REVIEWS OF GEOPHYSICS AND SPACE PHYSICS, VOL. 21, NO. 1, PAGES 125–138, FEBRUARY 1983

The Origins of Birkeland Currents

DAVID P. STERN

Planetary Magnetospheres Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771

A qualitative explanation is proposed for the main features of quiet-time Birkeland currents. The major conclusions are as follows: (1) The source of region 1 currents is the interplanetary electric field, linked to the polar ionosphere directly along open field lines near noon but indirectly, via the plasma sheet, at most other longitudes. (2) Region 2 arises from convective charge separation, owing to guiding center drifts. On the dayside, secondary charge separation may occur, extending the current pattern sunward. (3) The magnetopause boundary layer flow is not a major energy source of $j_{\parallel}$, and neither are dynamo processes in convected plasmas or in the plasma sheet. (4) The convection reversal is expected to occur (as observed) in the interior of region 1, often near its poleward edge. (5) The cusp currents are associated with an inhomogeneity of the polar electric field, correlated with interplanetary B_y and first observed by Heppner, though contributions by the boundary layer are not ruled out. (6) The four-lobed polar electric field patterns occasionally observed during times of northward interplanetary B_z are generated by a boundary layer dynamo and signify the temporary existence of a closed magnetospheric configuration. (7) The branching ratio between two routes by which magnetospheric space charge may be neutralized, via the ionosphere and via polarization currents, is estimated, with and without considerations of particle mirroring.

OUTLINE

Currents flowing along magnetic field lines which lead to the polar and auroral regions constitute the major link between the ionosphere and the magnetosphere. They were first proposed by Birkeland in 1908 [*Chapman and Bartels,* 1940; *Boström,* 1968] but were only observed directly from space by *Zmuda and Armstrong* [1974] and independently by *Sugiura* [1975]. More recently, additional details of Birkeland currents (or $j_{\parallel}$, as they will be denoted here) have emerged, but theories about their origins have so far been incomplete.

This study will begin by reviewing fundamental observations and 'zero-order theories' which explain them. It then goes on to describe more detailed recent observations, and from them a 'first-order theory' will be derived, which may be useful for planning and interpreting future observations and for constructing quantitative models for $j_{\parallel}$. In what follows, five distinct types of quiet-time $j_{\parallel}$ are distinguished (Figure 1 [from *Iijima and Potemra,* 1976*b*]):

1. 'Region 1' consists of current sheets which approximately follow the auroral oval. The current flows into the ionosphere at dawn and out of it at dusk.
2. 'Region 2' adjoins the above on the equatorward side, with opposite polarities.
3. Cusp currents are located poleward of region 1 in the cusp region and are also of opposed polarities.
4. An altogether different pattern of currents is sometimes observed following extended periods of northward interplanetary B_z.
5. A three-layer overlap of current sheets apparently exists in the Harang discontinuity, near midnight.

The main conclusions of the study are as follows:

