

# **The Earth's Magnetosphere**

*An article for the Encyclopedia of Geomagnetism and Paleomagnetism*

Stanley W H Cowley

Department of Physics & Astronomy

University of Leicester

November 2003

## The Earth's Magnetosphere

The Earth's magnetosphere is the region surrounding the planet, above the outer atmosphere and ionosphere, which contains and is controlled by the Earth's magnetic field. It extends from an altitude of  $\sim 500$  km above the Earth's surface to an outer boundary which is formed by the interaction of the planetary magnetic field with the solar wind, the plasma (charged particle) gas that streams continuously outward from the Sun. On the dayside, the compressive effect of the solar wind flow confines the field to a region extending  $\sim 10$  Earth radii from the Earth, while on the nightside the field is stretched into a long comet-like tail which extends typically in excess of  $\sim 1000$  Earth radii. The Earth's radius,  $R_E \sim 6400$  km, is thus a convenient measure of magnetospheric spatial scales. Electric currents flowing in these regions, whose magnetic effects can be observed at the Earth's surface, include those flowing at the solar wind-magnetosphere boundary, those produced by the drift of energetic charged particles inside the magnetosphere, and those associated with the coupling between the magnetosphere and the ionosphere. Magnetic perturbations observed at the Earth's surface due to these three sources peak typically at a few tens, a few hundreds, and a few thousands of nT, respectively, and vary on a range of time scales from minutes to hours and days. In this article we first concentrate on the overall structure and basic physics of the coupled solar wind-magnetosphere-ionosphere system, and then consider the dynamic processes which occur, leading to strong temporal variations.

### *Structure of the magnetosphere*

In Fig. 1 we show a cross-section through the Earth's magnetosphere in the noon-midnight meridian plane, where the arrowed solid lines indicate magnetic field lines. The small symbols indicate the main plasma populations, while the arrows show their principal flows. The small dashes show the solar wind plasma, together with the magnetospheric 'boundary layer' plasma populations derived directly from it. The direction of the Sun is towards the left, so that the solar wind flows from left to right as indicated. The solar wind derives from hydrogen gas in the solar atmosphere, which is heated so strongly in the solar corona (to temperatures in excess of a million degrees) that the atoms are fully broken up by collisions into protons and electrons, which then stream continuously out into the solar system at speeds  $\sim 500$  km s<sup>-1</sup>. At the orbit of the Earth the ion and electron number densities are  $\sim 10$  cm<sup>-3</sup>, which is sufficiently low that the particles in the gas are collision-free, meaning that the mean free path for particle collisions is comparable with or larger than the size of the system. The second source of plasma in the magnetosphere derives from the Earth's upper atmosphere, which is relatively cool ( $\sim 1000$  K) but is partially ionised by solar far-UV and X-rays at altitudes above  $\sim 100$  km, thus forming the ionosphere. These charged particles, consisting of protons and heavy ions, principally singly-charged oxygen, together with corresponding electrons, can also stream out of the topside ionosphere into the magnetosphere, as indicated by the small crosses in the figure. In the ionosphere, collisions between ions

and neutral atoms are significant at altitudes up to  $\sim 500$  km, while electrons are essentially collision-free throughout the ionosphere. Above  $\sim 500$  km, however, the ionospheric ions also become collision-free due to falling densities, this altitude thus being chosen (somewhat arbitrarily) above as the ‘base’ of the magnetosphere. Like the solar wind, therefore, the plasma in the Earth’s magnetosphere is also collision-free. However, as well as the plasma populations, a neutral hydrogen atmosphere of diminishing density (the geocorona) is found to extend away from the Earth to distances of at least  $10 R_E$ .

When collisions can be ignored, the motion of ions and electrons in the plasma is governed by the forces of the large-scale fields prevailing, and for both solar wind and magnetosphere the most important forces by far are electromagnetic. Solar and planetary gravity is only important in the regions close to the bodies concerned, i.e. in the solar corona near the Sun, and in the ionosphere near the Earth. The motion of charged particles in a large-scale electromagnetic field consists of a number of components which take place on generally widely-separated time scales. The first is a rapid gyration around the magnetic field lines in nearly circular orbits whose radius is generally tiny compared with the scale size of the systems. The second is a uniform motion along the magnetic field lines, which is subject to the magnetic mirror force. This repels particles from regions of increasing field strength along the field lines (i.e. is directed away from regions where field lines converge). The mirror force is produced by the action of the field gradient on the magnetic dipole formed by the gyrating particle. In the quasi-dipolar field of a planetary magnetosphere, for example, the field strength along a given field line is weakest at the equator and increases greatly towards the planet. This configuration can therefore trap charged particles almost indefinitely, the mirror force resulting in a bounce motion between mirror points in the northern and southern hemisphere. Only those particles that are directed almost along the field lines at the equator, and hence have small magnetic moments, can reach down to the upper layers of the atmosphere. The third motion then consists of large-scale plasma flows transverse to the magnetic field lines, such as that shown for the solar wind in Fig. 1. Such flows are associated with a large-scale electric field  $\mathbf{E}$  directed transverse to the magnetic field  $\mathbf{B}$ , and related to the plasma velocity  $\mathbf{V}$  by the vector relation  $\mathbf{E} = -\mathbf{V} \times \mathbf{B}$ . Inspection of the flows indicated in Fig. 1 then shows that the electric field is directed everywhere out of the plane of the diagram, though it is not everywhere of equal strength. In such ‘crossed’ electric and magnetic fields, charged particles drift perpendicular to the magnetic field with velocity  $\mathbf{V} = (\mathbf{E} \times \mathbf{B})/B^2$  independent of their charge or mass, this expression being exactly equivalent to the vector relation for  $\mathbf{E}$  given above. This ‘ $\mathbf{E} \times \mathbf{B}$  drift’ has a very special property, first discovered by Alfvén in 1942, that particles whose centres of gyration lie initially on the same field line as each other, remain on the same field line as each other for all time, thus showing why plasma structures are strongly organised by magnetic field lines, as indicated by various features in Fig. 1. We may picture the collective behaviour as one in which the plasma and field are ‘frozen together’. However, we can think either of the plasma as moving and carrying the field lines, or equivalently of the field lines as moving and carrying the plasma.

Which of these pictures is the most appropriate for a given system depends upon the relative energies in the plasma and magnetic field. Thus, for example, since the plasma energy exceeds the magnetic energy in the solar wind, the solar magnetic field is carried outwards ‘frozen into’ the expanding plasma flow, forming a large-scale interplanetary magnetic field (IMF) that pervades the entire solar system. The strength of the IMF at the Earth’s orbit is typically  $\sim 5\text{-}10$  nT, directed, on the average, in the ecliptic plane at an angle of  $\sim 45^\circ$  to the Earth-Sun line. The latter tilt is due to the rotation of the Sun, which winds the IMF into a spiral form as the field lines are carried out into the solar system, frozen into the flow. On the other hand, in the quasi-dipolar field regions of the magnetosphere and ionosphere, where the magnetic energy dominates, we instead think of the field lines as moving, transporting the plasma.

We may apply the frozen-in concept to the interaction between the solar wind and the planetary magnetic field. Since the solar wind and IMF are frozen together, as well as the planetary field and planetary plasma (e.g. from the ionosphere), then when these two media interact they will not mix. Instead, the solar wind confines the planetary field to a cavity surrounding the planet, around which it flows, as first deduced by Chapman and Ferraro in 1931. This magnetic cavity is the planet’s magnetosphere, whose outer boundary shown by the dashed line in Fig. 1 is called the ‘magnetopause’. A ‘bow shock’ also forms ahead of the cavity, as also shown in Fig. 1, due to the fact that the magnetosphere represents a blunt obstacle in the supersonic solar wind flow. Across the shock the solar wind is slowed, compressed, and heated, forming the turbulent ‘magnetosheath’ layer located between the shock and the magnetopause boundary.

