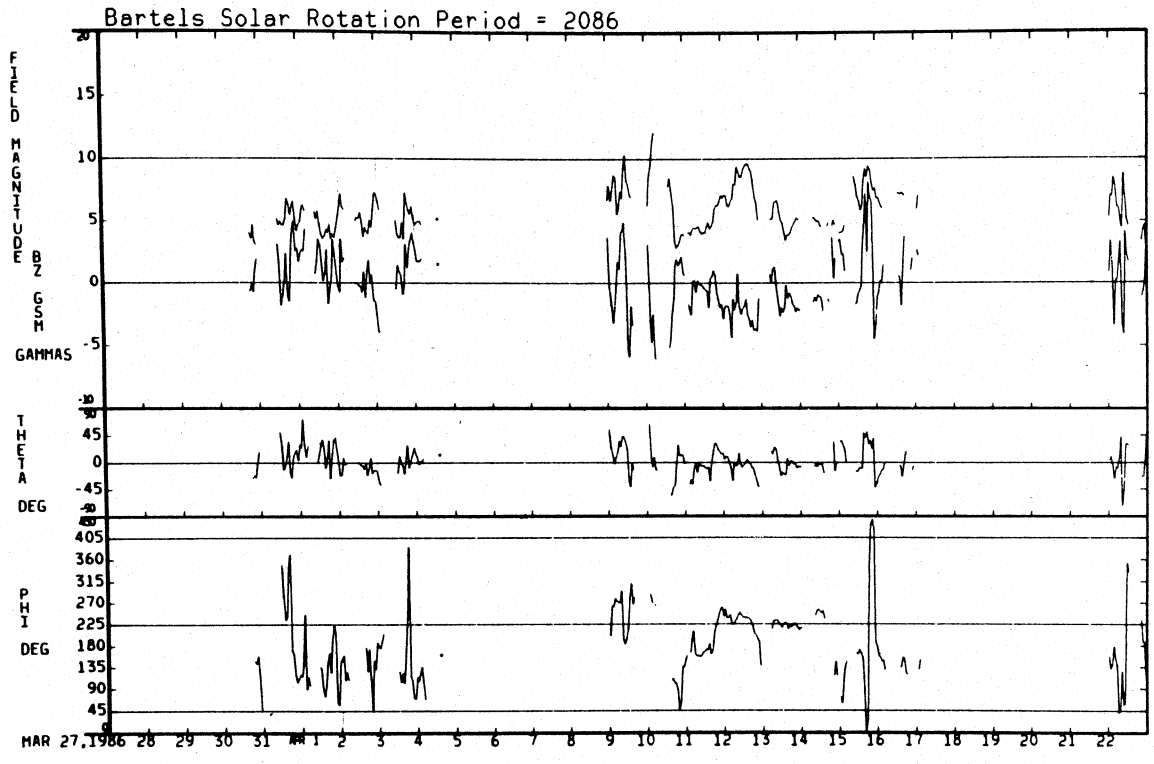


Figure 1.2 shows the observed magnetic properties of the solar wind during the period for which Figure 1.1 showed the fluid properties. The magnetic field associated with the solar wind is usually referred to as the interplanetary magnetic field (IMF) [Parks, 1991, p. 3]. The field magnitude near Earth is variable (here between about 3 and 12 nT) with 5 nT being a 'canonical' value. The indicated GSM Z component of the field has a mean which is slightly above zero in the period before 5 April and slightly below zero in the period 8-16 April. It is unsteady and crosses the zero line frequently. The GSM Z coordinate axis is parallel to the projection of Earth's dipole in a plane perpendicular to the ecliptic (see section 3.e) and the GSM Z component of the solar

wind field has been shown both theoretically and by statistical studies to be relevant to the question of energy coupling between the solar wind and the magnetosphere. In this figure, theta is the angle above the ecliptic of the field vector and phi is the angle of the ecliptic projection of the field from the Earth-Sun line. Careful inspection will show that the theta polar angle of the field vector shown does not correspond precisely to the arctangent of GSM Z: this angle is in fact the arctangent of GSE Z (again see section 3.e) and corresponds to the angle of the field to the ecliptic. The mean of theta seems to be near zero. The phi angle shown is that between the Earth-Sun line and the ecliptic projection of the field. Although subject to occasional large excursions, this angle is typically 135°. Although bulk flow is radial, the rotation of the Sun and the frozen-in field condition (see section 2b) cause magnetic field lines to be skewed at this angle from the Earth-Sun line. The magnetic field exerts a pressure and with a typical value of 5 nT, this pressure, $\frac{B^2}{2\mu_0}$, is 1.0×10^{-11} Pa, which is less than, although of the same order as, the thermal pressure. Both are small compared to the dynamic pressure or momentum flux of the bulk flow (which is $\rho v^2=4\times10^{-9}$ Pa)[1], as might be expected from the above discussion of energy content.

Figure 1.2 Solar wind magnetic properties observed with the IMP-8 satellite. Uppermost curve is field magnitude in nT (marked as gammas).

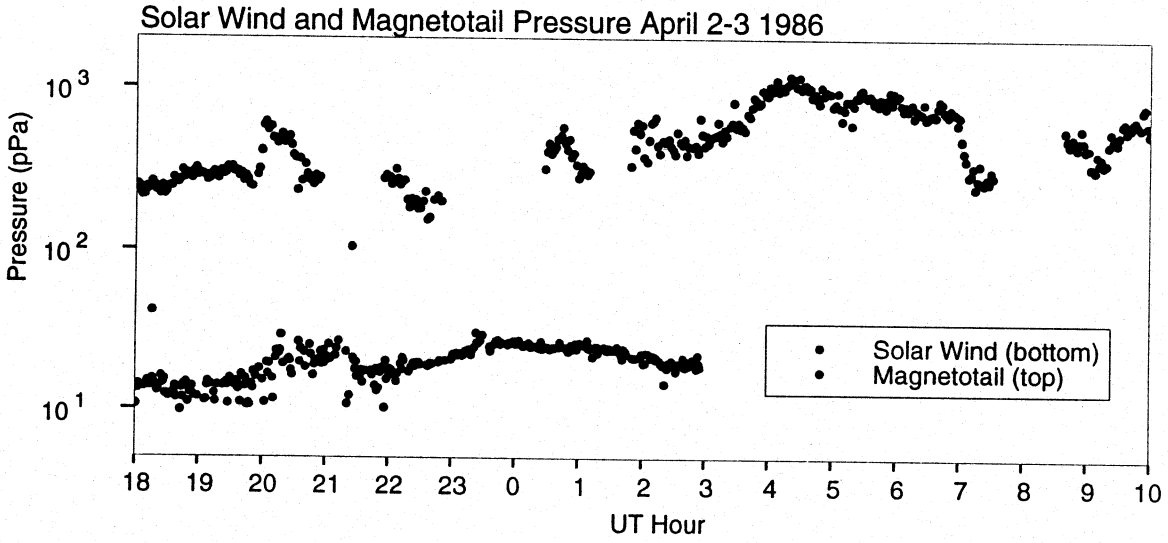


[1] Hundhausen [1995] cites 2.6×10^{-9} Pa. In SI units, 10^{-9} Pa=1 nPa.

The formation of a magnetosphere around the Earth may be understood by considering that the planet has an intrinsic magnetic field. The magnetosphere is a region whose existence is due to the tendency of this planetary magnetic field to exclude the solar wind. To a first approximation the Earth's intrinsic field is that of a dipole of magnetic moment 8×10^{15} Tm^3 or equivalently 30400 nTR_e^3 where R_e is one Earth radius (6371 km). In the magnetic equatorial plane such a field is entirely vertical (perpendicular to the plane) and its strength is Mr^{-3}, where M is the dipole moment and r the radial coordinate. The magnetic pressure, $\frac{B^2}{2\mu_0}$, is thus $\frac{M^2r^{-6}}{2\mu_0}$. Earth's dipole is inclined at only 11° from the spin axis. This in turn is inclined at the 23.5° obliquity from the perpendicular to the ecliptic, so the subsolar point always lies relatively close to the magnetic equator. The maximum magnetic latitude of the subsolar point is observed at solstices and is then roughly 35°. A rough idea of the size of the magnetosphere can be obtained by balancing the two dominant sources of pressure above the subsolar point (i.e. along the Earth-Sun line) somewhere between the Earth and the Sun. The dynamic pressure dominated region of the solar wind and the internal field dominated dayside magnetosphere could be thought in this first order picture to exist in static equilibrium. Equating the magnetic pressure to the dynamic pressure at the subsolar point, the pressure balance would be such that 3.7×10^{-4} = 4×10^{-9} L_s^6, where L_s is the 'standoff distance' expressed in units of Earth radii. This rough estimate[2] of standoff distance yields 6.7 R_e. In reality the dayside of the magnetosphere is compressed by the solar wind, enhancing the magnetic field over dipolar values, and thus the observed standoff distance is normally larger, about 10 R_e. It also follows that higher speed solar wind flows, with enhanced dynamic pressure, reduce the standoff distance. On the nightside the effect of the solar wind, rather than compressing the intrinsic field, is to sweep back plasma and associated magnetic flux into a long 'comet-like' magnetotail. That the dynamic pressure is the dominant effect in magnetotail formation is illustrated by Figure 1.3 which shows, using data from IMP 8 in the solar wind and ISEE in the magnetotail, that the static (thermal plus magnetic) pressure in the solar wind is one to two orders of magnitude less than that in the magnetotail. The much greater microscopic pressure in the magnetotail shows that solar wind dynamic pressure is needed to explain the confinement of the magnetotail. The component of solar wind dynamic pressure on the flanks of the magnetotail is almost entirely responsible for its confinement.

[2] The standoff distance varies as only the one-sixth power of dynamic pressure. With 2.6×10^{-9} Pa dynamic pressure, the standoff distance in this calculation changes to 7.4 R_e.

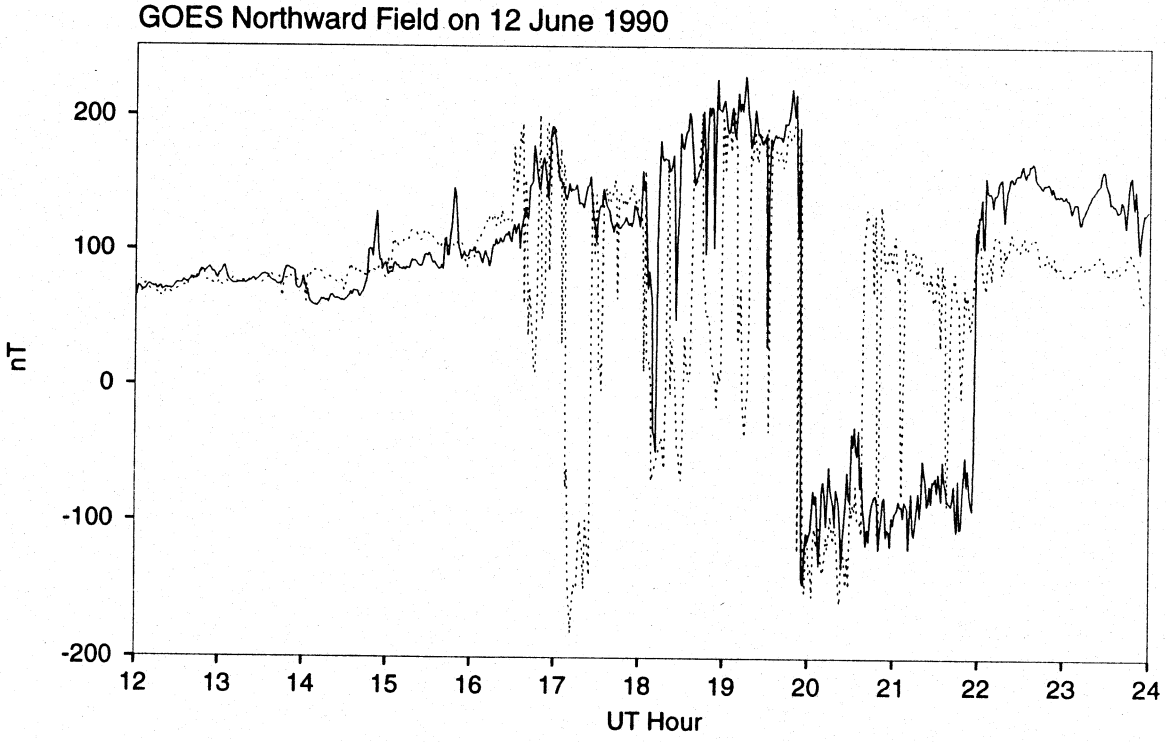
Figure 1.3 Thermal and magnetic pressure sum (in SI pPa units) in the solar wind and the magnetotail. Data gaps exist throughout the magnetotail data and in solar wind data starting at 3 UT.



In reality the interaction between the magnetic field dominated magnetosphere and the dynamic pressure dominated solar wind is more complex than the simple balance at an imagined interface described above. The diversion of the supersonic solar wind flow occurs at a bow shock inside which conditions abruptly change from those of the solar wind. The bow shock is observed to be in front of the actual boundary of the magnetosphere, which is the region dominated by fields directly associated with the Earth. This boundary is called the magnetopause and the transition between internal and external fields is associated with a magnetopause current, usually referred to as a Chapman-Ferraro current. Spacecraft traversals of this current may be seen as abrupt magnetic field changes in Figure 1.4, which shows a case on June 12 1990 when the magnetopause had been pushed back to geosynchronous orbit. The abruptness of the transitions indicates that this current is very thin: in fact it is typically only a few hundred km thick [Hughes, 1995]. The field is seen to basically reverse on transition, rather than dropping to the very small interplanetary field value: an exterior region exists having fields roughly comparable in magnitude to those inside the magnetopause. This region between the magnetopause and the bow shock is referred to as the magnetosheath. On the dayside it can be thought of as containing slowed-down solar wind plasma and a magnetic field which is related to that of the solar wind. After 16 UT two GOES spacecraft (GOES 6 and 7) both detected fields stronger than the nominal 100 nT field at the equatorial plane in synchronous orbit, presumably

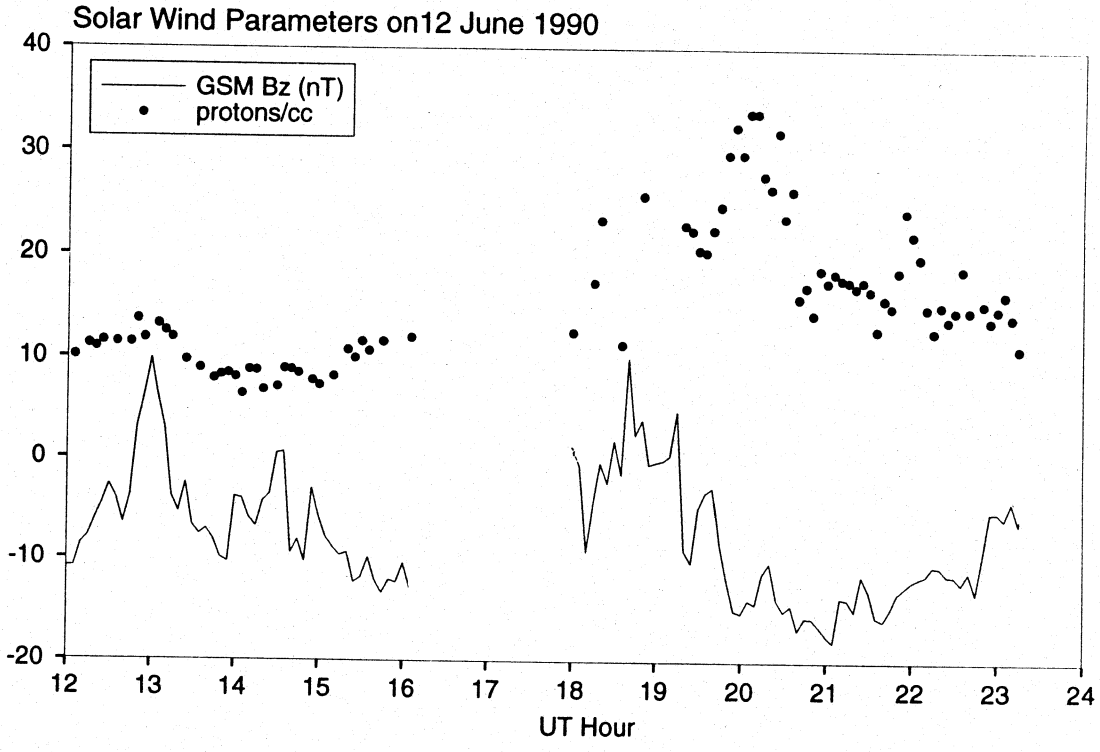
evidence of compression of the magnetosphere. That compression occurred is further evidenced by the observation of magnetopause crossings at synchronous orbit, that is, only 6.6 R_e from the Earth as opposed to a more typical 10 R_e. On this date, mean solar noon for GOES 7 occurred at 18.84 UT, and for GOES 6 at 21.00 UT. Prior to about 20 UT, magnetopause crossings occur mainly at GOES 7, which was closer to the subsolar point. Field values during these crossings attain about -50 nT, except during one crossing by GOES 7 at 17.0 UT. By 20.5 UT, this spacecraft had rotated away from noon sufficiently to not continuously observe the magnetosheath, while GOES 6 remained immersed in it. Before further discussion, it is instructive to examine simultaneous solar wind data.