1. Region 1 currents near noon may be driven by the interplanetary electric field on open field lines; away from noon the source is probably the part of the plasma sheet closest to its inner edge. The reversal of **E** is expected to occur inside region 1, often close to its poleward edge. The magnetospheric boundary layer is not an important energy source for Birkeland currents, under ordinary circumstances, and neither are bulk motions in the inner magnetosphere or in the plasma sheet, directed in the $\pm y$ direction.
2. Region 2 currents are driven by 'charge separation,' a modified version of the mechanisms proposed by *Schield et al.* [1969]. Such separated charges can be neutralized either through the ionosphere or by means of a polarization current in the magnetosphere: it is shown that the first channel is the dominant one, so that the assumptions of Schield et al. are justified. Where the current flow is completed by ionospheric particles drawn upward, it is shown that the particles' energy source can be ultimately traced to the solar wind. It is also argued that if particles of the partial ring current are energized by convection, then charge separation and region 2 currents are part of the closure of that current. Furthermore, the charge separation effect may be greatly enhanced by impulsive particle injection events. Finally, it is shown that the precipitation of separated charges creates a further imbalance, leading to further sunward extension of region 2 currents.
3. A necessary consequence of the asymmetry of polar **E**, related to interplanetary B_y and observed by *Heppner* [1972], is a B_y-related asymmetry in $j_{\parallel}$ [*Vasyliunas,* 1979]. This asymmetry can be explained in terms of the open magnetosphere and could explain the B_y-related asymmetry of cusp currents, observed by *Doyle et al.* [1981].
4. The four-lobed electric field and convection pattern claimed by *Burke et al.* [1979] and by *Horwitz and Akasofu* [1979] for times of northward interplanetary B_z can be explained if the magnetosphere is assumed to be at such times temporarily closed. This explanation requires the magnetopause to be relatively inefficient in transmitting the flow of electric charges between the dawn flank and the dusk flank of the magnetosphere.
5. The explanation of the current flow near midnight remains unclear. It appears that the middle layer there represents the 'Harang filament,' a continuation of the electric currents carried by the auroral electrojets. Such a filament has been proposed by *Rostoker* [1974], and its

Paper number 2R1661.

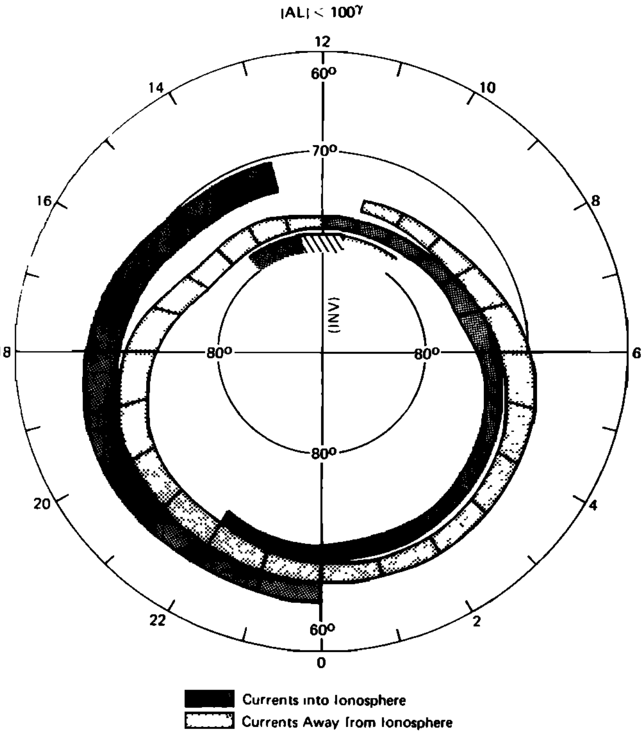

Fig. 1. Flow of quiet-time Birkeland currents into and out of the polar ionosphere according to *Iijima and Potemra* [1976*b*].

circuit appears to be completed by the near-earth plasma sheet. New evidence supporting this idea will be cited, but no explanation will be attempted.

Dynamo Action: Overview

There exist two main classes of sources which can drive $j_{\parallel}$. Dynamo action is the general name given to the emf developed in a magnetic field around a circuit, part of which is in motion relative to the rest. A simple example, for an open magnetosphere, is the circuit consisting of two field lines, the ionosphere connecting their 'feet' and the solar wind flow connecting their tops [*Stern*, 1977, Figure 18]. Viewed from the earth's frame, the moving medium is the solar wind, and the current flow in it is sustained by polarization drifts [*Stern*, 1973, appendix]. Dynamo action can also take place in a closed magnetosphere, where particles are convected nightward near the boundary and return sunward closer to the noon-midnight plane (Figure 2). Both nightward and sunward flows can in principle produce dynamo action and can drive Birkeland currents. The other type of source for $j_{\parallel}$ is charge separation, described below under that heading.