The size of the magnetospheric cavity is set by the condition of pressure balance. A simple estimate of the distance of the equatorial boundary at noon (the minimum distance in the direction facing the Sun), can be made by equating the solar wind ram pressure on one side of the boundary with the magnetic pressure ( $B^2/2\mu_0$ ) of the compressed planetary field on the other. With typical solar wind values the radial distance of the boundary is then estimated to be  $\sim 10 R_E$ , as is observed, a position which may vary by factors of up to two in either direction, in extreme circumstances. The strength of the compressed planetary field just inside the boundary is typically  $\sim 60$  nT, representing a planetary dipole field of  $\sim 30$  nT enhanced by a factor of  $\sim 2$  by the electric current flowing in the magnetopause boundary. This latter current is termed the ‘Chapman-Ferraro’ current, and inspection of Fig. 1 with Ampère’s law in mind shows that it flows out of the plane of the diagram in the equatorial region on the dayside, closing over the magnetopause into the plane of the diagram over the polar regions and on the nightside. These rings of magnetopause current produce a perturbation magnetic field in the Earth’s vicinity whose strength is typically a few tens of nT, directed northward (i.e. upward in Fig. 1) hence enhancing the horizontal field at low and middle latitudes at the Earth’s surface.

Frozen-in behaviour of the field and plasma is not, however, a universally valid description. In a collision-free medium it is broken in particular by the presence of particle drifts which are additional to the  $\mathbf{E} \times \mathbf{B}$  drift introduced above. The most significant of these effects causes ions and electrons to drift in opposite directions (thus producing a current) due to the presence of gradients in the strength and direction of the magnetic field. These drifts are proportional to the field gradients, and also to the particle energy, thus being more important for particles of higher energy. When the field gradients are weak, these additional drifts are important only for particles in the high-energy tail of the distribution, and frozen-in motion then represents a useful organising concept for the bulk of the plasma population. This limit applies essentially throughout the solar wind, and through most of the Earth's magnetosphere. However, when the field gradients are very strong, such that the motion of the bulk of the plasma particles is effected, then the frozen-in picture breaks down. One such place where this happens is the magnetopause boundary itself, where the field strength and direction in general switches rapidly from magnetosheath to magnetospheric values across the magnetopause current sheet. The simplest description of the consequence of frozen-flux breakdown is that the magnetic field diffuses through the plasma, locally, in the region of the strong gradient. As first pointed out by Dungey in 1961, this then allows magnetic field lines to become joined across the boundary, producing 'open' magnetic field lines which pass from the solar wind at one end, through the magnetopause, to the Earth's polar regions at the other. This process is called 'magnetic reconnection'. Two newly-reconnected open field lines are shown in Fig. 1 passing through the dayside magnetopause shortly after reconnection has taken place near the equator. Sharply-bent magnetic field lines exert a force on the plasma like the force of rubber bands (the force per unit volume being  $\mathbf{j} \times \mathbf{B}$ , where  $\mathbf{j}$  is the current density in the plasma), in this case accelerating the boundary plasma poleward away from the equator, such that the field lines also contract poleward, releasing energy to the plasma and allowing further reconnection to proceed at the equator. Subsequently, the open field lines are carried downstream frozen into the magnetosheath flow, and are stretched out into a long cylindrical comet-like tail. This tail consists of two lobes, D-shaped in cross-section, one connected to the northern polar region at Earth, the other to the southern, as indicated in Fig. 1. Observations show that the tail lobe field lines remain open typically for a few hours, such that with a downstream speed of  $\sim 500 \text{ km s}^{-1}$ , the tail is typically  $\sim 1000 R_E$  long.

The open field lines form magnetic pathways along which the magnetosheath plasma may enter the magnetosphere. Such plasma thus flows along newly-opened field lines to form a boundary layer adjacent to the dayside magnetopause, and the 'cusp' population as it then moves down towards the Earth (see the magnetospheric dashed regions in Fig. 1). The majority of the particles are, however, repelled by the magnetic mirror force as the field strength increases near the Earth, and hence move back out again towards the outer magnetosphere. Due to the anti-sunward motion of the open field lines, however, the cusp plasma flows back out into the lobes of the tail, in the region adjacent to the magnetopause, forming

the ‘plasma mantle’ population. As the open field lines are carried down the tail, so the field lines and mantle plasma sink in towards the centre plane, followed by further entry of anti-sunward flowing magnetosheath plasma at the tail magnetopause, such that the mantle grows thicker and with increasing density at larger distances. Plasma from the Earth’s polar ionosphere also flows into the lobes (cross symbols in Fig. 1), but because of its low velocity along the field lines ( $\sim 100 \text{ km s}^{-1}$ ), this plasma does not reach far down the tail on the few-hour time scale that the lobe field lines remain open. Overall, the plasma density in the inner part of the tail lobes is very low,  $\sim 0.01\text{--}0.1 \text{ cm}^{-3}$ , and with temperatures typically of order a few tens to a few hundred eV, most of the system energy resides in the lobe magnetic field.

The residency of the oppositely-directed open field lines in the two tail lobes is terminated when they sink into the centre plane of the tail and reconnect in the equatorial current sheet that separates them, as shown by the X-shaped field configuration on the right side of Fig. 1. On the tailward side of the tail reconnection site the lobe field lines become disconnected from the Earth, and accelerate the lobe plasma rapidly tailward, where it eventually rejoins the solar wind. On the Earthward side, the process forms new ‘closed’ field lines, connected to the Earth at both ends, which similarly contract rapidly Earthward under the effect of field tension, compressing and heating the lobe plasma as they do so. This plasma, containing both solar wind and ionospheric contributions, is termed the ‘plasma sheet’ population, and is shown by the dotted region in Fig. 1. In the near-Earth tail it is characterised by densities  $\sim 0.1\text{--}1 \text{ cm}^{-3}$ , and electron and ion temperatures of  $\sim 0.5$  and  $\sim 5 \text{ keV}$ , respectively, but both density and temperatures increase as the plasma flows sunward into quasi-dipolar inner magnetosphere and is further compressed and heated. Throughout these regions the hot plasma sheet population carries a current, due to the field-gradient effects mentioned above, which is directed westward, i.e. out of the plane of Fig. 1 on the nightside of the Earth, and into the plane of the diagram on the dayside. On the nightside within the tail, the current paths reach the tail boundary, and close over the magnetopause to form two D-shaped solenoids lying back-to-back, as required by the lobe field structure. Nearer the Earth, the current paths may close wholly round the Earth to form a plasma ‘ring current’. Overall, these currents produce a perturbation field at the Earth which is directed southward (i.e. downward in Fig. 1), and is usually a few tens of nT in amplitude. However, as we will discuss below, these currents become greatly enhanced during magnetic storms, producing perturbation fields an order of magnitude larger.

The closed field lines and hot plasma produced in the tail subsequently flow sunward around the Earth in the quasi-dipolar magnetosphere, eventually reaching the dayside magnetopause where they become open again, allowing the hot magnetospheric plasma to escape into the magnetosheath, as indicated by the dotted region outside the magnetopause. Overall, therefore, reconnection at the magnetopause and in the tail results in the formation of a large-scale cyclical flow of field lines and plasma through the Earth’s

magnetosphere, as first discussed by Dungey in 1961. The overall cycle time is ~12 hours, of which the field lines spend ~4 hours open in the tail lobes, and ~8 hours closed flowing sunward through the central magnetosphere. The solar wind is not, however, the only source of momentum that drives magnetospheric flow. A second source resides in the rotation of the Earth, and the coupled rotation of the Earth's upper atmosphere. Collisions between ions and neutral atmospheric particles in the ionosphere produce a frictional force on the feet of the magnetospheric field lines which tends to spin them up towards rotation with the planet, a flow called 'corotation'. In general, the overall magnetospheric flow results from a summation of the effects of the planetary and solar wind driving forces. The corotation effect tends to produce a flow in the equatorial plane increasing linearly with the distance from the planet (i.e. a flow with the planetary angular velocity), while in the simplest picture of a dipole magnetic field and uniform cross-system electric field, the equatorial Dungey-cycle flow increases as the cube of the distance. Reflection on these relative dependencies then shows that corotation will dominate close in, and Dungey cycle convection at larger distances. For Earth, the boundary between these flow regimes is located typically at ~5  $R_E$  in the equatorial plane, about half-way to the dayside magnetopause. The hot plasma flowing in from the tail is therefore excluded from this inner core of corotating field lines, as indicated in Fig. 1, which is instead filled to relatively high densities (~100–1000  $\text{cm}^{-3}$ ) by cold (~1–10 eV) plasma from the ionosphere. This population is termed the plasmasphere, and is indicated by the near-Earth region of dense crosses in the figure.