Figure 1.4 Dayside magnetopause crossings by GOES spacecraft on June 12 1990. Line representing northward component of magnetic field at GOES 6 is solid, at GOES 7 dotted.



The IMF had a southward component during much of the period shown, as seen in Figure 1.5. The proton density in the first part of the interval shown is a near-canonical 10^7 m^{-3} (shown as 10/cc), with stepwise increases to roughly three times this value by just after 20 UT. These data were obtained by the IMP-8 spacecraft which was in the solar wind at roughly 8 R_e sunward, 25 R_e duskward, and 20 R_e southward of the Earth (at 18 UT). The solar wind speed increased over several hours from a steady 550 km/s to nearly 800 km/s just after 18 UT and declined slowly thereafter (data not shown). Unfortunately the data gap seen in Figure 1.5, between roughly 16 and 18 UT, also is present in the speed measurements and does not permit details of the rise in solar wind speed to be specified.

Figure 1.5 Solar wind parameters on June 12 1990. The northward component of the IMF in nT is shown as a solid line, proton number density per cm^3 as dots with a common numeric scale.

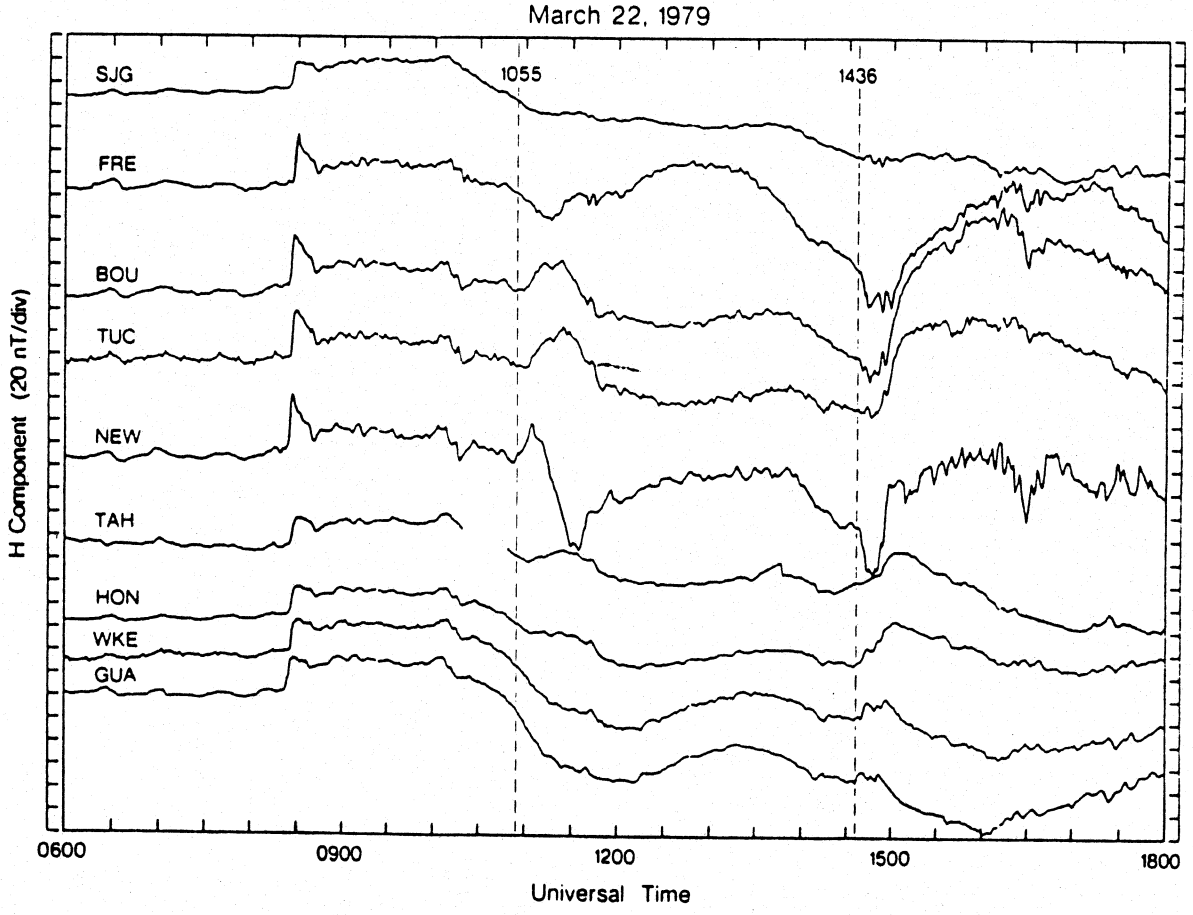


Taken together, the solar wind data and magnetopause crossings illustrate the three factors in the solar wind which may perturb the magnetosphere. The first two, already alluded to, are density and speed as manifested through dynamic pressure. From the figures shown, clearly the major perturbation of the magnetopause occurred between about 20 UT and 22 UT. During this period, the density was above average, and the speed well above average. The third factor is the orientation of the IMF, and in particular the northward component. During the major perturbation, as Figure 1.5 shows, this was solidly negative (i.e. the field was 'southward'). Under these circumstances 'reconnection' of interplanetary and planetary field lines can occur, stripping flux and plasma from the dayside magnetosphere and transferring it to the nightside [Hughes, 1995]. This process is likely of great importance in energy transfers within the magnetosphere.

The example above uses satellite data to illustrate changes in the magnetopause. Ground observations also permit detection of buffeting of the magnetosphere by the solar wind. When its propagating shock waves bring pressure pulses into contact with the magnetosphere, compression results, with increase of the magnetic field inside. This may be seen from two points of view (as turns out to often be the case in describing magnetic fields in plasma). One is that the magnetic flux is essentially frozen

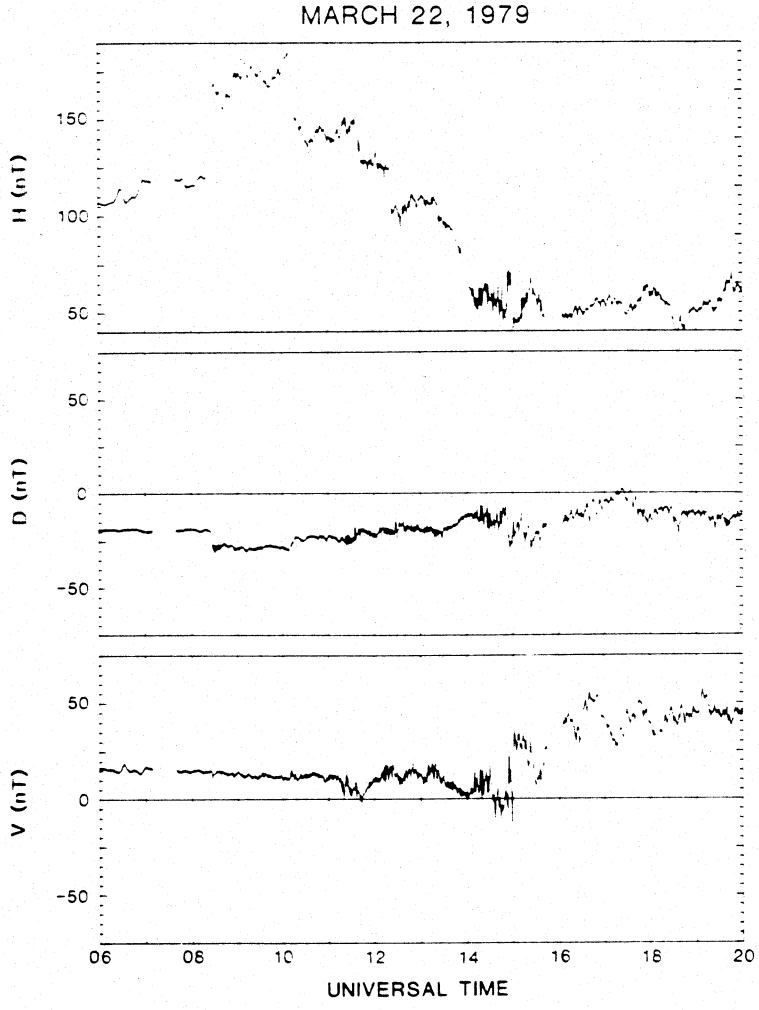
in to the highly conducting plasma, so that simple hydrodynamic compression would lead to more flux in less space and thus a field increase. The other is that the magnetopause boundary currents (Chapman-Ferraro currents) strengthen to move the magnetopause boundary inward. In any case, a northward component increase is seen inside the magnetosphere when it is compressed. Such a compression is referred to as an SI (sudden impulse) or an SSC if it is followed by a geomagnetic storm. A very clear example is provided in the well-studied CDAW 6 event of March 29, 1979 [McPherron and Manka, 1985]. A shock in the interplanetary medium caused an abrupt increase in the solar wind velocity and density, and thus an increase in the dynamic pressure, at 0826 UT as seen near Earth. Simultaneously, ground stations worldwide recorded an increase in the northward component as seen in Figure 1.6.

Figure 1.6 Midlatitude H components on March 22 1979. SI at 0826 UT is evident while decrease at 1008 is less so. Substorm onsets at 1055 and 1436 are marked by vertical dashed lines and are accompanied by mid-latitude H bays.



The GEOS 2 synchronous spacecraft, in the vicinity of noon, detected an increase in the northward component from a relatively normal 120 nT to 170 nT in a matter of minutes at essentially the same time as shown in Figure 1.7. Coincident with the passing of a shock-associated current sheet at 1008 UT, the density abruptly decreased by a factor of about 3 to its pre-0826 value of about 10 particles/cm^{-3} (data not shown). The resultant decrease in dynamic pressure led to an immediate decrease in the northward component at GEOS 2. Signals on the ground also show this decrease, although it is not as clear as at the satellite. This behaviour clearly illustrates the effects of solar wind in controlling the size of the dayside magnetosphere. The 1008 UT current sheet passage also turned the interplanetary magnetic field southward, and this led to substorm activity in the magnetosphere, of a type which will later be discussed and seen to be characteristic.

Figure 1.7 Magnetic field observations at the GEOS 2 synchronous spacecraft on March 22, 1979. The H (northward) component responds to compression of the magnetosphere by increased solar wind dynamic pressure at 0826 UT, while the other components are affected less. Pressure decrease at 1008 allows the northward component to decrease. Figure from McPherron and Manka [1985].



The effects noted above are associated with relatively rapid changes in the solar wind. Recurrence of magnetic activity also follows the synodic rotation period of (near-equatorial regions of) the Sun, which is 27 days. This indicates that the causes of magnetic activity in the solar wind are generated by long-lived regions, likely near the solar surface. Due to the influence of the north-south component of the interplanetary magnetic field, relative to the orientation of the Earth's dipole, there exists a seasonal effect also [Russell and McPherron, 1973]. On yet a longer timescale, it is well known that overall activity of the Sun varies periodically in an 'eleven-year cycle', which in fact is a 22 year cycle if the polarity of the Sun's reversing main field is taken into account. The number of sunspots varies dramatically through the eleven year cycle and the level of magnetic activity observed on Earth is roughly correlated with the number of sunspots although not as variable. The modern study of auroral phenomena may be considered to date from the coordinated studies during high solar activity in 1957, known as the International Geophysical Year (IGY) [Akasofu, 1964].

b. Ionosphere

The transitional region between the neutral atmosphere and space is known as the ionosphere. In this region matter is ionized to a significant but variable degree, the primary sources of ionization being photonic radiation and particle fluxes. Recombination processes in the ionosphere, primarily in response to excitation through precipitation of energetic particles, provide one of the more inspiring forms of evidence of the solar-terrestrial interaction: the light of the auroras. In the lower ionosphere interaction with neutral molecules is a significant factor and the ions are dragged along through the Earth's magnetic field. This creates a unipolar inductor and creates a corotational electric field around the Earth.

As a plasma in a magnetic field, the ionosphere may be characterized by a conductivity tensor whose characteristics are dependent on frequency. The properties of this tensor at low frequencies (low compared to particle gyrofrequencies or the ionospheric plasma frequency) are discussed in the following chapter. At higher frequencies, in the radio range, the ionosphere can be dispersive, birefringent, and absorptive, and these properties are of practical importance and interesting in introductory discussions (see Jackson [1975], chapter 7, for example). Here we will simply note that the ionosphere and its disturbances can be of practical interest. Further, the ionosphere may be monitored and studied at radio frequencies. At these relatively high frequencies, phenomena occur near the electron plasma frequency which result in reflection of incident waves. Since the electron plasma frequency is $\omega_e = \sqrt{\frac{ne}{\varepsilon_0 m}}$, where n is the electron density and the other quantities physical constants (e the electron charge and m the electron mass), this reflection frequency may be used to make a profile of electron density at various altitudes. The important physical parameter of conductivity is largely dependent on electrons, whose density may approach 10^{12} m^{-3}. The altitude

profile of the electron density may be ascertained by timing returns of pulses at varying frequencies giving reflections, and is called an ionogram [Davies, 1966]. The same technique also works if a satellite directs pulses downwards and is then usually referred to as topside sounding. Detection of increases in electron density may also be done by passive means, that is by monitoring natural radiations from distant space and noting that they decrease when precipitation of energetic particles causes the electron density to increase. Such detectors are called riometers. Sounding rockets and satellites also provide *in situ* measurements of ionospheric composition and density.

Ionospheric structure, as revealed by the above and other techniques, is complex. It is useful to speak of layers, these being, from bottom to top, the D region found below 90 km, the E region between 90 and 130 km, and the F region above 130 km. As will be discussed in the next chapter, for practical purposes, a height of 110 km is used as the 'nominal' height of ionospheric current flow in discussing low frequency magnetic perturbations. This height is used in modelling, with further justification below.

A further role of the ionosphere in magnetospheric studies is as a source or sink of particles. Since the Earth and Sun have differing cosmic histories, their bulk composition is different. In particular, Earth's atmosphere is abundant in oxygen and nitrogen, while the Sun, and thus the solar wind, is more abundant in hydrogen and helium (present in isotopic ratios not typical of Earth). Composition analysis can thus be useful in determining the sources of magnetospheric plasma.