It may be added here that dynamo action also produces intense currents connecting the corotating magnetosphere of Jupiter with the satellite Io. For the record, Faraday speculated [*Williams*, 1965, pp. 207–208] that the emf responsible for the polar aurora was generated by flow of the Gulf Stream across the geomagnetic field: his theory fails in numerous ways, but his concept of dynamo action was a valid one.

The Boundary Layer Dynamo

It has been proposed [*Eastman et al.*, 1976; *Sonnerup*, 1980; *Bythrow et al.*, 1981] that the boundary layer (BL) of the magnetosphere generates $j_{\parallel}$ by dynamo action. A schematic drawing of the process, applicable to a closed magnetosphere (further below) but also suitable for discussing BL dynamos, appears in Figure 2. The boundary flow passes here across AB and A'B', and its return flow (plus any return flow of reconnected magnetic flux) crosses BC and B'C'. Currents are expected to flow in the directions appropriate to region 1 near (B, B') and in those appropriate to region 2 near (C, C').

If the circuit of $j_{\parallel}$ is driven by the same voltage drop as is observed across the polar caps, then the energy input required is of the order of $1–3 \times 10^{11}$ W [*Nisbet et al.*, 1978; *Stern*, 1980]. The BL in contrast can provide only about one twentieth of this, for assuming that the layer contains plasma which has somehow breached the magnetopause [*Lemaire*, 1979] and that its entry front on each flank is 0.5 R_E thick and 10 R_E wide, with $n = 2.5$ cm^{-3} and $v = 200$ km/s, we find a total energy input of 6×10^9 W. Viscouslike interaction could perhaps supply additional energy, but it is more likely [*Vasyliunas*, 1979] that any BL source is only a minor addition to the main source of $j_{\parallel}$, which draws energy from the solar wind via open polar field lines.

A similar factor was also obtained by *Sonnerup* [1980, p. 2020], who noted for one of the models derived there that the voltage drop across the BL (~3 kV on each side) was only 10–15% of what has been observed across the polar cap, leaving the rest to be explained by open field lines; furthermore, the process operates only on the dayside [*Sonnerup*, 1980, Section 8, p. 2025]. It should be noted, however, that the properties of the BL in these models differ from those listed earlier. In Sonnerup's models the BL is thin and is driven by viscouslike forces, which may be important under special circumstances as noted further below.

In all this a clear distinction should be made between the energy source driving region 1 currents and the source of the particles which carry these currents. It is quite possible that an appreciable fraction of these particles originates in the boundary layer. However, the energy driving the circuit

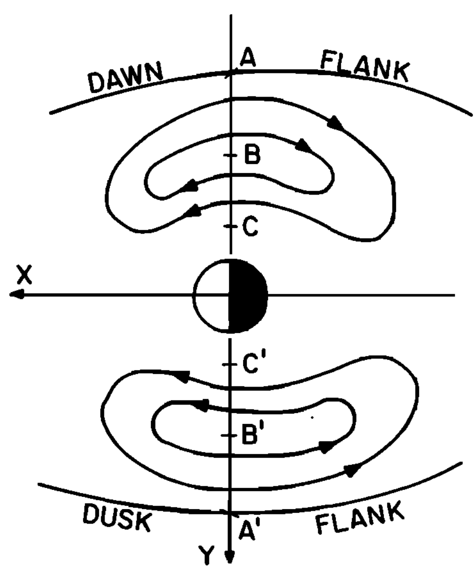

Fig. 2. Schematic view (not to scale) of convection near the equatorial dawn-dusk cross section of a closed magnetosphere. The flow lines are also electric equipotentials, and solar magnetospheric coordinates are used.

comes primarily from the large voltage drop across open field lines, with only a minor addition generated across the boundary layer.

What about the currents from (A, A′)? *Eastman et al.* [1976] identified them with observed cusp currents, and this was supported by *Vasyliunas* [1979], who noted that this required that 'the mapped boundary layer flow is faster than the polar cap flow.'