### ***Magnetosphere-ionosphere coupling***

The magnetospheric flows discussed above produce corresponding motions of the field and plasma in the polar ionosphere, resulting in large-scale current systems flowing between these regions. Fig. 2 shows a view looking down on the north polar region, with noon (the direction towards the Sun) at the top of the diagram, dusk to the left, dawn to the right, and midnight at the bottom. The central circular dashed line shows the outer boundary of open field lines (typically located at magnetic latitudes of ~75° at noon and ~70° at midnight), while the arrowed solid lines show plasma streamlines. The open field lines within the dashed line boundary, mapping to the northern tail lobe, flow anti-sunward from noon to midnight at speeds typically of a few  $100 \text{ m s}^{-1}$ , while the closed field lines at lower latitudes return sunward. This twin-vortex flow pattern is the signature of the Dungey cycle in the ionosphere. The  $-\mathbf{V} \times \mathbf{B}$  electric field associated with the flow is directed from dawn to dusk across the open field line region as indicated, reversing in sense to poleward at dusk and equatorward at dawn. The small symbols also show the magnetospheric plasma populations into which the ionospheric field lines map, the symbol code corresponding to that in Fig. 1. Ions and electrons from the cusp (dashes) and plasma sheet (dots) populations precipitate into the upper atmosphere, excite the atoms (typically at altitudes ~100–300 km), and cause them to emit photons, giving rise to one component of auroral light emissions. They also

ionise the neutral atoms and increase the plasma density of the ionosphere, a process which is especially important at night when the solar production mechanism is inoperative.

The Dungey-cycle flow drives two components of electric current in the lower ionosphere, both resulting from the effects of collisions between ionospheric ions and atmospheric neutral particles. At altitudes below  $\sim 120$  km, these collisions become sufficiently frequent that the ions do not move with the field lines at all, but are essentially tied to the generally smaller motions of the upper atmosphere (i.e. the winds in the neutral thermosphere). However, the electron motion remains almost unaffected by collisions, frozen to the field line motion throughout the ionosphere, to altitudes below  $\sim 100$  km. Consequently, in the layer between  $\sim 100$ – $120$  km, ionospheric electrons flow around the streamlines shown in Fig. 2, carrying a current in the opposite direction, while the ions in the layer remain almost static, tied to the neutral atmosphere. This current, directed transverse to both the electric and magnetic fields, is thus termed a Hall current, and with a height-integrated ionospheric conductivity of  $\sim 10$  mho and typical ionospheric flows of several hundred  $\text{m s}^{-1}$ , it produces equal and opposite magnetic perturbations on either side of the ionospheric Hall layer which are typically  $\sim 100$ – $200$  nT in strength. At the Earth's surface in the northern hemisphere these perturbation fields are roughly in the direction of the overhead ionospheric electric field (opposite in the southern hemisphere). These currents are non-dissipative (i.e.  $\mathbf{j} \cdot \mathbf{E} = 0$ ), and in principle can close wholly within the ionosphere (for uniform conductivity) by flowing round the plasma flow streamlines.

The second current component flows at a slightly higher altitude of  $\sim 120$ – $150$  km, where the ion motion is effected by collisions with neutrals, but is not yet wholly dominated by them. In this layer collisions provide the ions with some mobility in the direction of the electric field, while continuing to move approximately with the electrons along the plasma streamlines. An ion current thus flows in the direction of  $\mathbf{E}$ , termed the Pedersen current, and its magnitude is such that the  $\mathbf{j} \times \mathbf{B}$  force of the current, directed along the plasma streamlines, just balances the frictional drag on the ionospheric plasma due to ion-neutral collisions in the ionosphere. A key fact about the Pedersen current is that it cannot in principle close within the ionosphere itself, but is instead part of a large-scale current system which flows between the magnetosphere and ionosphere, which imposes the flow of the former onto the latter in the presence of ionospheric ion-neutral drag. It can be seen in Fig. 2, for example, that the Pedersen currents, directed along  $\mathbf{E}$ , flow away from the open-closed field line boundary on both sides at dawn, while flowing towards the boundary from both sides at dusk. Current continuity then requires the presence of field-aligned currents (FACs) which flow down the field lines from the magnetosphere to the ionosphere at the dawn boundary, and up the field lines from the ionosphere to the magnetosphere at the dusk boundary. These are called the “region 1” FACs, whose direction is indicated in Fig. 2 by the circled dots and crosses for upward and downward currents respectively. These FACs close across the field lines at

large distances, in the plasma which provides the momentum producing the flow, i.e. in the magnetosheath plasma at the tail magnetopause. The magnetic perturbations produced by the overall current circuit bends the magnetic field lines in such a way that the  $\mathbf{j} \times \mathbf{B}$  force slows the magnetosheath plasma, while transferring the momentum to the ionosphere to maintain the flow against collisional frictional drag. Overall, this current system is solenoidal in nature, such that the main magnetic effects are contained within the effective current solenoid between the ionosphere and magnetopause. Thus although the ionospheric height-integrated Pedersen conductivity is generally comparable with the height-integrated Hall conductivity, such that the two height integrated current components are of similar intensity in the ionosphere, the combined contribution of the ionospheric Pedersen current and the associated FACs approximately cancels underneath the ionosphere, and produces a much weaker magnetic perturbation on the ground than that of the Hall current. Above the ionospheric Pedersen layer, however, the perturbation field produced by the Pedersen-FAC system is directed approximately opposite to the plasma streamlines in the northern hemisphere, and along the streamlines in the southern hemisphere. Typical amplitudes in the region just above the Pedersen layer are again a few 100 nT. The total current flowing in the ‘region 1’ FAC system is typically  $\sim 2\text{--}3$  MA.

We also note from Fig. 2 that Pedersen current continuity at the equatorward boundary of the Dungey cycle flow (typically located at magnetic latitudes of  $\sim 70^\circ$  at noon and  $\sim 65^\circ$  at midnight), also requires the presence of upward FACs at dawn and downward FACs at dusk. These are called the ‘region 2’ FAC system, in which the upward currents at dawn close in the downward currents at dusk via field-perpendicular currents flowing westward in the inner edge of the equatorial plasma sheet population in the nightside magnetosphere (and then via ionospheric Pedersen currents, the ‘region 1’ FAC, and the tail magnetopause). The total current flowing in the ‘region 2’ FACs is a little less than that in the ‘region 1’ FACs, due to the fact that the latter currents are fed by Pedersen currents from both open and closed field line regions. We also note that regions of upward-directed FAC are typically associated with bright ‘discrete’ auroras (e.g. structured curtain-like forms), thus forming another component of the auroral emissions. This is because upward currents are carried primarily by warm plasma sheet electrons which flow down the field lines into the ionosphere. In order to produce a downward electron flux which is sufficient to carry the upward current, the electrons must be accelerated downward by an electric field directed upward along the field lines (this being another scenario in which the frozen-in picture breaks down). Typically the required field-aligned voltages are a few kV, such that the precipitating electrons carry a sufficient energy flux to produce a bright aurora.

### ***Magnetospheric dynamics***

The above discussion has focussed on the plasma physics principles which govern the structure and properties of the coupled magnetosphere-ionosphere system, focussing particularly on the nature of the

current systems that flow within it. For clarity of exposition we have viewed the system as being steadily driven, principally by the solar wind, but also by planetary rotation. However, it is now important to emphasise that the Earth's plasma environment is almost never in such a steady state, but generally varies strongly with time. There are two reasons for this. The first is that the interplanetary medium which impinges upon the Earth's field itself varies strongly with time, on time scales from a few minutes to the ~27-day rotation period of the Sun. There are also variations over the 11-year period of the solar cycle. The second factor is that even when the solar wind is relatively steady, the processes in the key interaction regions at the magnetopause and in the tail show a propensity for pulsed behaviour. We will now outline these behaviours, beginning here with those which are driven essentially directly by variations in the upstream medium.