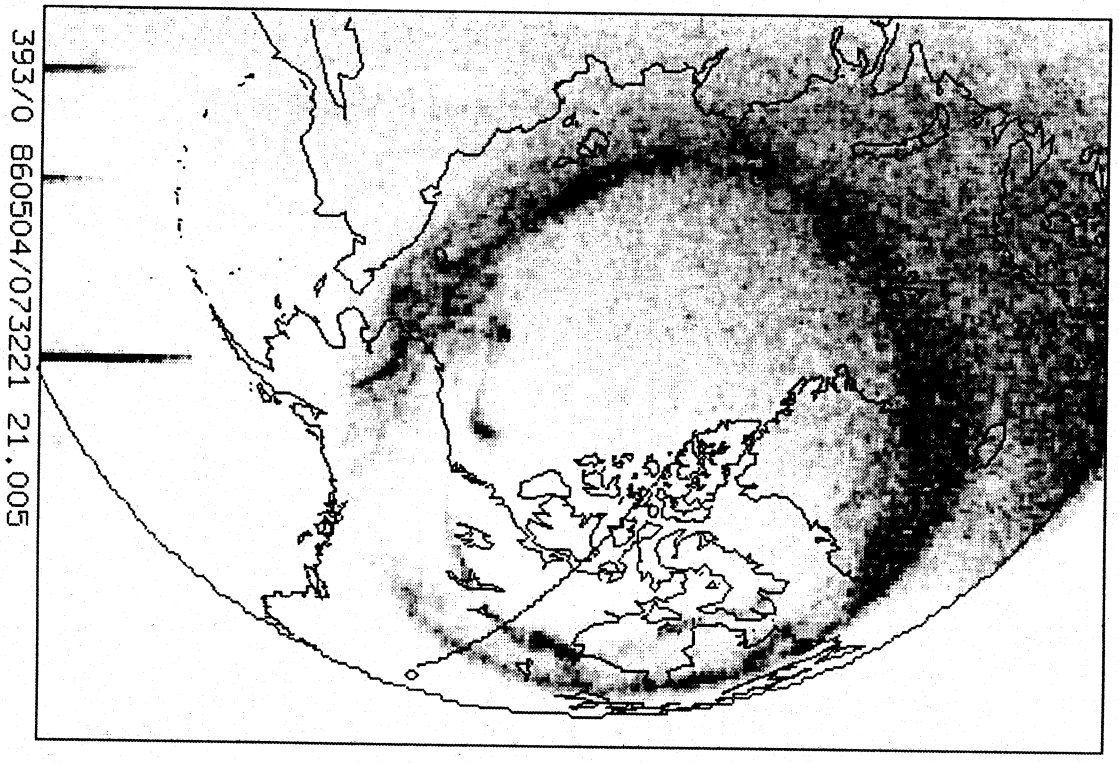
c. Magnetosphere-Ionosphere Coupling

Evidence of the interaction of the magnetosphere with the ionosphere is provided by the auroral emissions which result from the interaction of particles, accelerated in the magnetosphere, with the upper atmosphere. In addition to energy coupling by this mechanism of particle precipitation, there is electromagnetic coupling since currents flow through connected regions of the ionosphere and magnetosphere. Thus the two main types of Earth-based observations of the magnetosphere-ionosphere coupling have relied on detecting photons arising from the particle precipitation, or on detecting the magnetic fields associated with the currents. The currents are of course in most cases associated with particle precipitation in some way.

Both auroral emissions and magnetic perturbations are observed to be most pronounced in belts roughly centred on the magnetic poles of the Earth (see Figure 1.8), but usually about 20° equatorward. The large region where these phenomena are observed varies according to the level of the solar-terrestrial interaction, with the region where they are commonly observed being referred to as the auroral zone. The part of this region which is at any one moment active is referred to as the auroral oval as this is statistically, and for the most part instantaneously also, its shape. The most active part of the auroral oval is found near local midnight and the midnight sector is offset by at least 3° anti-sunward with respect to the magnetic pole [Akasofu, 1968].

Stations at typical auroral zone latitudes of about 65° in the magnetic dipole system are usually in the oval and detect activity most strongly around local midnight. Stations at higher latitude may cross the oval twice per day, as it remains relatively fixed with respect to the Sun-Earth line while Earth rotation carries them under it. In part due to a dearth of observatories, the highest latitude regions poleward of the auroral zone have historically been less observed than the zone itself. Even in the auroral zone investigations are limited by an insufficient number of observatories.

Figure 1.8 Instantaneous view of the entire northern auroral oval from the Viking spacecraft. Continental outlines have been superposed. Magnetic midnight is marked by a curved line through western Canada. Bars at left are imaging artifacts.



The magnetic effects of auroral currents have been used in the night sector of the auroral zone to trace the average position of the auroral oval [Rostoker and Phan, 1986]. These effects arise from electric current flow in the ionosphere in what is referred to as an electrojet, and the electrojet positions are within the auroral oval. The auroral oval is generally fixed in a sun-oriented coordinate system, that is, with respect to magnetic local time (MLT). Using data from a set of observatories along a meridian of geomagnetic longitude, the overall configuration of the auroral oval may be determined, since rotation of the Earth carries the observatories under it to sample all magnetic local times. As shown in Figures 1.9 and 1.10, the auroral oval borders traced in this way vary both with local time and with activity level.

Figure 1.9 Borders of auroral oval current flow in one-hour averaged magnetic data [Rostoker and Phan, 1986]. Dashed lines indicate borders of westward current flow, solid lines indicate borders of eastward current flow. AE 50 designates a period of globally low auroral oval current. Latitudes are referred to the magnetic pole.

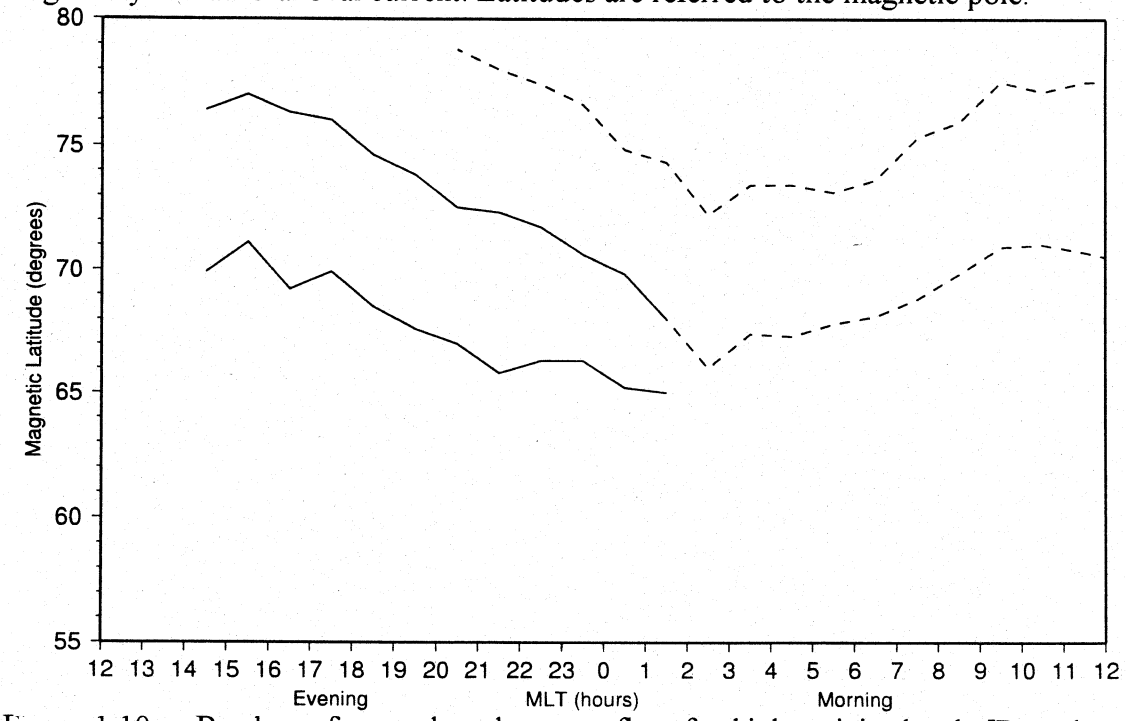
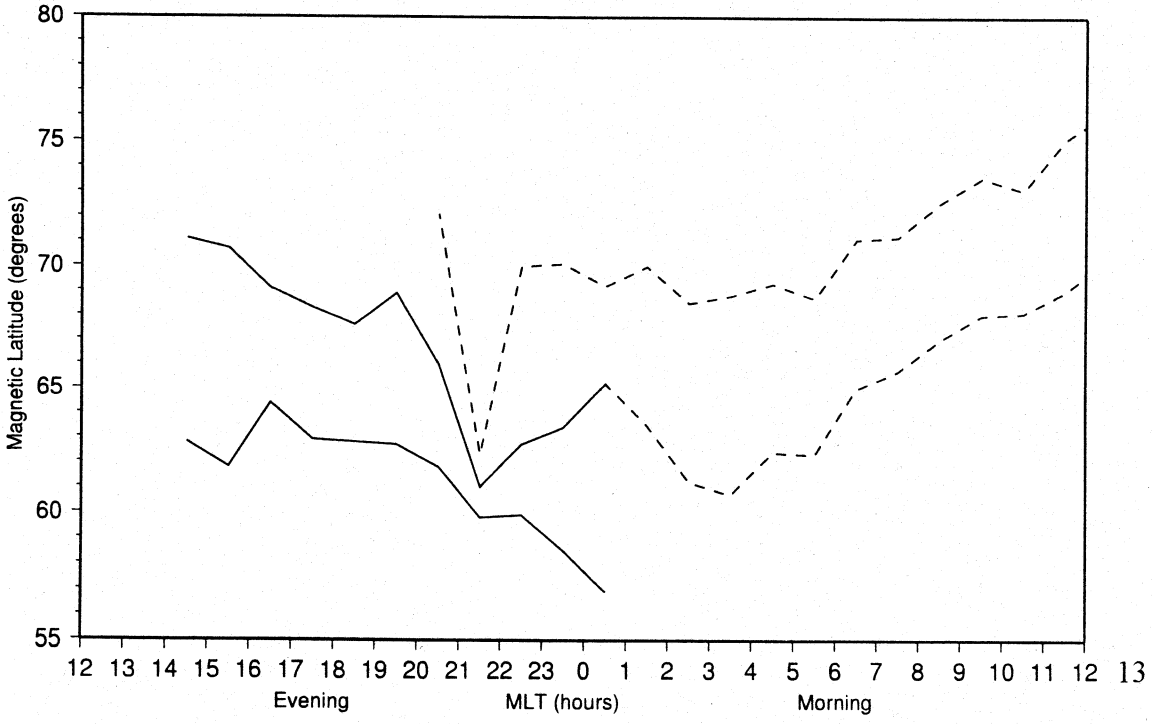


Figure 1.10 Borders of auroral oval current flow for high activity level [Rostoker and Phan, 1986]. Currents in this case are roughly ten times those in the previous figure. Evening sector electrojet in particular has shifted equatorward.



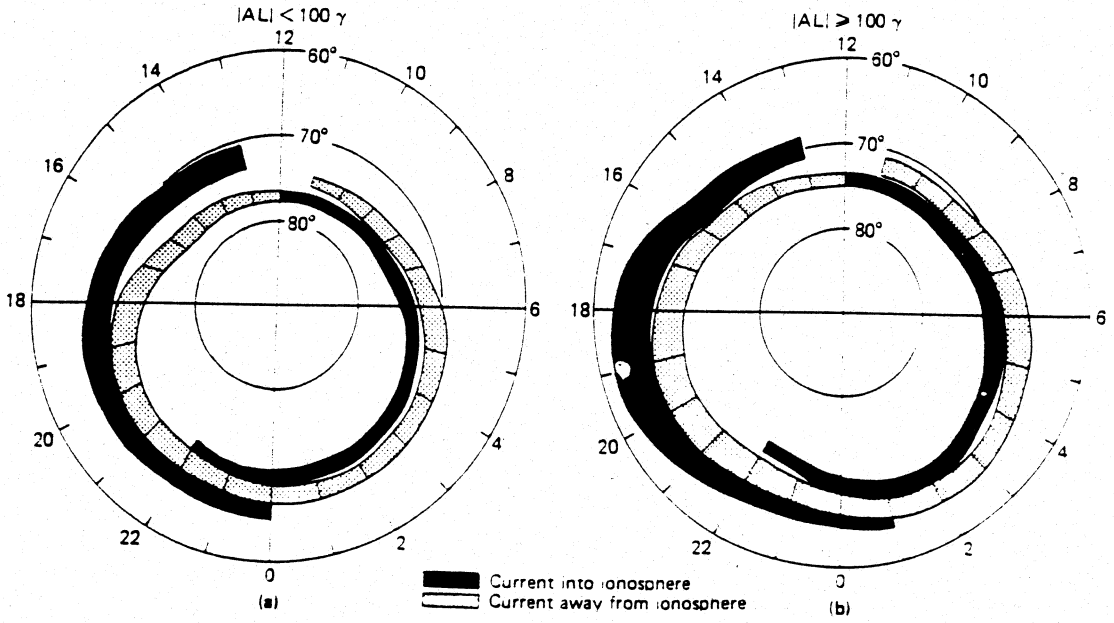
The evening sector is typified by an eastward flow of overhead electrical current at the lowest latitudes, with westward flow poleward of it as the midnight sector is approached. The morning sector has primarily westward current flow. Both in the morning and evening sectors, there is a decrease in latitude as midnight is approached. The overall pattern shifts equatorward by up to nearly 10° with increased activity, with most of the shift being in the evening sector, as seen in figure 1.10. Generally similar results came from studies of ionospheric disturbance at substorm maximum [Besprozvannaya *et al.*, 1991] as a function of local time. These disturbances (including auroral absorption of radio waves) are associated with westward electrojets at high latitude in the evening sector, with a sudden decrease in latitude at 22 MLT, so that the southern border of the active westward electrojet is found at roughly 60° (geomagnetic) latitude. The eastward electrojet was typified by a different sort of ionospheric disturbance and was found by these authors to be south of the westward electrojet in the evening sector, but apparently rarely beyond 22 MLT.

d. Configuration of the Magnetosphere

In this section the overall configuration of the magnetosphere is discussed. The actual causes of the near-Earth currents are the subject of ongoing research. Even the establishment of connections between near-Earth and more distant regions (primarily inferred through attempts to 'map' along field lines, but also deduced from particle distributions) is not generally agreed upon. Here an attempt is made to discuss accepted aspects of magnetospheric structure and the basis for this acceptance.

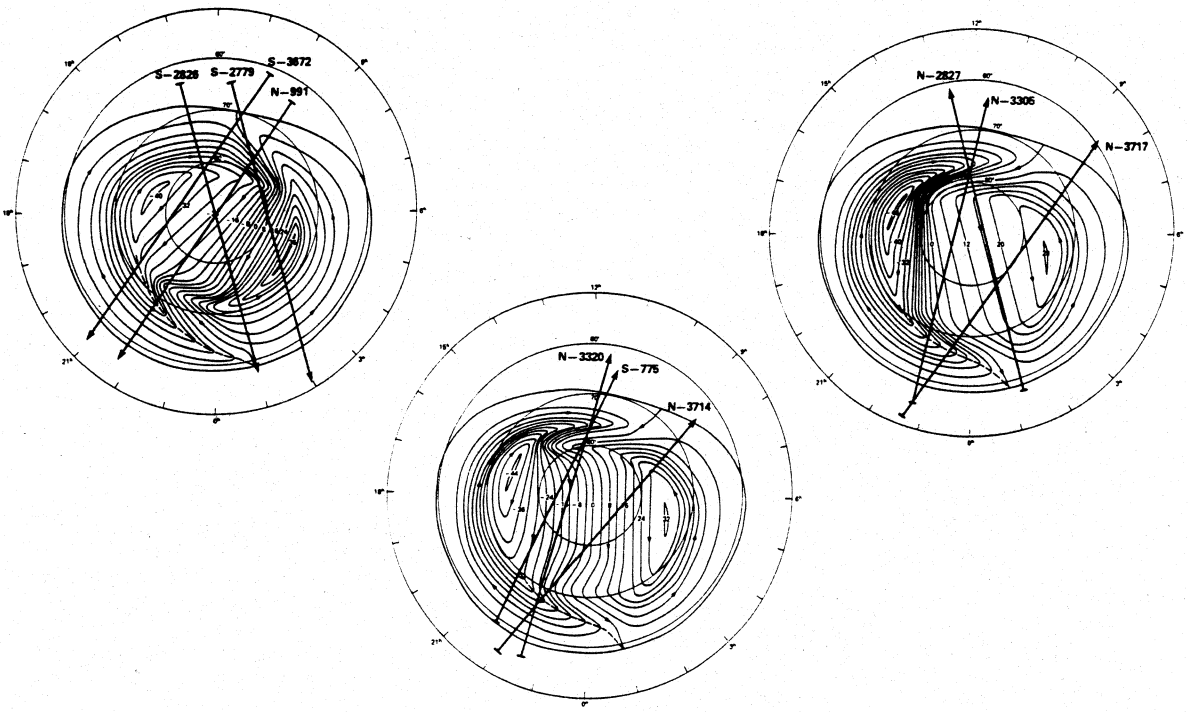
A major historical controversy in space physics concerned the structure of near-Earth space and provenance of auroral particles. As will be discussed in later chapters, magnetic observations from the ground do not allow unambiguous inversion to a three-dimensional causative current system. Despite the seemingly obvious connection between particle precipitation and near-vertical currents, space observations were realized to be needed [Boström, 1964] to distinguish between models of auroral electrodynamics with or without such currents. The initial relevant observations were done in the mid 1960s (see discussion in an historical review by Dessler [1983]), the currents were discovered, and within a decade the results were being systematized.