If one inserts observed magnitudes, the condition is indeed fulfilled. Estimating the magnetic field in the low-latitude BL as 30 nT gives the observed layer (using figures cited earlier) a voltage drop of 20 kV. The electric field of the BL, as mapped into the ionosphere, is then much stronger than the polar field on open field lines. Physically, the reason is that the latter field is reduced, because open field lines are stretched into the tail, to a value ~15 times below what might be expected in the absence of such stretching [*Stern,* 1973].

The cusp currents could thus be generated as suggested by *Vasyliunas* [1979], and their intensity would (as observed) rapidly decline away from the noon meridian because of the limited energy available. Two questions remain. First, the boundary is not observed to undergo appreciable deceleration, and hence only a part of its energy (at most) is lost; this might be explained if the flow is partially decoupled from the ionosphere [*Stern,* 1981, section B4*a*]. Second, recently observed correlations between cusp currents and interplanetary B_y [*Doyle et al.,* 1981] suggest an altogether different explanation, discussed further below. Some observers (T. Potemra, private communication, 1982) claim that both effects occur in superposition, that is, that part of the cusp current is permanent and symmetric while another part is related to B_y; further studies should clear up this point.

In conclusion, then, BL dynamo action may perhaps be a source of cusp currents, but (as suggested by *Vasyliunas* [1979] and *Sonnerup* [1980]) it plays only a minor part in driving the currents of regions 1 and 2 under ordinary circumstances.

Other Dynamos

It has been proposed that Birkeland currents are driven by a dynamo process on closed field lines, due to plasma convected sunward from the plasma sheet [*Akasofu et al.,* 1981, Figure 8; *Akasofu,* 1982, Figure 9]. This would be quite similar to the dynamo action in the return flow of the BL as described in the previous section, taking place across BC and B′C′ of Figure 2 and generating currents of both region 1 and region 2.

Again, the process is possible, but the energy available from it is even less than before, because quiet-time convection velocities are small compared to those of the BL. Let the plasma be thrust into the near-earth magnetosphere from the inner edge of the plasma sheet, after being set in motion in the plasma sheet itself. Furthermore, let the width of the sheet be 10^5 km, its magnetic field 20 nT, and the potential drop across it 50,000 V: then the plasma starts at 25 km/s. Assuming furthermore an extent of 30,000 km in z and a density $n = 0.5$ cm^{-3}, the energy input is 2×10^7 W, too small by several orders.

It should be noted that the rate at which energy is injected from the plasma sheet into the inner magnetosphere is about 5000 times larger [*Stern,* 1980]. However, almost all this amount is contained in the individual energy of particles, while the dynamo process can tap only the relatively small amount of energy contained in the plasma's bulk motion.

A different dynamo process has been advocated by *Rostoker and Boström* [1976], *Hughes and Rostoker* [1979], and *Rostoker and Hughes* [1979] and involves bulk motions in the plasma sheet, parallel and antiparallel to the dawn-dusk axis (y axis) of the magnetosphere. Such tail flows have not been reliably observed [*Frank and Ackerson,* 1979, Figure 5] and would seem hard to sustain. Furthermore, the total kinetic energy of plasma sheet motions, from which such dynamo action is derived, is again small compared to the requirement of $j_{\parallel}$. Assuming for the flows $v = 20$ km/s, the kinetic energy of the plasma sheet is of the order of 5×10^{11} J, equivalent to what is delivered by $j_{\parallel}$ in 2–4 s. Thus a mechanism is needed that pumps energy into this particular mode at an equal rate and very steadily.

In both cases, if the magnetosphere is open, the solar wind dynamo and the charge separation mechanism, which have been long accepted by a large part of the science community [e.g., *Vasyliunas,* 1979], are also expected to exist. Their effects will then mask those of the much weaker dynamos described here.