Although the density and velocity of the solar wind determine the size of the magnetospheric cavity and the speed with which open flux tubes are transported to the tail, by far the most important upstream parameter which determines the magnetospheric interaction with the solar wind, and hence the dynamics of the Earth's plasma-field environment, is the direction and strength of the IMF. This determines how much flux is reconnected at the magnetopause per unit time, where on the magnetopause it is reconnected, and how the newly-formed open tubes move into the tail. Although the IMF on average lies in the ecliptic plane at the ~45° spiral angle mentioned above, pointing either toward or away from the Sun along this direction, variable north-south fields are also produced by a variety of effects, such as waves and other disturbances propagating in the wind outward from the Sun. Open flux production is largest, and the Dungey-cycle flow strongest, when the IMF is directed southward opposite to the Earth's equatorial field (i.e. points downward in Fig. 1, as is shown), and is weak when the IMF is directed northward (i.e. points upward in Fig. 1). From Faraday's law the overall magnetic flux transfer in the Dungey cycle, in  $\text{Wb s}^{-1}$ , is equal to the voltage of the cross-system electric field associated with the flow, in Volts. This cannot be measured directly in the magnetosphere, but can routinely be measured at ionospheric heights using a network of ground-based radars which determine the flow in the polar ionosphere. These observations show that when the IMF, of strength ~5-10 nT, points south, the voltage associated with the flow is ~100 kV, i.e. the flux throughput in the Dungey cycle is ~100  $\text{kWb s}^{-1}$ . This compares with voltages in the solar wind across the magnetospheric diameter of about five times this value. Reconnection is thus ~20% efficient during such southward field intervals, in the sense that ~20% of the interplanetary magnetic flux which flows towards the magnetosphere in the solar wind becomes reconnected with the terrestrial field. At the same time, of course, ~80% of the interplanetary magnetic flux is deflected around the magnetosphere as predicted by Chapman and Ferraro. Nevertheless, it is this breakdown of perfect deflection at the ~20% level which is critical to magnetospheric dynamics. As the IMF direction rotates away from southward, however, reconnection becomes less efficient. For example, when the IMF lies in its average direction in the ecliptic plane, the voltage values fall to ~50 kV, while

when it rotates to northward they fall to ‘background’ values of  $\sim 20$  kV or less. The Dungey-cycle flow in the magnetosphere-ionosphere system is therefore strongly modulated by the sense and magnitude of the north-south component of the IMF, and is found to respond rapidly to changes in this component on time scales down to a few minutes.

The form of the flow is also found to change not only in response to the north-south field component of the IMF, usually referred to as the ‘z’ component (with  $B_z$  positive northward), but also to the component which points into or out of the plane of the diagram in Fig. 1, usually referred to as the ‘y’ component (with  $B_y$  positive out of the plane of the diagram). While the  $B_z$  component of the IMF typically fluctuates in sense on time scales of minutes to a few tens of minutes, driving similarly variable Dungey-cycle flows as described above, the  $B_y$  component is mainly determined by the large-scale structure of the solar field in the coronal source plasma (pointing either toward or away from the Sun), and hence typically remains steady in polarity, if not in strength, over intervals of several days (i.e. some fraction of the solar rotation period). When dayside reconnection takes place in the presence of IMF  $B_y$ , the field tension effect on newly-opened flux tubes pulls the open tubes ‘sideways’ (i.e. east-west) as well as poleward, oppositely in the two hemispheres. This is illustrated schematically in Fig. 3, which shows a view of the dayside magnetosphere looking from the direction of the Sun shortly after reconnection has taken place with an IMF having a positive  $B_y$  component. The field tension (rubber band) force pulls the newly opened field lines westward (to the left) in the dayside cusp in the northern hemisphere, and simultaneously eastward (to the right) in the dayside cusp in the southern hemisphere. The open field lines are correspondingly carried asymmetrically into the tail lobes, preferentially into the dawn side of the northern lobe, and into the dusk side of the southern lobe, an effect which tends to twist the tail structure on the nightside. The corresponding ionospheric flows in the northern hemisphere are shown in the upper row of diagrams in Fig. 4, for near-zero or negative IMF  $B_z$  and various IMF  $B_y$  as indicated. Again noon is at the top of each diagram and dusk to the left. In this case the region of open field lines is relatively expanded and the twin cell Dungey cycle flows well-developed due to the negative IMF  $B_z$ , while the IMF  $B_y$  component produces dawn-dusk asymmetries in the flow with newly-opened cusp field lines flowing west for  $B_y$  positive (as indicated in Fig. 3), and east for  $B_y$  negative. The simultaneous east-west flows are opposite in the southern hemisphere. These variations of the flow in the cusp region (and beyond) produce corresponding variations in the Hall and Pedersen-FAC current system that can be sensed both on the ground and above the ionosphere according to the general principles outlined above. The perturbations observed at magnetic observatories lying under the cusp (at  $\sim 75^\circ$  magnetic latitude near noon) can be used in particular to reconstruct the large-scale structure of the IMF over long intervals from historic magnetic records.

Observations indicate that reconnection at the dayside magnetopause tends to occur preferentially in regions where the magnetospheric and magnetosheath field adjacent to the magnetopause are anti-parallel, such that the ‘magnetic gradients’ across the boundary are largest. When the IMF turns from southward to northward, these regions migrate from low to high latitudes, such that reconnection can then take place between northward IMF and open field lines in the tail lobes that were produced by earlier intervals of southward field. Such ‘lobe reconnection’ does not change the amount of open flux in the tail, but causes it to circulate within the region of open field lines, as the newly reconnected field lines flow around the dayside magnetopause and back into the lobe. The patterns of lobe circulation produced by this effect are shown in the lower row of diagrams in Fig. 4, which also indicates that the amount of open flux present is usually smaller under these conditions. Again, these patterns of ionospheric flow are associated with corresponding patterns of magnetic perturbations above and below the ionosphere, produced by the corresponding Hall and Pedersen-FAC currents that are driven.

### ***Geomagnetic storms***

Usually the  $z$  component of the IMF is a few nT in magnitude and fluctuates in sense on time scales of minutes to a few tens of minutes, as mentioned above. This drives variable Dungey-cycle flows as already outlined, as well as impulsive magnetospheric responses in the nightside tail that will be discussed in more detail in the next section. However, under some specific circumstances the IMF can become very strong (e.g. several tens of nT), and can remain southward pointing for extended intervals of several hours. Under these circumstances a geomagnetic storm is produced at the Earth in which the horizontal (northward) field is depressed globally over intervals of hours and days, the depression typically maximizing at a few tens to a few hundreds of nT. These effects are masked at high latitudes, however, by magnetic perturbations occurring on shorter time scales (typically several tens of minutes to an hour or two) which can reach values an order of magnitude larger. The origins of the impulsive high-latitude magnetic effects will be discussed in the next section.

For reasons that will be outlined below, the occurrence of magnetic storms tends to follow the 11-year solar sunspot cycle, as first noted by Sabine in 1852, but on average there are roughly ten storms each year whose mid-latitude field depression exceeds 50 nT, and one that exceeds 250 nT. The largest depression observed since systematic records began in 1957 was of ~600 nT during a storm in March 1989, which occurred near solar maximum. The magnetic depression typically grows fairly gradually over an interval of several hours (~2–12 hours), termed the ‘main phase’, and then decays even more gradually over the following several days (~1-5 days), termed the ‘recovery phase’. These effects result from a prolonged enhancement of the Dungey-cycle flow which is produced by the prolonged interval of strong southward fields. During such intervals the boundary between corotating and Dungey-cycle flow in the inner equatorial magnetosphere shrinks significantly in size, such that the outer layers of the pre-

existing plasmasphere are stripped off and flow out to the dayside magnetopause. Correspondingly, the hot plasma produced in the tail flows inward to replace it, significantly enhancing the westward-directed ‘ring current’ produced by the differential drift of energetic ions and electrons in the gradients of the quasi-dipolar magnetic field, as outlined above. The perturbation field produced by the hot inner plasma is a major contributor to the ‘main phase’ field depression at Earth. However, the picture is complicated by the fact that the inflowing hot plasma is usually asymmetric in its distribution around the Earth, leading to partial current ‘rings’ that close in the ionosphere via ‘region 2’-type FACs. These FACs also contribute to the overall magnetic disturbance, as do the combined ‘fringing’ fields of the enhanced plasma sheet and tail currents on the nightside. During the recovery phase, however, Dungey cycle convection is reduced, so that these current systems also decline. Hot ‘ring current’ plasma now marooned within the newly-expanded corotating region slowly decays due to wave-scattering followed by precipitation into the mid-latitude atmosphere, and due to charge-exchange with neutral atoms in the geocorona. The outer newly-corotating flux tubes also refill with cold plasma from the ionosphere on time scales of a few days, thus re-forming a more extended plasmasphere.