Figure 1.11 Canonical view of patterns of field-aligned currents observed with the Triad satellite [Iijima and Potemra, 1978]. Data from 439 passes during weakly disturbed conditions (left) and from 366 passes during active periods were averaged to produce the patterns shown.



The establishment, through polar-orbiting satellite observations, that large-scale electric currents flow between the magnetosphere and ionosphere, allowed studies of those currents and their relation to other phenomena to be undertaken. Figure 1.11 shows the configuration of field-aligned currents which exist in a statistically repeatable fashion around the north polar regions of the Earth [Iijima and Potemra, 1978]. These currents were deduced from perturbations observed by the Triad satellite, in polar orbit at 800 km altitude. Interpretation of such perturbations is discussed in further detail in section 3.a.1.c, however here it will suffice to say that a satellite-borne magnetometer at 800 km basically allows determination of currents flowing at or near that altitude. Observations suggest that such currents largely flow parallel to the magnetic field, that is to say, are field-aligned. The basic configuration in both quiet and disturbed times is downward current poleward of upward current on the morning side, while the reverse is true on the evening side. In the noon and midnight sectors a more complex configuration is found. The poleward current zone is usually referred to as Region 1, while the equatorward currents flow in Region 2. Perturbations poleward of the well-defined Region 1 zone are sometimes referred to as Region 0 currents and do not have the same degree of statistical repeatability as those categorized by Iijima and Potemra.

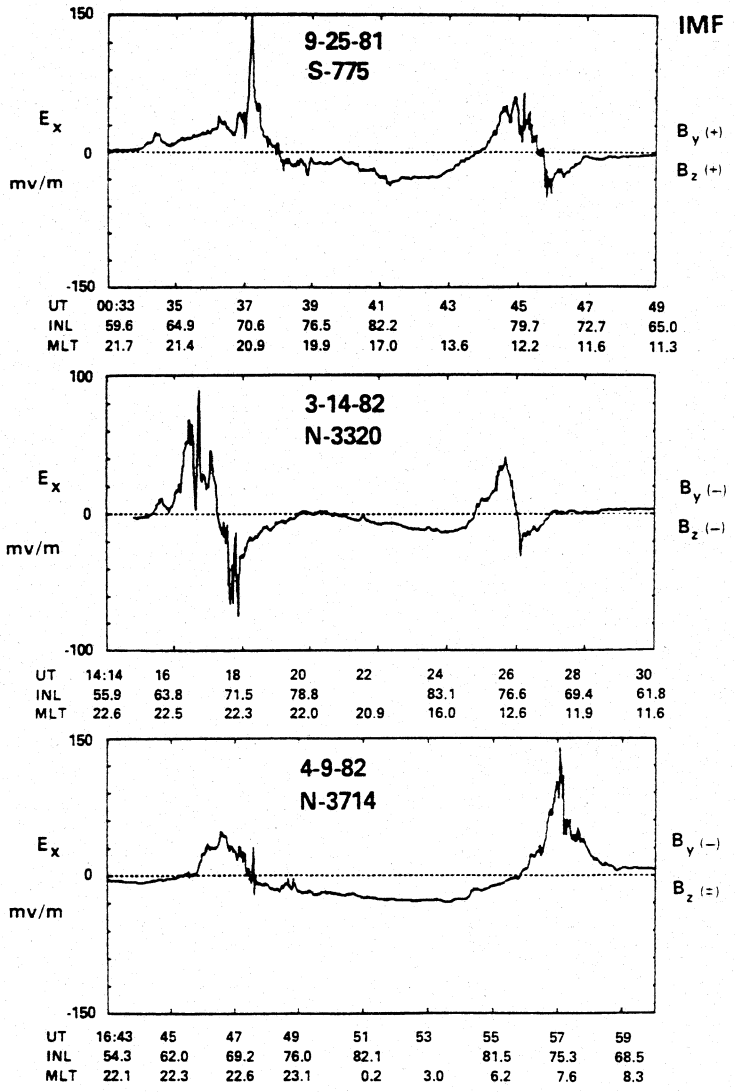
Figure 1.12 Potential patterns as projected onto the northern hemisphere. Equipotential curves are marked with potential in kV. Model patterns vary with activity level and direction (sign) of azimuthal interplanetary magnetic field (Y IMF).



Observations of electric fields from satellites in low polar passes complement the magnetic studies. Examples of potentials and electric fields observed with the Dynamics Explorer satellite are shown in figures 1.12 and 1.13 [Heppner and Maynard, 1987]. Large fields and large field gradients are seen to be approximately colocated with the Region 1 and Region 2 currents in the auroral oval at roughly 60° to 70° invariant (dipole magnetic) latitude. From many passes the statistical picture of Figure 1.12 has been built up, in which the electric potential is presented. The electric field is the (negative) gradient of the potential, so that regions with closely spaced contour lines are those of high electric field. The units on the contour lines are kilovolts (kV) and it is seen that in this case there is an observed electric potential difference of roughly 80 kV across the polar cap. The models presented are for varying orientation of the toroidal (Y) component of interplanetary magnetic field. Here the similarities in patterns, rather than their differences, will be considered. The overall electric field within the polar cap is roughly from dawn to dusk, and 80 kV across the roughly 3000 km polar cap corresponds to an electric field of about 30 mV/m. However, within the auroral oval near dusk and dawn the potential gradient, and thus electric field, reverses. Near dawn the auroral oval electric field is equatorward while near dusk it is poleward. Individual spacecraft passes measured electric fields as shown in Figure 1.13. Reversals of the electric field correspond to reversals in plasma flow direction (see below) and this occurs in a pronounced fashion in the evening sector. This flow reversal had previously been found to occur based on

studies of auroral arc drift and the dashed line shown to indicate it in Figure 1.12 depicts what is referred to as the Harang discontinuity.

Figure 1.13 Electric fields as measured on polar passes of the DE-2 satellite. Large fields and reversals occur in the auroral zone while the polar cap has small fields.



In a global sense, the above electric and magnetic observations are consistent in that ionospheric currents presumably, for the most part, connect the Region 1 and 2 currents to provide current closure. In the morning sector, downward currents are observed at the poleward border of the auroral oval, and the equatorward ionospheric electric field is that required to allow completion of the circuit to upward field-aligned current in the equatorward portion of the auroral oval. The situation is reversed in the evening sector. The more involved question of where the field-aligned currents close far from the Earth is not addressed here. It is simply worth noting that the auroral oval

lies in the vicinity of the interface between closed field lines which can be traced from one magnetic pole to the other on Earth, and so-called 'open' field lines for which this is not possible [Lyons, 1992].

The electric fields are also consistent with what had previously been deduced about ionospheric plasma flow within the polar cap. The arrows on the equipotential lines in Figure 1.12 indicate plasma flow which would be inferred from the electric fields. The contour lines which indicate the equipotentials are also the plasma flow streamlines. This is due to the good conductivity of the upper F region ionosphere measured by the spacecraft at 800 km altitude, meaning that the 'frozen-in' condition applies: $\boldsymbol{E} = -\boldsymbol{v} \times \boldsymbol{B}$, where **E** is the electric field, **v** the plasma flow velocity, and **B** the magnetic induction. Since the electric field is also the (negative) gradient of the potential ϕ, $\boldsymbol{E} = -\nabla\phi$, its component along **v** is $-\nabla\phi \cdot \boldsymbol{v} = -\boldsymbol{v} \times \boldsymbol{B} \cdot \boldsymbol{v} = \boldsymbol{v} \times \boldsymbol{v} \cdot \boldsymbol{B} = 0$. Thus $\nabla\phi \perp v$ and the equipotentials (orthogonal to the potential gradient) are locally parallel to the flow and are thus flow streamlines. The frozen-in condition applies throughout much of the magnetosphere (breaking down only on thin boundaries or small regions such as those above auroral arcs [Lyons, 1992]) and a further consequence is that the electric field is perpendicular to the magnetic induction (since $\boldsymbol{E} \cdot \boldsymbol{B} = -\boldsymbol{v} \times \boldsymbol{B} \cdot \boldsymbol{B} = -\boldsymbol{v} \cdot \boldsymbol{B} \times \boldsymbol{B} = 0$). This implies that magnetic field lines (the loci of connected points whose tangent lines are parallel to the magnetic induction at any point) are also equipotentials. A potential difference observed in the ionosphere will map out along the field lines and correspond to an electric field in the magnetosphere. This in turn indicates that flow must occur there. The corresponding magnetospheric and ionospheric flows are known as 'convection' by analogy to thermally driven flows in fluids (although the magnetospheric flows are not thermally driven).

The ionospheric flow associated with convection, provided that collisions with neutrals do not dominate ion motion (see Chapter 2), features ions and electrons moving at the same velocity [Carlson and Egeland, 1995]. In this case, the plasma velocity and the species velocities (indicated by subscript e for electrons, i for ions) are simply the 'E×B drift velocity':

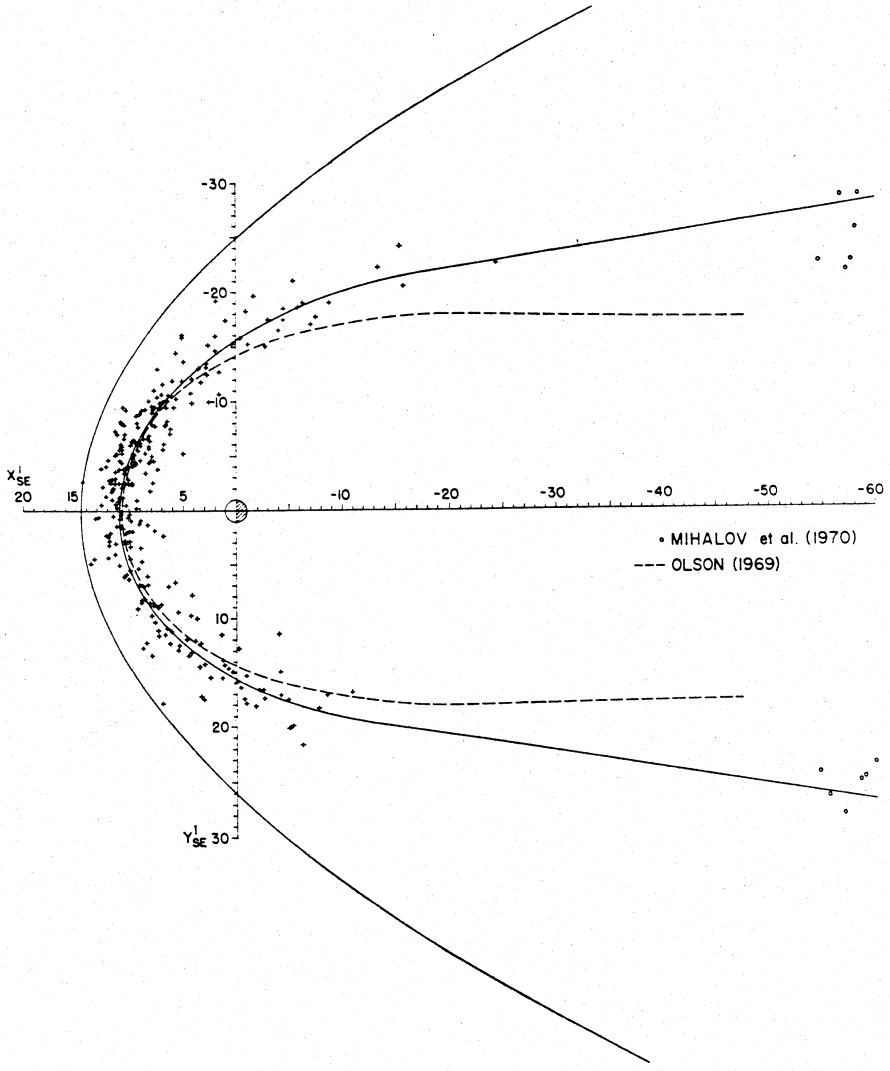
$$\boldsymbol{v} = \boldsymbol{v}_e = \boldsymbol{v}_i = \frac{\boldsymbol{E} \times \boldsymbol{B}}{\boldsymbol{B}^2} \approx \frac{|\boldsymbol{E}|}{|\boldsymbol{B}|} \boldsymbol{e}_\perp,$$

where the approximation at the right applies to horizontal flow in a near-vertical magnetic field and is parallel to the unit vector perpendicular to both. In the northern hemisphere that unit vector is parallel to a 90° counter-clockwise rotation of the **E** field. This approximation, which holds well in the auroral zone, is often used to allow conversion of drift velocities of ionospheric irregularities, which are readily observed with radars, into electric fields. With the known values of magnetic and electric fields in the auroral zone, the speeds of plasma flow along the ionospheric streamlines can be calculated. The behaviour of the main near-dipolar field at ionospheric height is given by equations below, but here it suffices to note that the auroral zone field is nearly

vertical with a strength of about 50,000 nT. With the average 30 mV/m cross polar cap electric field, the associated plasma speed is 600 m/s. In regions where the electric field is stronger than this average the speed can be higher. About 1 km/s could be considered a nominal auroral zone ionospheric flow speed under active conditions. These considerations apply well in the upper F region of the ionosphere, where Carlson and Egeland point out that recombination lifetimes for ions are also long enough that these flows cause net transport of ions. In the E region, due to frequent ion-neutral collisions, the drift approximation breaks down, ions are not transported in this manner, and resistive effects are important. Discussion more appropriate to that region (which is, as mentioned above, where most current flows) is deferred to Chapter 2.

The configuration of the magnetosphere implies connections of one sort or another between its various regions. We have seen that increased solar wind pressure can push back the magnetopause so that the magnetosheath may be detected at synchronous orbit. Nevertheless, it appears that the basic structure is preserved. The existence of such structure, in this case in the form of a persistent boundary, is also deduced from spacecraft moving through the more normal position of the boundary at quieter times as shown in Figure 1.14. The position of the bow shock is also shown in the figure although the positions of the individual bow shock crossings are not shown. The structures seen in these repeated spacecraft passes are entirely consistent with those seen when the boundary was pushed back to synchronous orbit by increased dynamic pressure. When that occurs, the compression of an overlying region has caused deformation of an underlying one, but the overall topology has remained the same. The changes in boundary positions can be well explained by considering fluid dynamics with magnetic effects added. This interaction between clearly defined regions may be considered to be one type of connection between them.

Figure 1.14 Magnetopause positions as projected onto the plane of Earth's orbit as observed by six IMP spacecraft in 1963 to 1968 [Fairfield, 1971]. Solid curves are best fits to magnetopause (inner curve) and bow shock (outer curve). Dashed curve is a theoretical magnetopause position.