CHARGE SEPARATION: OVERVIEW

The second main type of mechanism producing Birkeland currents is charge separation, the disruption of charge neutrality by magnetic drifts (gradient and curvature), which move charges of opposite signs in opposing directions. As plasma flows earthward from the tail of the magnetosphere [e.g., *Stern,* 1977], to the lowest approximation its magnetic drifts are negligible, so that both ions and electrons share the same electric drift

$$\mathbf{u} = \mathbf{E} \times \mathbf{B}/B^2 \qquad (1)$$

and neutrality is preserved. A better approximation, however, recognizes that magnetic drifts are also present, although they are slow, since they are roughly proportional to the energy W. Such magnetic drifts (unlike $\mathbf{u}$) move particles across electric equipotentials and increase their W [*Hines,* 1963] until, for those orbits which penetrate closest to the earth, magnetic drifts do become comparable to $\mathbf{u}$. In practice, this growth of W is most easily derived by using the adiabatic invariance of μ and J.

Here a simple approximation of this motion will be used, in which corotation is neglected and equatorial particles (with W proportional to B and $J = 0$) are taken to represent in their motion all convected plasma particles (a somewhat similar approximation was used by *Alfvén* [1939]). Furthermore, the argument will be simplifed by adding the assumption that all convected particles of either sign have come from the geomagnetic tail, where they have belonged to some standard initial energy distribution, at the time when they passed the (equatorial) line with $B = B_0$, where B_0 is some appropriately chosen initial intensity.

As such particles approach the earth, magnetic drifts tend to deflect ions toward dusk and electrons toward dawn, so that most ions pass on the duskside and most electrons on the dawnside. This suggests a surplus of ions at dusk and of electrons at dawn: such a separation was suggested by *Alfvén* [1939] (see also *Karlson* [1962, 1963] and *Helmer* [1963]), but it was first put in a form applicable to observations by *Schield et al.* [1969]. The latter authors pointed out that unhindered sunward drift of particles of any single energy produces 'Alfvén layers' accessible only to particles of one sign: on the duskside a layer containing only ions and on the dawnside one with only electrons. With the assumptions used here, outside these layers the density at the given energy balances, because both ions and electrons in the range (W, dW) have started from the same range (W_0, dW_0) at $B = B_0$ and had the same density there.

Of course, each range (W_0, dW_0) tends to produce different Alfvén layers, but in all cases there exists an excess of positive charge near dusk and of negative charge near dawn. Since the plasma will not tolerate any appreciable space charge, it has been postulated that such excesses will be either precipitated or neutralized: in either case a current flows to the ionosphere. The validity of the above assumption and the possibility that neither of these two mechanisms is rapid enough to prevent the buildup of appreciable space charge will be discussed separately later. The sense of the currents agrees with that of region 2, and therefore the above may be viewed as a zero-order explanation of such currents.

In summary, then, it will be assumed here that $j_\parallel$ in region 1 is produced by the source of polar **E**, namely, by dynamo action on open field lines. As noted further below, on the nightside the linkage of $j_\parallel$ to that source probably proceeds through the inner plasma sheet. Region 2, on the other hand, is caused by charge separation in plasma convected sunward from the tail.

Neither of these zero-order mechanisms is new [*Vasyliunas,* 1979] (see also *Schield et al.* [1969] and *Stern* [1977]). Alternative dynamo processes have been proposed, taking place either in the BL, in plasma convection sunward, or in the tail. However, it was shown here that none of those alternative sources can supply the rather considerable energy requirements of the main Birkeland current system. Thus if they exist, their role is secondary.

Below, various details of the flow of $j_\parallel$ will be studied, including the expected structure of $E_\parallel$, the role of polarization currents, secondary charge separation processes, four-lobed polar electric fields, and asymmetries correlated with interplanetary B_y. Before all that, however, the main pertinent observations will be listed.

OBSERVATIONS

(The list that follows does not necessarily cite the initial discovery, and omissions may merely reflect gaps in the writer's awareness.)