As indicated at the beginning of this sub-section, the extended intervals of strong southward IMF which excite geomagnetic storms are associated with specific phenomena in the interplanetary medium which link storms with processes on the Sun and the solar cycle. Two principal mechanisms are responsible. In the first, a large loop-like field structure located in the solar corona suddenly becomes unstable and is ejected away from the Sun, forming a ‘coronal mass ejection’ (CME). The ejection speed is sometimes very rapid, exceeding  $\sim 1000 \text{ km s}^{-1}$ , such that the CME plasma and frozen-in field plough into the slower solar wind ahead of it. This creates a giant shock wave which propagates through the solar system compressing and heating the ambient solar wind medium ahead of the CME ‘driver gas’. Behind the shock the IMF is also compressed to high field strengths. When such a shock impinges on the Earth, the magnetosphere is impulsively compressed, such that the horizontal field at the Earth’s surface is suddenly increased, typically by a few tens of nT, corresponding to the effect of the increased Chapman-Ferraro current at the magnetopause. What happens after this depends on the direction of the enhanced IMF in the compressed solar wind and CME driver gas. These fields may have any orientation depending on the individual circumstances of the event, and may also fluctuate in direction during the event. A magnetic storm ‘main phase’ occurs only if the enhanced IMF points southward during some extended interval after the shock wave has passed. If this occurs (typically in about one in six events), the impulsive field enhancement at the Earth’s surface produced by the shock is called the ‘storm sudden commencement’ (SSC), while the variable interval between the SSC and the onset of the main phase (when the upstream IMF happens to turn southward) is called the ‘initial phase’ of the storm. If such an enduring interval of southward field does not occur, and with it no ‘main phase’ field depression, then the shock-related compression event is simply termed a ‘sudden impulse’. The connection between these

events and the solar cycle results from the fact that  $\sim 3$  CMEs occur each day near solar cycle maximum (only a small number of which impinge on Earth), reducing to one every several days during solar cycle minimum. The connection with solar flares, first noted by Carrington in 1860, results from the fact that flares often occur near the sites in the solar corona where CMEs have formed.

During the declining phase of the solar cycle, extended intervals of strong southward IMF can also be produced by another interplanetary mechanism, which thus can also generate geomagnetic storms. The Sun typically produces two types of quasi-steady solar wind outflow (as opposed to CMEs), a ‘slow’ variable wind of  $\sim 300\text{--}500 \text{ km s}^{-1}$  from regions surrounding closed field regions of the solar corona, and a ‘fast’ steady wind of  $\sim 800 \text{ km s}^{-1}$  from open-field ‘coronal holes’. During solar maximum, the solar field is typically very disordered, and with it the solar wind outflow. During the declining phase of the cycle, however, large-scale off-axis coronal holes tend to form, which migrate slowly toward the poles of the Sun as activity decreases towards solar minimum. Initially, however, these coronal holes may extend down to the solar equator, such that during each solar rotation the Earth experiences a pattern of fast and slow solar wind output that may persist in form for many solar rotations. The corotating variable solar wind source regions on the Sun thus give rise to ‘corotating interaction regions’ (CIRs) in the interplanetary medium, in which fast plasma outflow from an equatorial coronal hole ‘runs into’ slower plasma that was emitted earlier into the same direction from a non-coronal hole source region on the rotating Sun. This compresses the plasma and may enhance the frozen-in field by factors of  $\sim 5\text{--}10$  above usual values. As the solar wind velocity subsequently drops, however, after the passage of the material from the coronal hole, a rarefaction region is created in the solar wind, and with it regions of weak IMF occur. Geomagnetic storms can be excited by the enhanced fields of the compression phase of CIRs, during intervals in which the solar wind speed is increasing locally at the Earth, if the enhanced field happens to point southward for a significant interval. Such storms, however, are typically not so intense as those generated by CMEs because the field directions tend to fluctuate more during these events, as opposed to the more ordered field structures produced by CMEs. In addition, these events are not associated with interplanetary shocks, such that no SSC occurs prior to the onset of the main phase.

### ***Magnetospheric substorms***

In the above sections we have so far principally discussed the flows and currents that are driven in the coupled magnetosphere-ionosphere system by the Dungey cycle, viewed as a quasi-steady process that is strongly modulated by variations in the strength and direction of the IMF. However, observations show that even when interplanetary conditions are relatively steady, the driving processes in the key interaction regions at the magnetopause and in the geomagnetic tail are not. Reconnection on the dayside, in particular, often occurs as waves that propagate east-west over the magnetopause from the dayside towards the tail, which recur on time scales of  $\sim 5\text{--}10$  min. These ‘flux transfer events’ (FTEs) give rise to

pulsed magnetic and plasma signatures at the magnetopause, and pulsed flow, current, plasma injection, and auroras in the cusp ionosphere on the dayside. Due to the typically overlapping nature of the effect of these pulses, however, the overall flow and current modulations are generally modest.

The pulsed behaviour in the tail takes place on longer time scales of around 1–2 hours, however, and produces major perturbations of the flow and current in the nightside magnetosphere-ionosphere system, termed magnetospheric substorms. These are the principal focus of our discussion in this section. Substorms are again driven by intervals of southward-directed IMF, but can be excited by southward fields of a few nT lasting for a few tens of minutes, which happens very frequently, as opposed to the few tens of nT fields lasting for a few hours that are required for a geomagnetic storm, which happens rather rarely. Such relatively brief intervals of modest southward field are sufficient, however, to produce enhanced open flux production at the dayside magnetopause which in turn enhances the radius and field strength of the lobes of the tail. The current carried by the plasma sheet at the centre of the tail is also correspondingly intensified, while the current layer is also found to thin, within the near-Earth region, to thicknesses of only a few thousand km. After ~30-40 min of such tail development, termed the ‘growth phase’ of the substorm, a disruption is observed to occur within the plasma sheet. The origin and development of the disruption remain controversial, but it involves the formation within a minute or two of a new reconnection region within the plasma sheet in the central tail at down-tail distances of 20–30  $R_E$ . Reconnection at this site first pinches off the tailward portion of the pre-existing plasma sheet, forming a closed-loop plasmoid field structure, illustrated schematically on the right hand side of Fig. 1, which is expelled tailward into the solar wind at speeds in excess  $\sim 500 \text{ km s}^{-1}$ . Subsequently, open field lines in the tail lobes are also reconnected, reducing the amount of open flux in the system. Earthward of the new reconnection region new closed field lines collapse toward the Earth, ‘dipolarising’ the extended tail field lines, and heating and compressing the plasma sheet plasma. Precipitation from this hot plasma into the nightside ionosphere produces a patch of bright active auroras, called the ‘substorm auroral bulge’, and also considerably enhances the density of the nightside ionosphere. Consequently the height-integrated Hall and Pedersen conductivities in this ‘bulge’ are also strongly enhanced, from  $\sim 1 \text{ mho}$  or less in the pre-substorm nightside ionosphere, to values of  $\sim 10\text{--}100 \text{ mho}$ . This has the effect of suppressing the flow and electric field within the perturbed region, but even so intense currents flow within the region, consisting principally of a westward-directed Hall current or ‘substorm electrojet’. The total current flow is typically  $\sim 0.5\text{--}1 \text{ MA}$ , yielding magnetic perturbations on the ground underneath the electrojet of several hundred nT. The disturbance field is directed equatorward, weakening the horizontal component of the planetary dipole field in both hemispheres. Current continuity is maintained by FACs which flow from the tail plasma sheet into the ionospheric electrojet at its eastern end, and return from the ionosphere to the plasma sheet at its western end, as shown in Fig. 5, forming the ‘substorm current wedge’. The FACs then close in the cross-tail current on either side of the dipolarised region, and from

thence over the tail lobe magnetopause as shown. It can therefore be seen that the current wedge is associated with a reduction in the cross-tail current within the dipolarised region, the FACs being associated with the shears in the tail field that occur in the interface between the dipolarised field within the wedge, and the as yet undipolarised tail-like field outside.