Another type of connection between regions is that based upon mapping along field lines. The importance of such mappings may be looked at from two points of view. From a macroscopic point of view, field lines may at most points be considered to be equipotentials for reasons (mainly high parallel conductivity) mentioned above. This has implications for the mapping of electric fields, whose strength between equipotentials is related to their distance from each other and thus is closely linked to the magnetic field geometry. The integral effect of the mapped electric fields is expected to be a cross-tail potential of roughly 80 keV since field lines from opposite points across the polar cap presumably map to opposite points across the tail. As we have seen, the velocity is related to both electric fields and magnetic fields due to the frozen-field condition, and we will return to discussion of velocity fields at the end of

this section. From the complementary microscopic point of view (the guiding centre approximation), relatively low energy particles follow field lines closely. Thus field lines indicate where particles can easily move from one region to another. Since field line positions are influenced by local currents (including those flowing along the field lines themselves), mapping is not as straightforward as it might originally seem, and in fact will be to a large extent avoided in the present work. It is nonetheless important to realize that points in the ionosphere may be connected to more distant regions of the magnetosphere as viewed both from a macroscopic and microscopic point of view. Details of the connections are found in mapping along field lines. The lowest order approximation to the field is that of a dipole. External to the Earth, the dipole field is curl-free and not the result of local currents. The actual mappings are made by its superposition with fields caused by external currents. Near the Earth, the dipole field dominates and mapping along its field lines is similar to mapping along the total field. Further out (starting at about synchronous orbit) the other fields can equal or exceed the dipole field and dipolar mapping should not be considered accurate. On the nightside of the Earth the currents generally cause true mappings from a given surface latitude to arrive at the equatorial plane further out than would a dipolar mapping. On the dayside, due to compression, the opposite is true. Here the considerations in tracing a dipolar magnetic field line are briefly discussed. In the modelling work presented later, currents are assumed to follow dipolar field lines as described in the next paragraph. Ground data are used in this work and the field near the surface is dipolar to a high degree of accuracy. That part of the surface fields which is due to currents flowing along field lines is well approximated by treating them as dipolar.

In quite general fashion, a magnetic field line is, at a given time, the locus of points whose tangent is parallel to the local magnetic field. To lowest order the Earth's internally generated field is that of a dipole: in Cartesian coordinates with the dipole moment M_z along the z axis this has components:

$$B_x = 3xzM_z r^{-5} \quad B_y = 3yzM_z r^{-5} \quad B_z = (3z^2 - r^2)M_z r^{-5} \quad B = \sqrt{3z^2 + r^2}\,M_z r^{-4}$$

where $r = \sqrt{x^2 + y^2 + z^2}$. As noted above, for the terrestrial dipole the value of M_z is 8×10^{15} Tm^3 or 30400 nTR_e^3, the latter value being useful if distances are expressed in Earth radii (R_e) and magnetic induction in nanotesla (nT) as is usual in geophysics. Spherical coordinates are a useful system in which to express the equation of a field line [Parks, 1991, p. 55], which is $\phi = \phi_0; r = r_0 \sin^2\theta$. If the constant of integration r_0 is expressed in R_e it is often denoted by the symbol L and referred to as the MacIlwain L parameter of the field line. This is the radial distance from the centre of the Earth to the field line in the equatorial plane. A field line may be traced from the surface of the Earth, where $r=R_e$, and often it is convenient to use the polar angle of the start point as a parameter: calling this θ_0, one has $r_0 = \dfrac{R_e}{\sin^2\theta_0}$, or $L = \dfrac{1}{\sin^2\theta_0}$. In this case the

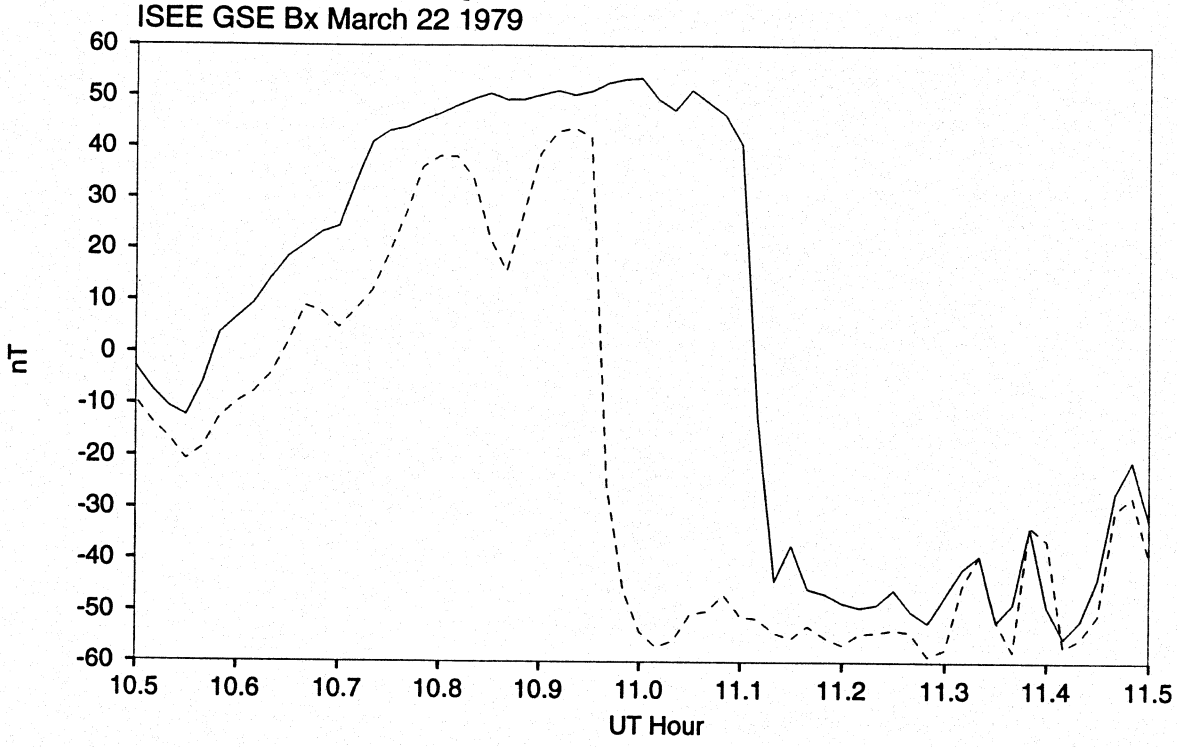
solution for radius as a function of polar angle is $r = \frac{R_e \sin^2\theta}{\sin^2\theta_0}$ or in units of Earth radius $r = \frac{\sin^2\theta}{\sin^2\theta_0}$. Rather than the polar angle (colatitude) θ, it is often more convenient to use latitude as a coordinate, and in degrees the latitude $\lambda=90°-\theta$. Mapping in a dipolar field implies that field lines starting at latitude 45° have L values of 2, those starting at 60° have L values of 4, and those from 75°, L values of about 15. Typical auroral oval field lines should map into the equatorial plane of the magnetosphere at roughly 6 to 10 R_e. Magnetic field lines are distorted by the intensified currents which flow in active times and mapping is a hotly debated topic under these conditions.

The dominance of the internal dipole-like field is no longer complete beyond roughly 10 R_e. On the dayside, as already discussed, the magnetopause is at about this radius, and outside it one is in the magnetosheath or solar wind. On the nightside, the dynamic pressure of the solar wind creates the magnetotail (often referred to simply as the 'tail'). The rapid falloff of the dipole field means that the structure of the tail is dominated by local currents, with the dipole field ultimately forming a boundary condition. Consistent with this boundary condition is the observation that the northern half of the tail is dominated by fields which point Earthward[3] and northward, and that the southern half is dominated by fields pointing away from the Earth and also northward. Beyond 10 R_e downtail, the northward component is small compared to the total field except in the zone of transition between Earthward and anti-Earthward field. The field is referred to as 'tail-like' where this is true. The zone of transition is necessarily (through Ampère's Law) occupied by a current (oriented dawn-to-dusk) across which the field reverses. Since the field is small in this region the zone of current flow is referred to as the 'neutral sheet'. Magnetic pressure considerations dictate that lower field regions have a higher contribution from particle pressure than adjacent regions: the low field region in which the neutral sheet is found is thus called the 'plasma sheet'. The adjacent, higher field regions are referred to as the 'lobes'. Figure 1.15 illustrates the oppositely directed lobe fields through simultaneous observation at twin ISEE spacecraft. The spacecraft are located only 0.6 R_e apart in the direction perpendicular to the location of the inferred neutral sheet, although at respective radial distances from Earth of 15.4 R_e and 13.9 R_e. ISEE-2 goes from north to south relative to the current at 10.95 UT, while ISEE-1 does not cross it until 11.1 UT. During this interval, ISEE-1 sees the Earthward-directed field of the north lobe while ISEE-2 detects the oppositely directed field of the south lobe. The small spacecraft separation allows inference of 'plasma sheet thinning' during the CDAW-6A substorm, March 22, 1979. In this case [Fairfield, 1984] it was inferred that the plasma sheet had become only 300 km thick (strictly speaking, using only magnetic

[3] Note that the main field as observed at the surface in the northern hemisphere is also 'Earthward', that is to say that its vertical component points radially inward.

measurements, this can only be inferred as the thickness of the current sheet embedded within the plasma sheet). During the short time interval shown in the figure, motion of the spacecraft is relatively small. In most cases, including this one, the crossing is not that of the neutral (current) sheet by the spacecraft, but a traversal over the spacecraft location by a moving current sheet. The term 'neutral sheet crossing' is often used when reversals of the tail field, as seen in this figure, are observed, but should not usually be taken to imply spacecraft motion.

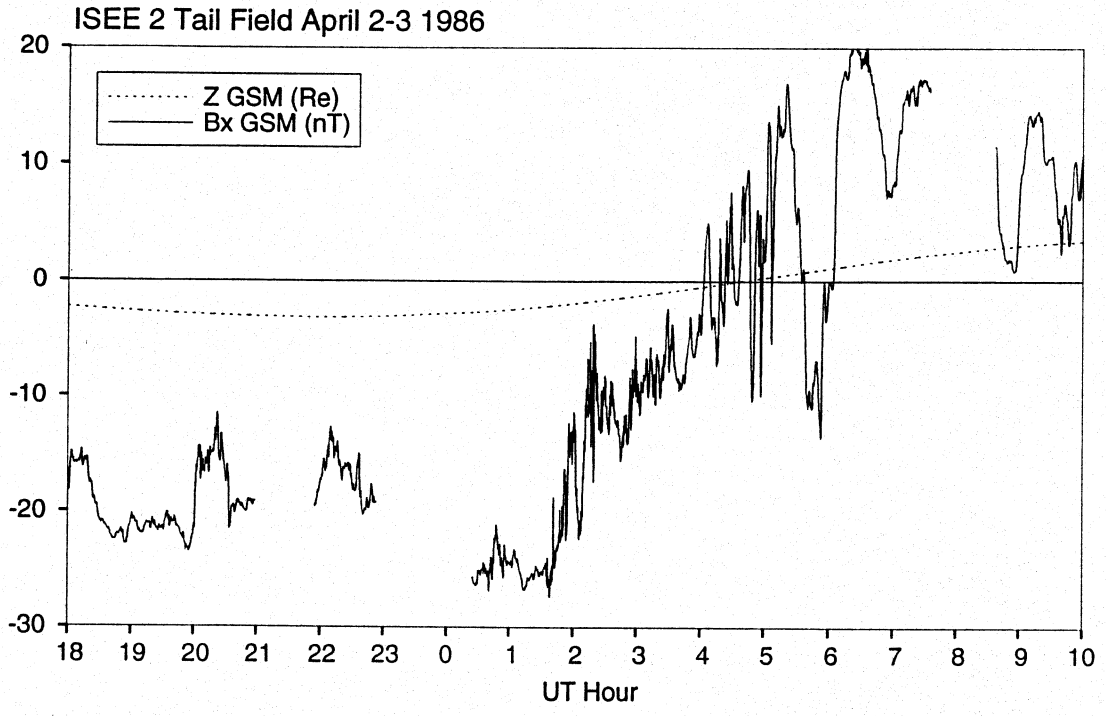
Figure 1.15 Simultaneous observations of the two lobes from ISEE-1 (solid) and ISEE-2 (dashed). The GSE B_x component is along the Earth-Sun line and accounts for most of the field magnitude except when near zero.



A traversal through a less thinned current sheet is illustrated in Figure 1.16, in which data from the single spacecraft ISEE 2 are shown as its GSM Z position changes with time throughout an interval of sixteen hours. Initially, in the south lobe, the spacecraft detects a field directed away from Earth. This field is about -10 to -20 nT, smaller than in the CDAW 6 case above, partly because the spacecraft is further from Earth (at about 20 R_E) and the tail field is known to decrease with distance, and likely partly due to a lower strength of the current sheet. With much variation, the X component of the field reverses sign approximately as the spacecraft passes through the zero position of GSM Z, becoming about +20 nT on average after the crossing. At least part of the 'noise' in the plot is due to motion of the current sheet. This is best seen when the field briefly reverses for about one half hour starting at about 5 UT. The most likely explanation is that the mobile current sheet moved to positive Z GSM values. Thus, with respect to the current sheet, ISEE 2 was to the south and detected earthward

field. This example also serves to illustrate the nature of spacecraft data taken at a single point. Even if structures (such as the current sheet) are present at all times, they may move and it can be difficult to disentangle their motion from that of a spacecraft. The reader may consult Section 3.e for details of the GSM coordinate system which provides a useful reference plane (Z=0) in this example.

Figure 1.16 ISEE 2 pass through the current sheet on April 2-3 1986. Z GSM coordinate very roughly indicates distance from the nominal current sheet position. B_x GSM is basically Earthward when negative. Data gaps are present in these data.



Since the energy density and pressure in a magnetic field are equivalent and equal to $\frac{B^2}{2\mu_0}$, energy is stored in the lobe magnetic field. Since the field is antisymmetric about the neutral sheet and zero at its centre, $B_X = \frac{\mu_0}{2} I_y$ in an infinite current sheet approximation, where I_y is the linear current density (A/m in SI units). An increased neutral sheet current thus leads to larger lobe fields and increased energy storage. This capacity for energy storage and the reversed field configuration, which is thought to be conducive to the reconnection process of energy conversion, have led to the concept of 'unloading' of tail energy in a bursty fashion. Much as discussed previously in the context of the upper ionosphere, the 'frozen-in' property of plasmas implies motion under the conditions in the tail. Thus energy may be stored in the form of plasma kinetic energy in the tail as well as directly in the magnetic field.

The final aspect of configuration to be dealt with here is the distribution of particle populations of the magnetosphere. As mentioned in the section on the ionosphere, it can be a source as well as a sink of particles. Characteristic species from that region (for example oxygen) can be injected into the magnetosphere to serve as tracers, and affect its structure and dynamics [Hughes, 1995]. Here, not much attention will be paid to the varying composition in the magnetosphere. The primary distinctions made will be of temperature and density of particles in various regions. It has already been established (recall Figure 1.3) that the total pressure in the magnetotail is much greater than the gas plus magnetic pressure in the solar wind, and can only be balanced by components of the solar wind dynamic pressure on its flanks. We have also seen that the north and south tail lobes, separated by a current sheet and associated plasma sheet, have oppositely directed fields. On the dayside, when the magnetopause was pushed earthward, another layer was noted, the magnetosheath. In fact, the magnetosheath plasma flows around the entire magnetosphere and has complex interactions with it [Hughes, 1995]. Inside the magnetopause there are boundary layers containing mixtures of magnetosheath and magnetospheric plasmas. The low latitude boundary layer (LLBL) is the region between magnetopause and plasma sheet. With the further subdivision of the plasma sheet into a boundary layer (between it and the lobe - this could be quite narrow) and a central plasma sheet, Hughes [1995] supplies typical parameters of these regions for the near tail (about 30 R_e out), which have been augmented by data from Hundhausen [1995] and Wolf [1995]) to form Table 1.1.