As indicated above, substorm tail reconnection and field dipolarisation begin in a localised region near the centre of the tail (generally displaced to the dusk side of midnight), but then spreads in both directions across the tail. The bright auroral and electrojet current region in the ionosphere correspondingly starts as a small oval region in the pre-midnight sector, located typically at  $\sim 65^\circ$  magnetic latitude, and then spreads towards dusk and dawn, as well as poleward as flux continues to be reconnected and plasma compressed and heated in the nightside tail. This is called the ‘expansion phase’ of the substorm, and was first described by Akasofu in 1964. Sometimes the auroral bulge will expand to cover much of the nightside polar ionosphere, though it does not generally reach to the magnetic pole itself. After 20-30 min the auroral bulge reaches its maximum size, and the auroral intensity and the currents then start to decline, signalling the start of the substorm ‘recovery phase’ which typically lasts a further 30-40 min. In the tail, ‘recovery’ is associated with the down-tail propagation of the reconnection region to large distances from Earth, such that the plasma sheet re-forms in the near-Earth tail. On the ground underneath the bulge, where the magnetic effects are dominated by the overhead electrojet current, the magnetic depression in the horizontal field typically grows rapidly and impulsively during the expansion phase, and then declines somewhat more gradually during recovery, giving rise to a signature which is called a ‘magnetic bay’ in high-latitude magnetic records. Substorm signatures are also observed in nightside mid-latitude magnetic fields, but in this case both the ionospheric currents and the FACs of the current wedge contribute to the form of the overall disturbance.

Several substorm cycles often occur each day, driven by individual episodes of modest southward field in the IMF, producing magnetic disturbances in the polar region under the electrojet of  $\sim 100\text{--}1000$  nT which grow and decay over intervals of  $\sim 1$  hour. During rarer extended intervals of strong southward field, however, impulsive electrojet activity is essentially continuously present in the high-latitude nightside ionosphere, with variable magnetic disturbances on the ground underneath peaking at  $\sim 1000\text{--}2000$  nT. These high-latitude perturbations occur simultaneously with the world-wide storm-time field depressions of typically  $\sim 50\text{--}250$  nT which are produced under the same conditions, mainly by hot plasma inflow from the tail as discussed in the previous section. The largest high-latitude disturbances observed since systematic records began in 1957 have been  $\sim 3000$  nT, this type of disturbance thus being the largest produced at the Earth’s surface due to external currents, corresponding to  $\sim 5\%$  of the planetary field. At other times, however, the IMF may remain small and northward-pointing for extended intervals, leading to prolonged intervals of ‘magnetic quiet’ on the ground, when only the variations due to the

effects of the Sq-current system are present, driven by the daily thermal-tidal motions of the upper atmosphere.

## Figure Captions

Figure 1: Sketch of a cross-section through the Earth's magnetosphere in the noon-midnight meridian. The direction of the Sun is towards the left, such that the solar wind blows from left to right. The arrowed solid lines show magnetic field lines, while the heavy dashed lines show locations of principal discontinuities at the bow shock and magnetopause. The small symbols indicate the principal plasma populations present, while the arrows indicate their direction of flow. The dashes indicate the solar wind, and the plasma populations derived directly from it in the magnetosheath and magnetosphere (cusp and plasma mantle). The crosses indicate plasma originating from the Earth's ionosphere (plasmasphere and polar wind), while the dots indicate the heated plasma of mixed origin that originates in the tail (plasma sheet and ring current).

Figure 2: Sketch looking down onto the northern polar ionosphere, with noon (the direction towards the Sun) at the top of the diagram, dusk to the left, dawn to the right, and midnight at the bottom. The dashed line shows the boundary between open and closed magnetic field lines, while the arrowed solid lines shows the plasma streamlines of the Dungey-cycle flow. The short arrows marked  $E$  indicate the directions of the electric field associated with the flow,  $E = -V \times B$ , while the circled dots and crosses indicate regions of upward and downward field-aligned current (FAC) flow, respectively. The small symbols indicate principal plasma regimes into which the field lines map in the magnetosphere, with the same format as Fig. 1.

Figure 3: Schematic of newly-opened field lines on the dayside following low-latitude reconnection in the presence of an IMF with a positive  $B_y$  (directed toward the right) component. The view is from the direction of the Sun. Under these conditions magnetic tension pulls open field lines westward towards dawn (left) in the northern hemisphere, and simultaneously eastward towards dusk (right) in the southern hemisphere, as indicated.

Figure 4: Sketches of the ionospheric flow in the northern hemisphere ordered according to the direction of the IMF, in a format similar to Fig. 2. The dashed line shows the open-closed field line boundary, while the arrowed solid lines show the plasma streamlines. The upper row show flows for IMF  $B_z$  near zero and negative, while the lower row shows flows for IMF  $B_z$  significantly positive. The three columns show flows for IMF  $B_y$  negative (left), near zero (centre), and positive (right).

Figure 5: Sketch showing the form of the nightside 'current wedge' current system that is excited during the expansion phase of magnetospheric substorms. A portion of the plasma sheet current in the centre of

the tail is diverted along the field lines north and south towards the Earth, and closes in the ionospheric substorm westward electrojets.

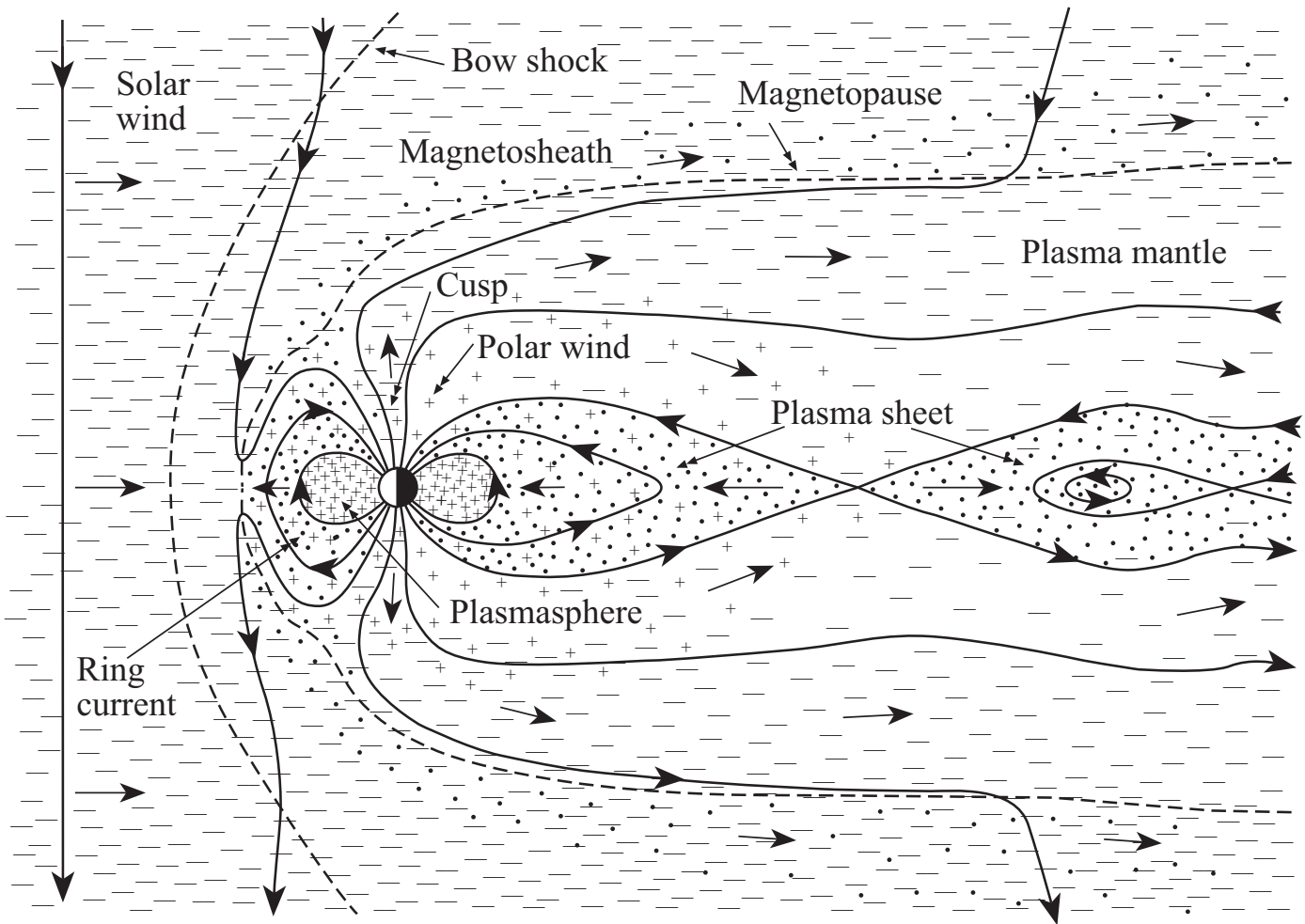


Figure 1

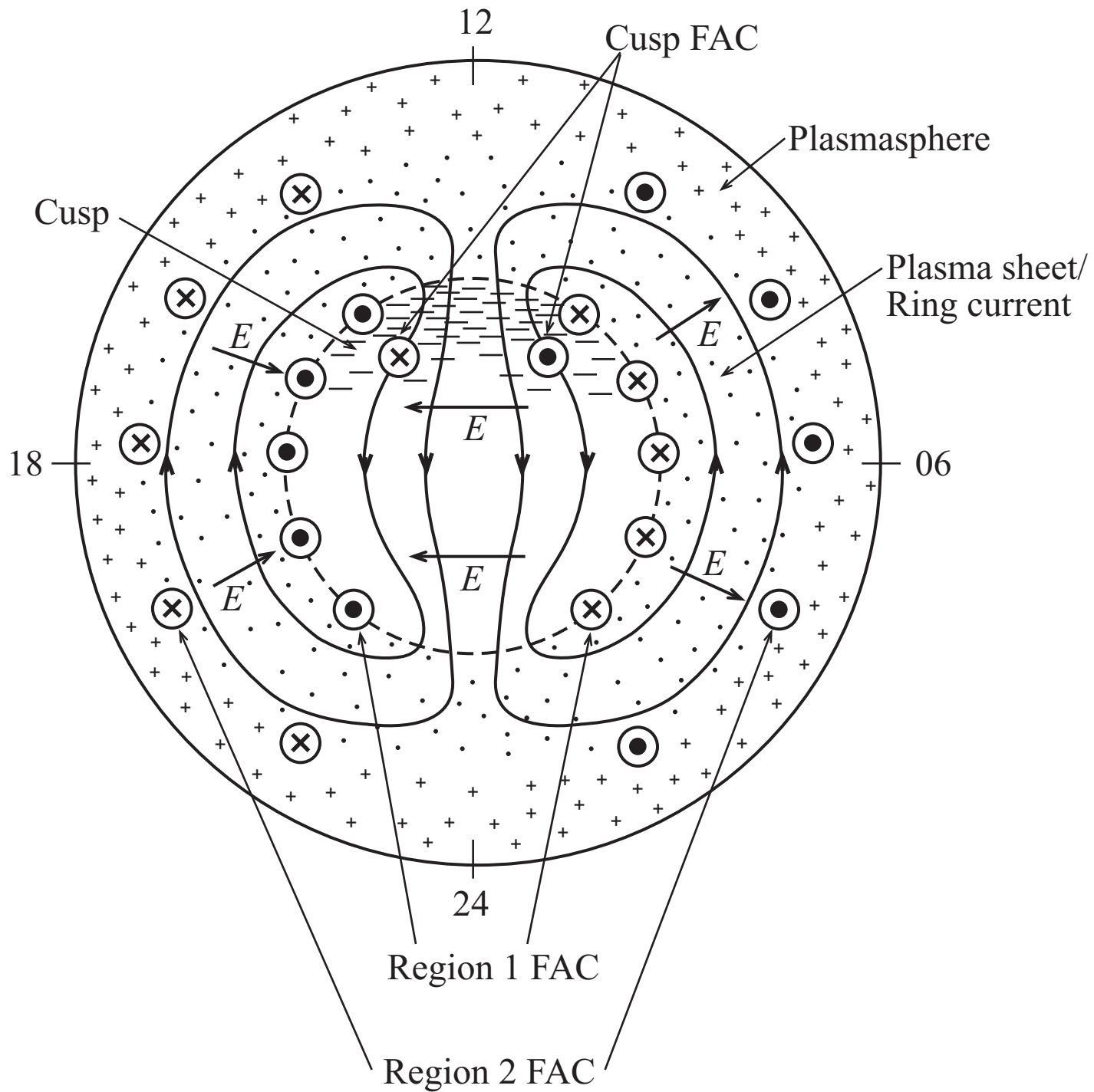


Figure 2

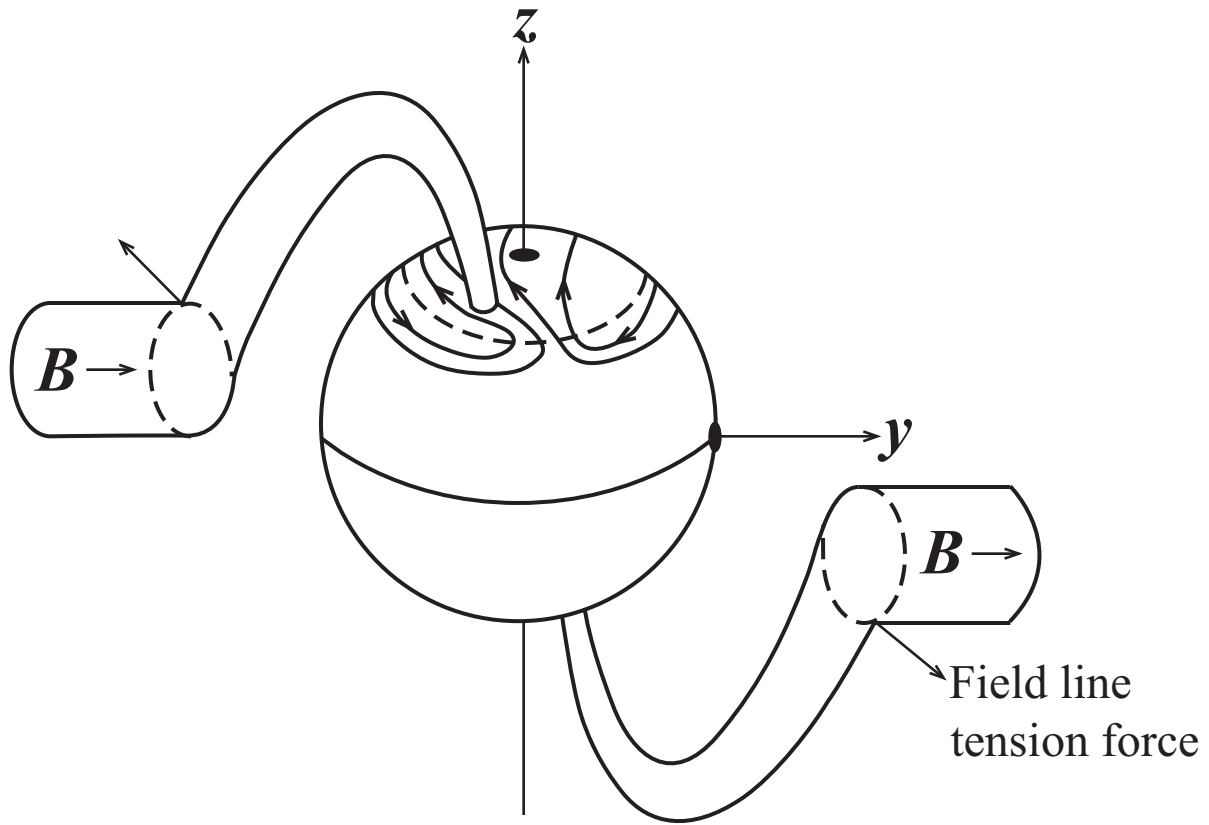