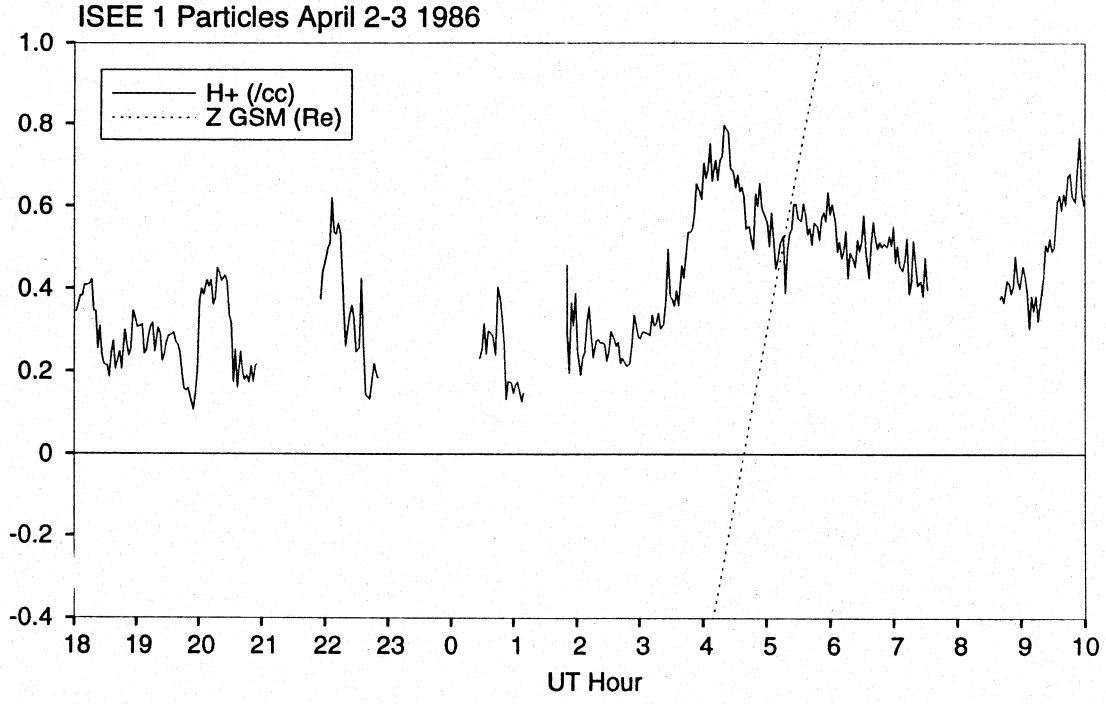
Table 1.1 Typical Near-Tail Plasma and Field Parameters

Region	n (m^{-3})	T_i (eV)	T_e (eV)	B (nT)	β
CPS	3×10^5	4200	600	10	6
PSBL	10^5	1000	150	20	0.1
Lobe	10^4	300	50	20	0.003
Magnetosheath	8×10^6	150	25	15	2.5
Solar Wind	7×10^6	10	12	7	2
Plasmasphere	10^9	1	1	1000	0.001

The last column indicates the plasma beta (β), which is the ratio of thermal to magnetic pressure. The relatively dense but cold plasma of the magnetosheath (density values cited here, in the tail, are considerably less than in the subsolar region) and the solar wind, and the less dense but hot plasma of the central plasma sheet (CPS), have high β and are dominated by thermal pressure, while the lobe is dominated by magnetic pressure and contains very few particles. The plasma sheet boundary layer (PSBL) is the transitional region between CPS and lobe and has intermediate characteristics. A distinguishing feature is sustained field-aligned ion and electron flows both earthward and tailward, indicating that acceleration processes are active in or near this boundary layer [Hughes, 1995]. Once more examining real data with all of its nonideal aspects, we consider Figure 1.17 which shows ISEE 2 particle data for the

same period as the magnetic data of the previous figure. As the magnetic data suggested, the plasma sheet does not appear to be particularly thin since there appears to be no trace of the very low lobe densities in these data, even when the spacecraft is at ±2 R_E. The density appears to be higher when the spacecraft is traversing the neutral sheet (dashed trace for Z position crosses the axis), and certainly the density values are within the range indicated in the table for the PSBL and CPS.

Figure 1.17 Particle data as ISEE 2 traverses the neutral sheet on April 2-3 1986. Proton density (in units of 10^6 m^{-3}) increases near nominal neutral sheet (Z GSM=0).



As might be expected, particle behaviour in regions dominated by the dipolar field is somewhat different. Rather than a tail-like structure enforced by the streaming by of the solar wind, a cylindrical symmetry is dominant in the inner regions, imposed by the symmetry of the dipole field about the magnetic pole. In the innermost regions, called the plasmasphere, plasma is dense (up to about 10^{10} ions/m^3) and corotates with the Earth. Corotation of the bulk plasma may be regarded as a manifestation of the 'frozen-in condition' in the region where there is very low β so that the strong intrinsic field dominates. More correctly, in this region there is a corotation electric field established through charges (to first order) present at ionospheric level due to the rotation of that conductor enforced by atmospheric drag. Studies of dispersive electromagnetic waves ("whistlers") and *in situ* measurements from satellites reveal a usually abrupt plasmaspheric outer boundary referred to as the plasmapause [Wolf, 1995]. Here densities decrease by one or more orders of magnitude to those (10^6/m^3 or less) typical of the tail (CPS). In addition it has been observed that this boundary

varies in position, being at lower L values for high levels of magnetic activity. At the highest levels it is at an L value of about 3.5 while in quiet times it retreats to about L=6. The plasmapause thus generally maps to subauroral latitudes.

Particles in the magnetosphere may be characterized by single particle equations, which for 'high energies' may give results differing from those of the bulk plasma. The guiding centre approximation of Alfvén is very suitable for examining the motions of such particles and a large and clear literature exists, of which two useful examples are Lyons and Williams [1984] and Wolf [1995]. Here only the general results of the theory, as needed to understand qualitative and some quantitative aspects of particle motion, are presented. The guiding centre is a fictitious moving point associated with a particle. Particle motion may be viewed as consisting of motion of the guiding centre and motion of the particle around the guiding centre in a circular orbit. Inherent in use of the guiding centre approximation is that the particle has free motion in the presence of external force fields which change on a scale much larger than its orbit about the guiding centre (both spatially and temporally). This is observed to be the case almost everywhere in the magnetosphere as generally the orbits are much smaller than any other length scales and the orbital frequency is one of the highest frequencies in the system. Frequent collisions (implying brief application of strongly varying forces) among particles would prevent the guiding centre approximation from remaining valid, but magnetospheric plasmas, right down to the upper ionosphere, are essentially collisionless. In this case, the effects of a force field **F** on a particle are to impart a velocity parallel to the field lines and to distort the orbits from circular: this latter causes motion of the guiding centre relative to the magnetic field. In a general sense, then, the motion of the guiding centre will take place at a drift velocity perpendicular to the local magnetic field [Parks, 1991, Section 4] which is $\boldsymbol{W_D} = \frac{1}{q}\frac{\boldsymbol{F}\times\boldsymbol{B}}{B^2}$, where q is the particle charge. A special case of external field is an electric field, in which the force on a particle is **F** = q **E**, and the charge cancels out in the drift equation, leaving the 'E×B' drift already alluded to in discussion of the ionosphere in the preceding section. Corotation may also be understood by considering all of the components of the plasma to E×B drift due to the induction electric field imposed by the ionosphere's behaviour as a unipolar inductor. Corotation applies to low energy constituents (the bulk of the plasma in the plasmasphere in practice) and we must now explain what is meant by low energy as opposed to 'high energy', the term with which this paragraph began. It has been noted that the cross polar cap potential is roughly 80 kV and that a potential of this order could be expected to map out into the magnetosphere. Particles having an appreciable fraction of the energy associated with acceleration through this potential could be thought of as 'hot' or 'high energy' since they would not change their energy appreciably in moving a small distance in the magnetosphere, while those of much lower energy would be expected to be more affected by electric fields (the observed cross-polar cap field, the corotation field, and local fields). The plasmaspheric particles, with an average energy of 1 eV, are most definitely 'low energy' and effectively react to the electric field through 'E×B drift'. Higher energy

particles, although perhaps mixed with the cold particles, have significantly different dynamics. One should recall that these populations are not coupled by collisions and in many circumstances can be considered to be noninteracting. Exceptions to this statement can occur and can be important in generating magnetic pulsations [Samson, 1991]. We proceed now to consider more terms of the single particle drift equation, which are needed to understand high energy particles which will later be used in studying substorm onsets .

A particle (mass m) whose motion is controlled by the electric and magnetic fields is subject to the Lorentz force $\mathbf{F}_L$ and obeys Newton's second law, so that

$$m\frac{d\boldsymbol{v}}{dt} = \boldsymbol{F_L} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}).$$

The orbits referred to above in the guiding centre approximation are part of the solution of this equation in a uniform magnetic induction field. It is readily shown that these orbits occur at the cyclotron frequency $\Omega_c = \frac{qB}{m}$, which is independent of particle energy, and with the Larmor radius (or gyroradius) $\rho_c = \frac{v_\perp}{\Omega_c} = \frac{mv_\perp}{qB}$, which increases with particle (perpendicular) energy ($v_\perp$ is the particle velocity along the orbit, which is in the plane perpendicular to the magnetic field). For all but the highest energy particles, and possibly near certain boundaries, the gyroradius is much smaller than other lengths in the magnetosphere. The electron cyclotron frequency is generally higher than other frequencies encountered, and due to inverse mass dependence is 1837 times higher than the corresponding proton (often loosely called simply 'ion') cyclotron frequency. Another aspect of the solution is that if there is a component of the electric field along the magnetic field, there is acceleration parallel to the field which is proportional to it. Empirically, the field lines are in most places very close to equipotentials, so that such an electric field is small. As seen above, the perpendicular component of the electric field causes an E×B drift, which may be considered as a fundamental type of particle motion.

A second type of motion characterizes particles in a nonuniform, and in this case, dipolar, field. Consider a particle in the equatorial plane in such a field and without any parallel component of the electric field, but having a possibly non-zero velocity ($v_\parallel$) parallel to the field. The particle's overall velocity vector will, at such a time, lie at an angle α to the magnetic field, where $\alpha = \tan^{-1}\left(\frac{v_\perp}{v_\parallel}\right)$. This angle is the 'pitch angle'. If the parallel velocity component is in fact non-zero, the particle will move away from the equatorial plane. In so doing, it will enter a region of stronger field (recall that $B = \sqrt{3z^2 + r^2}\, M_z r^{-4}$, which has a minimum in the plane z=0). In this motion, where the change in B will be taken to be slow over one gyroperiod, an adiabatic invariant applies [Wolf, 1995]. With a perpendicular coordinate's momentum-displacement integral conserved, the quantity $\frac{1}{2}mv_\perp^2 \frac{2\pi m}{qB} = \frac{2\pi m}{q}\mu$ is conserved, which means that,

for constant charge/mass ratio, the so-called first adiabatic invariant $\mu = \frac{mv_\perp^2}{2B}$ also is. This quantity is equal to the perpendicular kinetic energy per unit of magnetic induction, and is the magnetic moment of the particle. In the small displacement considered here, the increase in B while μ is conserved means that the perpendicular kinetic energy must increase. Since it is easily shown that the magnetic field can do no work on a particle, its total kinetic energy does not change, so that any increase in perpendicular kinetic energy must be matched by a decrease in parallel kinetic energy. As the particle moves, due to its initial velocity, into regions of stronger field, it loses parallel energy (and thus speed) until it has zero parallel speed. At this point it reverses the process. The overall periodic motion between points of zero parallel speed is known as 'mirroring'. Where a particle mirrors depends on its equatorial pitch angle. If this is too large, in the case of the Earth, entry into the atmosphere may occur. Processes exist which can cause pitch angle changes ('pitch-angle scattering') and a population of mirroring particles may see its members enter the atmosphere if such processes are effective.

It is often stated that 'particles follow field lines' when the guiding centre approximation is valid. This is true of low energy particles, which in the general case make many 'bounces' and perhaps perform other motions on a much longer timescale. It should be borne in mind that even in the relatively simple geometry of the dipole field, particle trajectories are non-integrable. However, the problem has been solved repeatedly by numerical methods for even complex fields, and the following of field lines, with consideration of drifts perpendicular to them which are discussed below, is an established property. It can be shown for a dipole field that the general (any energy) case is for particles to orbit traversing the surface formed by rotation of the dipole field lines about the axis [Chamberlain, 1964, section 1.6]. In fields close to dipolar, closely analogous 'drift shells' may be defined using adiabatic invariants. For higher energy particles, in addition to the simple bounce motion, drifts may be important. The major drifts are the gradient and curvature drifts. The former is important for particles whose perpendicular energy leads to a large enough gyroradius that regions of varying induction field are sampled and the orbit deformed slightly. The latter is important for particles having large parallel energy, so that the action of following a field line implies a net inward centripetal force which is not negligible. By considering the orbital deformation from a general force **F**, the aforementioned drift velocity $\boldsymbol{W_D} = \frac{1}{q}\frac{\boldsymbol{F}\times\boldsymbol{B}}{B^2}$ results, even if we do not consider in detail the origin of the force but merely note that by following a drift shell the particle must have been subjected to a centripetal force (whose magnitude can be calculated based on the acceleration). The general drift formula in the absence of external non-electromagnetic forces thus becomes [Wolf, 1995] a sum of E×B, gradient, and curvature drifts, respectively:

$$\boldsymbol{W_D} = \frac{\boldsymbol{E}\times\boldsymbol{B}}{B^2} + \frac{W_\perp \boldsymbol{B}\times\nabla B}{qB^3} + \frac{2W_\parallel \hat{\boldsymbol{r}}_c \times \boldsymbol{B}}{qR_c B^2}.$$

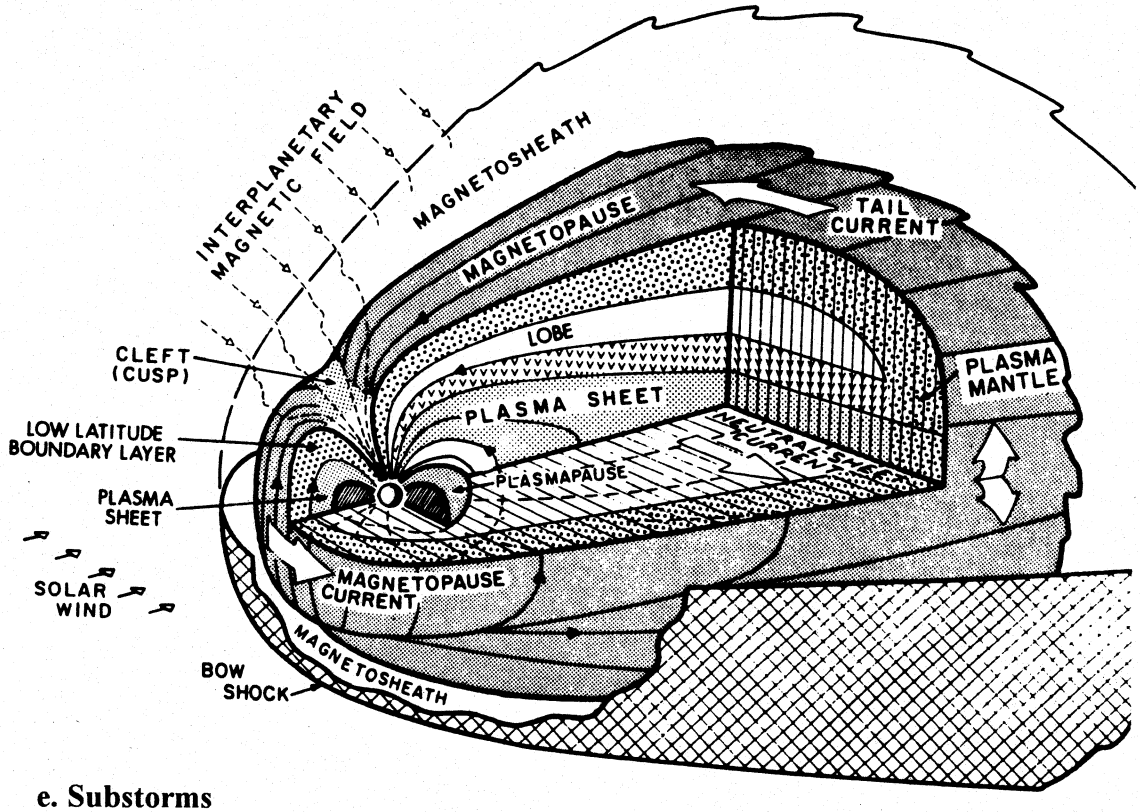
In this equation, $W_\perp$ and $W_\parallel$ are the perpendicular and parallel kinetic energies, respectively, and $\hat{r}_c$ the unit vector direction and R_c the magnitude of the radius of curvature. In a dipole field the field falls off outward while the radius of curvature is inward, so that the gradient and curvature drifts act in the same direction. This direction depends on the sign of the particles: electrons drift eastward while ions drift westward. For high energy particles, the second two terms are most important: their drifts are energy dispersive and capable of causing currents since opposite charges drift in opposite directions. During active times the ring current due to such drifts of high energy particles relatively near the Earth can contribute non-negligibly to the magnetic field observed at the surface.