Figure 3

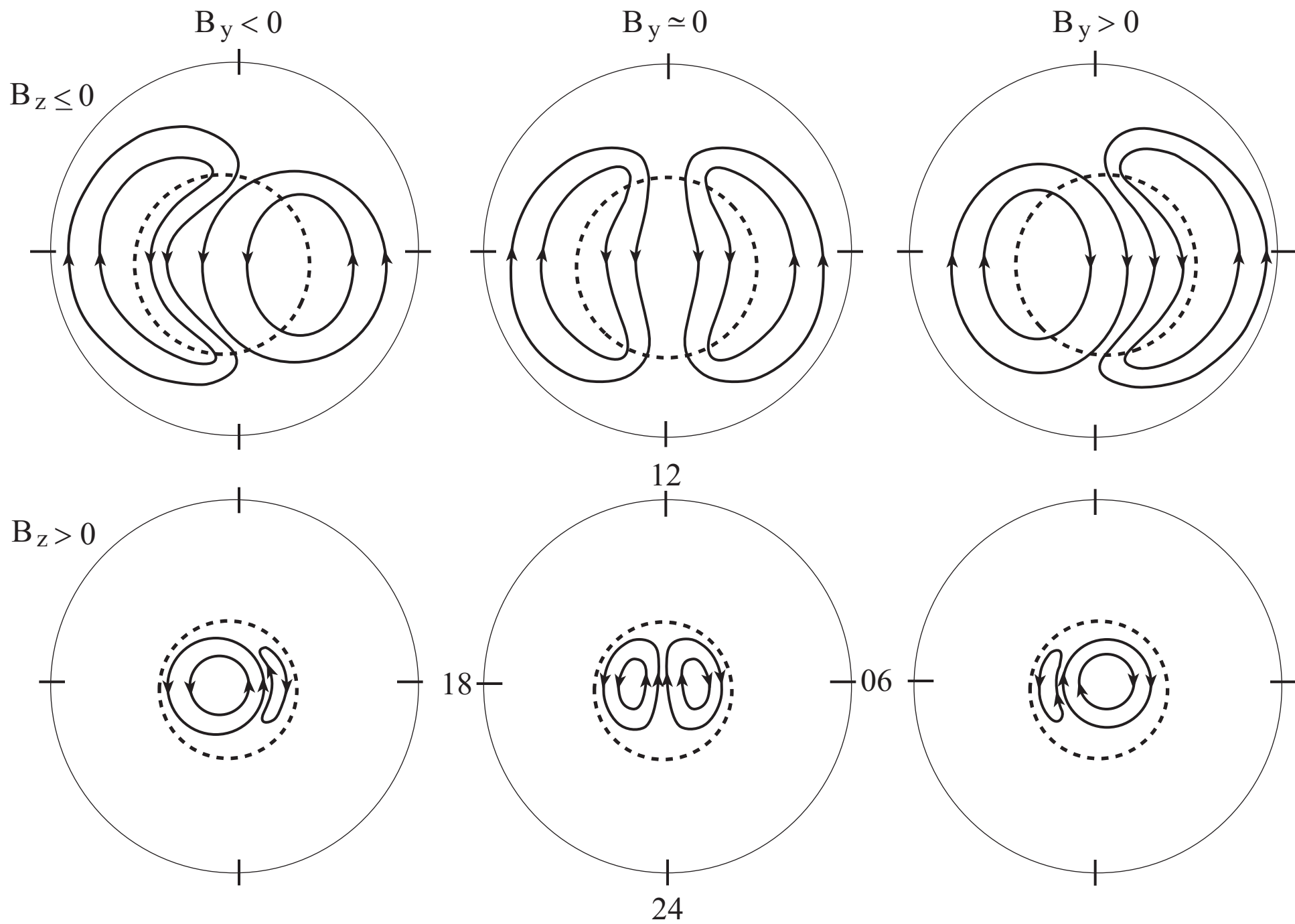


Figure 4

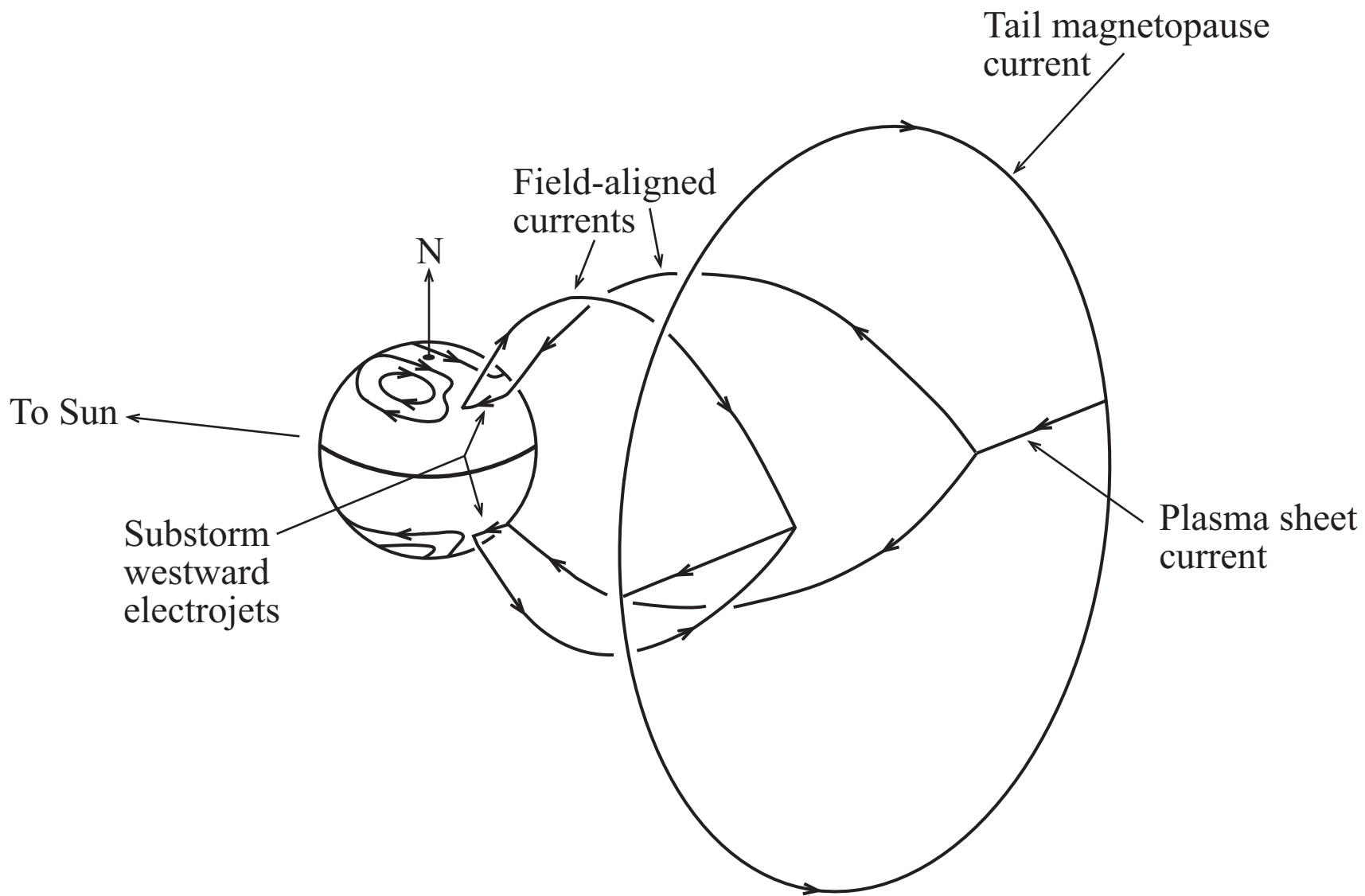


Figure 5