As mentioned above, if the mirror points are low enough, bouncing particles may collide with particles in the atmosphere and not be able to return. This is known as particle precipitation and leads to excitation of atmospheric atoms and molecules, which upon recombination or deexcitation release radiation. The auroras observed at many wavelengths throughout the electromagnetic spectrum are one result of the precipitation. In a diagnostic sense, auroral emission may be used to locate areas of precipitation. Electric fields above the atmosphere may also remove atmospheric constituents (in a charged state), so that a process opposite to precipitation also occurs. Such upwelling would not usually be accompanied by emission but could carry net current. Study of upwelling ion beams with multiple spacecraft have established that under quiet conditions there is an auroral acceleration region which has a several kV parallel potential drop between about 1400 km and 2 R_e altitude [Reiff *et al.*, 1993]. In certain active regions the acceleration appeared to be at higher altitude, possibly in the magnetotail.

The low-altitude acceleration regions will not be discussed further here but lead naturally to the concept of acceleration in the magnetotail, which appears to have both continuous and impulsive components. The presence of magnetic and electric fields and plasma implies that there must be plasma drifts in the tail. Thus there is continuous motion. As discussed near the beginning of this section, this is also the case in the polar ionosphere and the flows are referred to as 'convection'. It stands to reason that there must be some connection between the ionospheric flows and those in the magnetotail: however this connection falls into the domain of mapping and here will not be considered to be known. Nevertheless, the structure of the tail and its flows can be consistently laid out.

This section has attempted to portray the structure or quasi-static configuration of near-Earth space with a strong observational basis. In closing, a pictorial synthesis of results is shown as Figure 1.18. The drifts of high energy particles alluded to above create a current which is not shown in the diagram: this is the ring current which is generally inside the plasmapause. The high energy particles are often referred to as the 'van Allen' radiation belts.

Figure 1.18 Cross-sectional pictorial synthesis of Earth's magnetosphere.



e. Substorms

To this point, the focus has been on static characteristics of the coupled solar wind-magnetosphere system. Although it has been made abundantly clear that variations occur (for example in dynamic pressure), the overall picture which has emerged is one of regions of space with boundaries which may change but which nevertheless remain identifiable. By contrast, the substorm process appears to necessarily involve changes in the magnetospheric configuration, but it is not clear what these changes are, in which regions they occur, or to what degree the regions remain distinct.

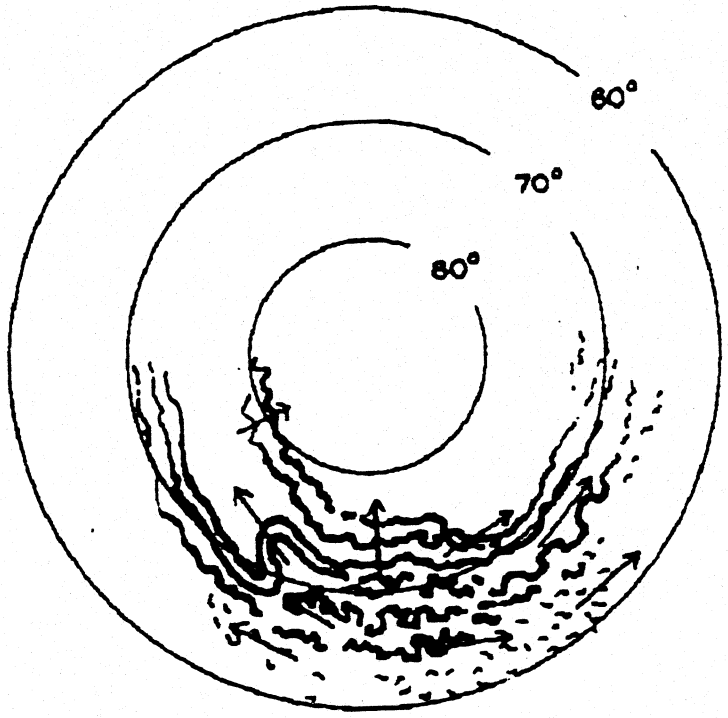
Substorms are characterized by certain observational manifestations, which will be discussed here, some of which have relatively direct known connections to physical processes. The comprehensive physical picture of substorms, however, remains unclear, as the physical processes involved are complex and multifaceted, and the data available often ambiguous even when voluminous.

The term 'substorm' came into use around 1960, when Chapman and Akasofu, studying magnetic storms, found that the polar disturbance field varied on a shorter timescale [Stern, 1991]. In fact Birkeland in the period 1899 to 1903 had already identified what he called 'elementary polar magnetic storms', and made the association of increases in auroral luminosity with increases in magnetic disturbance locally in the

auroral zone. The occurrence of substorms on the roughly half hour timescale and of storms on the half day and longer timescale is now accepted and the relationship between the two phenomena is still under investigation. Substorms can occur outside of storm periods but may be involved (in most cases) in producing particles causing storms. Storms are not discussed in detail here although their effect through the ring current will be seen below (Section 2.d) as manifested in the D_{st} index.

The various aspects of substorms led at one time to a set of terms describing their manifestations. It is now clear that all are related and thus the simple term substorm is in general use to describe the phenomenon. The systematic study of the visual auroral observational aspects of substorms on a near-global scale was first undertaken in the late 1950's [Akasofu, 1964]. Beginning from a quiet phase (it was later realized that there is usually characteristic activity during this phase also), in which homogenous arcs dominate the auroral zone, there is a sudden brightening of an arc and subsequent poleward motion. This 'expansive phase' leads to the formation of a region of intensified emission, usually centred on midnight, called the 'auroral bulge'. Figure 1.19 shows the later stages of formation of the bulge as inferred through ground-based imaging.

Figure 1.19 Stage III of the auroral substorm expansion phase according to Akasofu [1964]. Active auroras characterize the midnight sector (bottom).



From instantaneous satellite imaging, it has been verified that the bulge is indeed roughly of the shape deduced by Akasofu (see particularly Chapter 7 below). A distinctive loop form in the evening sector is referred to as a 'surge' and subsequent

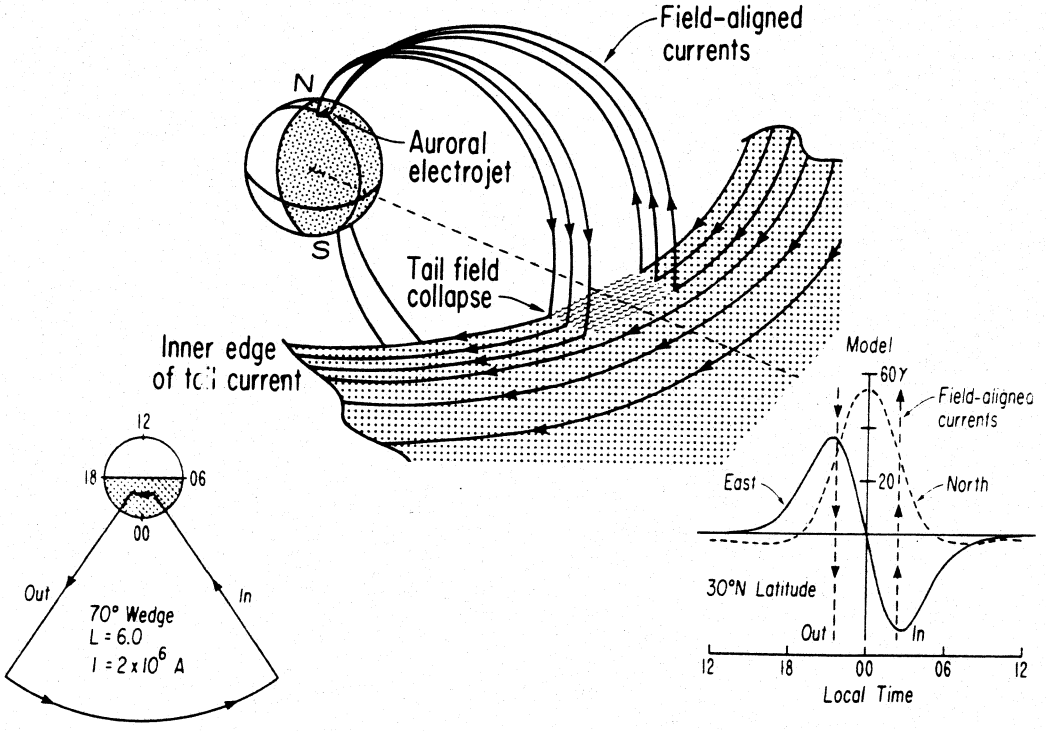
more westerly appearance of surges on all-sky camera images at different stations led to the conclusion that the surge moved westward, hence the term 'westward travelling surge'. It has been subsequently shown that the westward motion is not necessarily smooth and that there may be more than one surge present at a time (these aspects are illustrated in Chapter 6 below). Seen in this figure but not emphasized by Akasofu is an evening sector band of enhanced luminosity at high latitude. A finding in the present study is that activity having substorm-like character can occur along such a high latitude band in the evening sector. Akasofu found that the maximum of the expansive phase was followed by the equatorward retreat of auroral arcs, with westward travelling surges degenerating into loops (including groups of loops according to Akasofu) and with eastward moving features known as omega bands characteristically found in the morning sector. This second phase of auroral activity was called the 'recovery phase'.

The expansive phase near midnight has as a very characteristic auroral zone signal a large (relative to other types of magnetic variation although small relative to the total field) decrease in the northward component of magnetic field. These signatures are often referred to as bays, and along with an analogous signature in midlatitudes, will be discussed below. Another signature of expansive phase onset is pulsational activity, particularly the Pi 2 damped wave trains in the 6.7 to 25 mHz band [Rostoker, 1979], observed in midlatitudes and increasing in strength near the auroral zone electrojet. Within the auroral zone, the northward magnetic perturbations of 'negative H bays' may show a very rapid change of up to 1000-2000 nT in a timespan of a few minutes. The Z (vertical) component can also show large changes but these are more complex as they depend critically on the latitude of the observer relative to ionospheric currents which are inferred to be enhanced in the ionosphere during a substorm. The east-west component near the auroral zone is also perturbed near surges in such a way as to suggest north-south current flow to be associated with them. Both ground and satellite data suggested that during a substorm there was formation in the near-midnight auroral zone of a region of enhanced ionospheric current, with field-aligned currents connecting it to somewhere in the magnetotail. Not only current closure, but also detection of the magnetic field of these field-aligned currents from the Earth's surface, supported this view [Clauer and McPherron, 1974]. That such a three-dimensional current configuration could exist was by that time tenable due to the discovery (in confirmation of Birkeland's hypotheses) of field-aligned currents, which had occurred by 1970 [Dessler, 1983]. This configuration, apparently active during the expansive phase, is called the 'substorm current wedge' and Figure 1.20 shows the canonical view of it.

Some auroral zone geomagnetic bay variations are known to be associated primarily with the near-Earth substorm current wedge. Similar perturbations at midlatitudes may also be explained by the wedge. Studies in the 1960's indicated the primacy of the westward electrojet in the midnight sector during substorms and equivalent current systems of such an electrojet could explain them [Rostoker, 1966]. In the D (eastward) component,

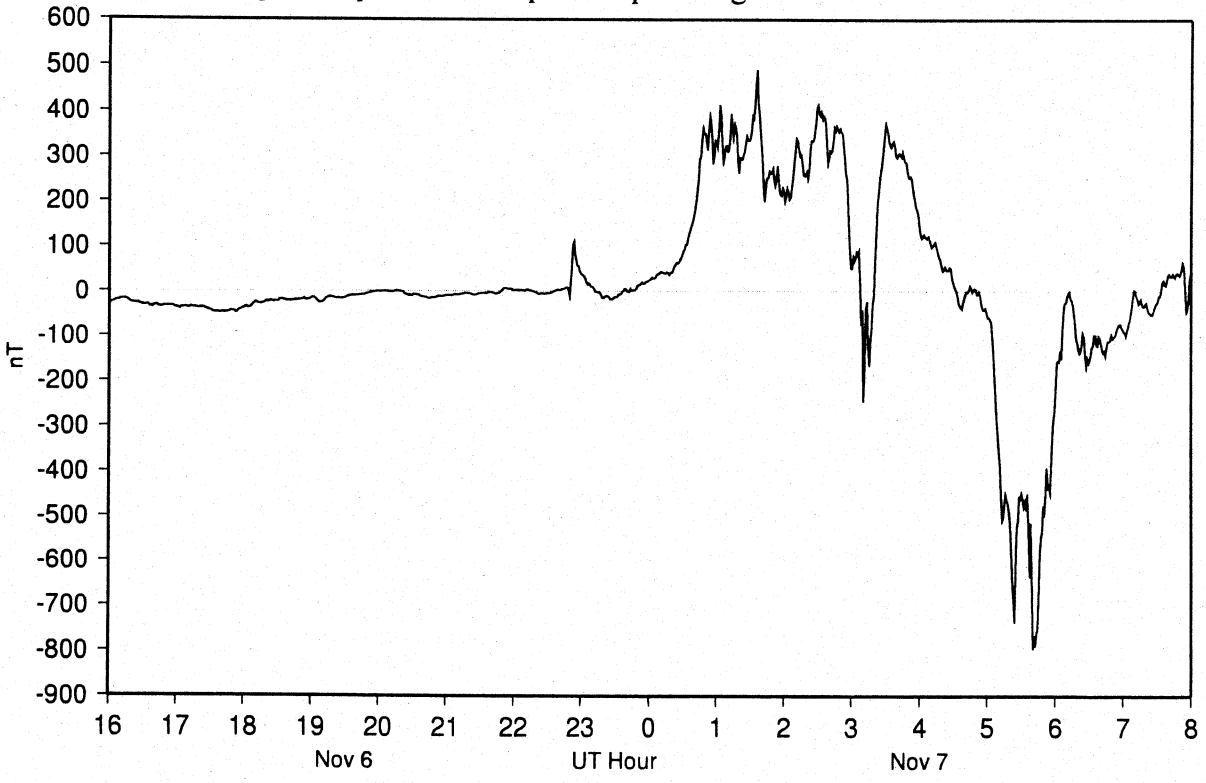
positive bay occurrence peaked strongly about 3 hours before midnight and that of negative bays about 2 hours after. Viewed with the modern knowledge of the importance of field-aligned currents to the mid-latitude D component, these statistics are entirely in conformity with the Clauer and McPherron view as presented in Figure 1.20. These midlatitude perturbations, appropriately weighted, can be used with a magnetic modelling routine to constrain the position of the ionospheric currents further poleward. Rostoker's statistical data also allows deduction of the rough size of the substorm current wedge as being 5 hours of local time or 75°, comparable to that given as typical by Clauer and McPherron [1974].

Figure 1.20 Canonical view of the substorm current wedge as deduced by Clauer and McPherron [1974]. Midlatitude perturbations suggest downward current east of midnight and upward current west of midnight.



The modern 'canonical view' of substorm phenomenology adds a third phase to the two noted by Akasofu. This is the 'growth phase' first clearly identified by McPherron [1970]. The initial motivation for the identification of this phase, preceding the expansive phase, was the presence of disturbances in ground magnetograms. A good example is seen in Figure 1.21, which shows the perturbations to the northward component during an intense event, as recorded at Meanook, 140 km north of Edmonton. An SI at 2300 UT on November 6, 1997 is followed by positive signatures in the early hours of the following day, while the station was in the evening sector.

Figure 1.21 Northward magnetic perturbations at Meanook on November 6-7 (UT) 1997. An SI, growth phase, and expansive phase signatures are seen.



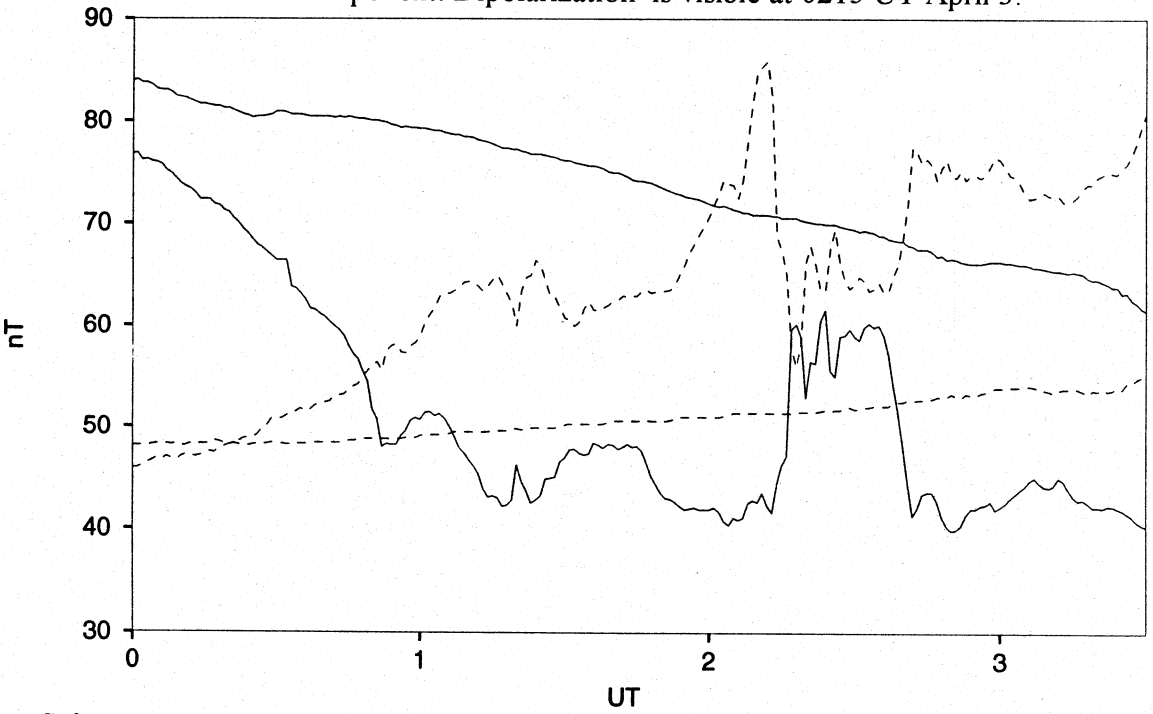
Expansive phase signatures of rapid decline in the northward component (bays) are seen at about 0300 and 0500. These primarily show the effects of the near-Earth electric currents of a near-midnight substorm current wedge similar to that in Figure 1.20. The positive perturbations cannot originate from this system and indicate a growth phase current system which was present in the evening sector. The early studies showed that in addition to ground signatures, distinct magnetospheric signatures were found to occur during the period before the expansive phase and these are now regarded as characteristic of growth phase (see below). It is interesting to note that bursts of Pi 2 band pulsations are taken as signals of the expansive phase, but McPherron [1970] noted that they are also seen during the growth phase. At that time it appeared that the growth phase signatures were generally of the same sign as those of the expansion phase and that a separate current system was involved. It is now generally thought that the substorm current wedge, subsequently proposed by McPherron as noted above, is in fact a separate current system, and that the growth phase signatures primarily arise from the 'driven' system [Akasofu, 1983]. The driven system is a pattern of currents on the global scale. Initially suspected in the 1930's, it has historically been referred to as the S_D system, DP or DS. In the view of Chapman the global effects seen were interpreted as ionospheric currents everywhere. The present three-dimensional view is that the effects in the auroral zone are primarily those of ionospheric Hall currents, while those at low latitudes are generally those of field-aligned currents. The basic configuration of the driven system is one of westward electrojets in the morning

sector and eastward electrojets in the evening sector, consistent with what the perturbations in Figure 1.21 suggest. To create the global driven system, net field-aligned current input occurs near noon and return currents to the magnetosphere are near midnight. These global scale systems may also have Region 1/2 currents associated with them, but the perturbations they cause on the ground are generally subtle. Here, it is shown that they can be detected and modelled, in the polar cap primarily. A discussion of the driven system in terms of relation to magnetospheric convection and electric fields is given by Lyons and Williams [1984, pp. 60-63], in which its relation to active conditions is emphasized (as this means it cannot be used to study general quiet-time convection). The driven system's net convection electrojets have also been referred to as the DP2 system (see for example [Kamide, 1988, pp. 671-672]) while the substorm electrojet is referred to as DP1. The term 'driven system' implies a physical mechanism and similarly the diversion of field-aligned current into the ionosphere in the current wedge model is referred to as an 'unloading' process. The relation of the driven and unloading systems is not yet clearly established and is a subject treated later in this work. A brief discussion of satellite observations of substorm phenomenology will now be given.

In the midnight sector at geosynchronous altitude (relatively near Earth in terms of position downtail) the magnetic field may be characterized as transitional between dipole-like and tail-like. The dipole-like internally-generated field suffers a 300 fold decrease in magnitude in going from the surface to 6.6 R_e in the equatorial plane, resulting in about 100 nT total field there, most of which is directed northward under quiet conditions. Tail fields are variable but can at times attain almost this value, but mostly in the radial component. Figure 1.22 shows the field detected at the GOES 5 spacecraft under quiet and disturbed conditions. The quiet day curves correspond very closely to those found for synchronous orbit at this time of year by Skone [1994]. The spacecraft is in the northern magnetic hemisphere, and thus the earthward component is positive, that is to say, the overall field direction is toward the Earth. On the disturbed day, the earthward component exceeds the quiet day value at all times after 0030 UT. From 0215 UT until 0245 UT there is a notable decrease in the earthward component, and a corresponding increase in the northward component, with both approaching, but not attaining, quiet day values. An increase in the earthward component corresponds to a more 'tail-like' field and departure from dipolar geometry. The decrease in earthward component and rise in northward component indicate a return to more normal dipolar conditions at synchronous orbit. This process of 'dipolarization' in the near-Earth tail occurs at the same time as many indicators of substorm onset, notably major ground perturbations in the late evening sector below the satellite (see Chapter 6). This example is consistent with the observations of Nagai [1982] using earlier GOES-series satellites in that substorm onset corresponds to dipolarization. Another well-studied example is the case of GOES 3, located at 0145 LT (i.e. near midnight) during the CDAW 6A substorm onset [McPherron and Manka, 1974]. That case is more dramatic than that shown here in that the earthward component decreased its magnitude from 100 nT to about 30 nT in only a few minutes at onset time. In the event presented here, the dipolarization is temporary, whereas the

events studied by Nagai were selected to be simpler and to show a long-lasting dipolarization (as was the case for CDAW 6A also). The data here support the conclusion of Nagai: that there is a current flowing in the equatorial plane near synchronous orbit which initially produces a more tail-like, distorted configuration, and that that current is disrupted by substorm onset. The large perturbations observed on the ground during this disruption have led to the view that the current is diverted to the ionosphere as suggested by the canonical current wedge as pictured in Figure 1.20.

Figure 1.22 Northward and earthward magnetic components at GOES 5 on April 3 and 4 (smoother curves) 1986. Solid lines indicate the northward component, dashed lines the earthward component. Dipolarization is visible at 0215 UT April 3.



Substorms are also observed to accelerate particles within the magnetosphere. Phenomenologically, such acceleration is detected through flux increases observed at or near the beginning of the expansive phase. While some flux changes are due to changing magnetic fields and their influence on particle motions, overall enhancements of the flux levels above previous levels also occur. Observations of drifting populations and ion beams in regions near the plasma sheet boundary layer indicate that acceleration actually occurs and appears to be linked to induction from rapidly changing magnetic fields [Lopez and Baker, 1994]. Further, this occurs in the tail at roughly 10 R_e downtail, and up to several hundred keV of energy may be involved, although the bulk of the plasma remains at far lower energies. This acceleration is quite distinct, in location and energy, from that which occurs at much lower altitude above auroral arcs. Particles traversing the lower acceleration region are typically accelerated to 1-10 keV, and account for brightening of the auroras at expansion phase onset. Aspects of the use of both high and low energy particles in studying

substorms will be discussed in Chapter 3. Here it will suffice to note that a large particle data set exists with which substorm injections may be studied [Baker *et al.*, 1982], and it has been amply demonstrated that sharp flux increases are associated with substorm onsets.

Figure 1.23 Electron fluxes in the 30 to 300 keV range at synchronous satellite 1984-037. Lowest energy is at top in each set of curves. UT is marked on bottom axis, local time on top horizontal axis.

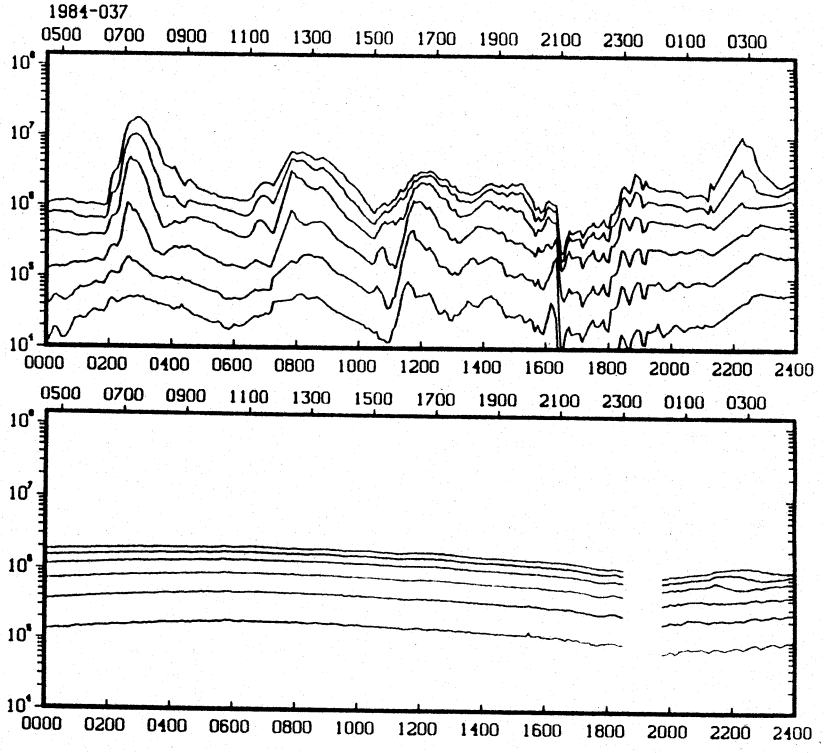


Figure 1.23 shows part of the large synchronous orbit dataset and compares fluxes on a quiet day with those from a more active period. In general, on the active day the highest energy fluxes are depressed and the lowest energy fluxes experience large transient increases. Such behaviour is typical of substorm injections, and with several spacecraft the energy dependent drifts allow determination of the injection region. In this example a flux dropout which is likely due to a change in magnetic field configuration is seen at roughly 1620 UT. The impulsive flux increases at earlier times would most likely be associated with injections. Further examples of injections associated with expansive phase onset may be found during the major substorms studied in later chapters.

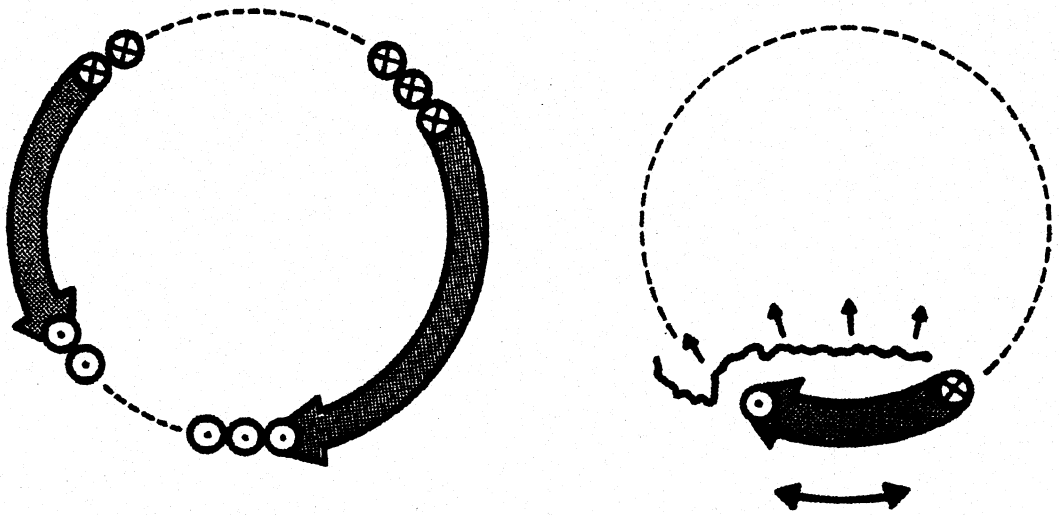
This section has briefly reviewed aspects of substorms, which appear as transient events within the larger magnetospheric context presented earlier in the chapter. We now examine the essential characteristics of electric current systems that flow during substorms, which form the basis of global modelling undertaken later in the thesis.

f. Electric Current Systems for Substorm Models

In Section 1.c above, basic aspects of magnetosphere-ionosphere coupling have been detailed. Figure 1.8 illustrated the auroral oval, and the results of Rostoker and Phan [1986] in outlining its average position through observations of ground magnetic perturbations were given. Ground observations form the basic data for modelling in this thesis. As was alluded to above, the ground perturbations reflect the positions of electrojets, that is, regions of electric current flow in the ionosphere. It will be shown in Section 4.d that two basic types of current system exist and that they are to a large degree colocated. Systems whose ionospheric currents are largely perpendicular to the trend of the auroral oval, such as those associated with the field-aligned currents whose patterns are shown in Figure 1.11, do not produce large perturbations at ground level. Systems with large longitudinal extent and current flow basically parallel to the auroral oval do produce relatively large perturbations on the ground. From observations of these perturbations, a general picture of three main current systems has been built up, and is shown in Figure 1.24. In this diagram the leftmost sketch shows schematically the three-dimensional current system associated with the electrojets usually present in the ionosphere, described above as DP2 or the 'driven' system. Downward net field-aligned current (shown by circled crosses) on the dayside feeds ionospheric currents flowing toward the nightside, where it leaves the ionosphere as upward field-aligned current (shown by circled dots). This results in large regions of nearly meridional electric current flow in the ionosphere which account for the overall pattern deduced by Rostoker and Phan. The figures showing their results also give an idea of the latitudes at which the large-scale currents flow. Generally there is eastward current flow over the evening sector of the Earth, and this gives rise to a northward perturbation at ground level, as was noted in association with Figure 1.21. The morning sector is typified by westward current flow and southward perturbations at the surface.

On the right hand side in the figure, the current associated with the substorm current wedge is shown, once more with a region of downward current shown as a circled cross and a region of upward current as a circled dot. The wedge current may be compared to the canonical view of Figure 1.20. The dynamic development is indicated by small arrows indicating poleward motion and east-west spreading of the region of current flow. These are generally similar to the arrows indicating the development of visual auroras as seen in Figure 1.19. The two types of systems shown are felt to schematically represent the patterns of electric current present in the auroral zone. The degree to which one or both types of current system are present, and the relation between them, are subjects of active research. The morning and evening system electrojets forming the 'driven system' appear to exist independent of substorm activity although they certainly are involved in substorms as cursory examination of magnetometer data has already shown (refer once more to Figure 1.21). As discussed, the current wedge is one of the most prominent features in substorm development.

Figure 1.24 Simplified ionospheric current systems for substorm models. Modified from Kamide and Baumjohann [1993].



The sketches shown inspire the approach to global modelling in this thesis. If, as appears to be the case, most magnetic perturbations associated with the auroras arise from such systems, then it should be possible to use models based on three-dimensional representations of the driven system currents and the wedge to reproduce the magnetic data observed during active times. The present study takes this approach by using a modelling routine with three basic current systems, and attempting to vary their parameters until a good match to the input data is found. The current systems parametrized in this manner constitute an inversion of ground magnetic data in terms of realistic physical sources. The feasibility and utility of this approach to inverting magnetic field data associated with substorms will be evaluated in what follows. After further discussion of relevant physics and instrumentation, as well as of other techniques allowing inversion of observations of magnetic field perturbations into representations of causative current systems, a new technique allowing direct inversion of the Biot-Savart integral in three dimensions will be presented. This technique, Automated Forward Modelling (AFM), will be used in five event studies for which good data are available. Comparison of AFM results with other information about the events allows deduction of numerous aspects of auroral phenomenology. AFM's advantages, including the fact that it may be used globally or regionally as appropriate, will be exploited, and its place among techniques for studying the Solar-Terrestrial interaction will be evaluated.