1. A thorough survey of the average properties of $j_\parallel$ by *Iijima and Potemra* [1976*a*, *b*] and *Potemra* [1979] (see also *Saflekos et al.* [1982, Figure 16]) revealed the following (Figure 1): (1) On the average, the total region 1 current in each flank of each polar cap is about 1.5×10^6 A at quiet times and 2.7×10^6 A at disturbed times. Region 2 currents have typically 75% of the intensity of region 1 currents. (2) The current density of $j_\parallel$ in either direction is of the order of 1 μA/m^2, and current sheets are typically 2.5° wide in latitude. Regions 1 and 2 generally (though not always) adjoin each other. (3) Region 1 currents are most intense on the dayside, at MLT 14–16 and 6–9, with a decline toward noon and a broad minimum around the Harang discontinuity (MLT ~ 22), where the sheets overlap. (4) Region 2 currents are strongest near midnight and gradually decline toward noon. (5) Both current systems intensify with increased magnetic activity. However, at very quiet times, region 2 currents may become extremely weak, while region 1 tends to 'bottom out' at a finite level. (6) There exist cusp currents paralleling region 1 on the poleward side, with flow directions opposing those of the adjacent region 1 sheets. (7) There exists a triple overlap of sheets in the Harang discontinuity region, with an outward current flanked by two earthward sheets.

2. The preceding results support the hypothesis that the auroral electrojets [*Rostoker,* 1980] are due mainly to the Hall currents associated with the ionospheric Pedersen current connecting regions 1 and 2. Because contributions to ground magnetic disturbances from $j_\parallel$ tend to cancel those arising from the related ionospheric Pedersen currents [*Fukushima,* 1969, 1976], the ground magnetic signature of the current system involving $j_\parallel$ is believed to be mostly due to the electrojets. What was not realized before $j_\parallel$ was observed from space was that an appreciable Birkeland current (and therefore also, presumably, appreciable electrojets) existed even at quiet times. The magnetic signature of these electrojets has now been observed from space [*Iijima et al.,* 1982], and it indeed peaks in the transition between regions 1 and 2.

3. The outflowing middle sheet in the Harang discontinuity does not seem to belong to either of the outflowing sheets adjoining it, but rather to be a current filament issuing from the confluence of the auroral electrojets, carrying at least part of their current into space. In what follows, such a current will be referred to as the Harang filament (HF). Such a filament was deduced from observations, for substorm periods, by *Kamide and Rostoker* [1977]. Even earlier, *Rostoker* [1974] suggested that a shift of the HF during substorms would produce the magnetic signature commonly associated with the diversion of the cross-tail current through the ionosphere, which has been proposed as occurring during substorms [*McPherron et al.,* 1973]. Indeed, a very simple form of the idea occurs in the model of *Alfvén* [1939].

A recent observation which supports this notion is one by *Inhester et al.* [1981], who deduced a strong upward current (i.e., downward electrons) from the leading edge of the westward traveling surge of a substorm. If the HF moves with the surge, it would have properties like those predicted by *Rostoker* [1974]. *Shuman et al.* [1981] have also claimed, from observations of the S3-2 spacecraft, that an intense upward current existed just poleward of the Harang discontinuity.

4. The location of the reversal of **E** at the edge of the polar cap is often viewed as the boundary of open field lines and was expected to occur near the boundary between regions 1 and 2. Instead, recent comparisons of **E** and $j_\parallel$ found that the reversal occurred near the poleward edge of region 1, barely inside it. Such observations were performed by S3-2 [*Smiddy et al.,* 1980], which carried electric probes, and also by AE-C [*Bythrow et al.,* 1981], which measured plasma flow, from which **E** could be deduced.

5. A seasonal variation of $j_\parallel$ was observed by *Sugiura and Potemra* [1978] and was confirmed by *Fujii et al.* [1981], with the largest currents flowing in summer. The latter authors noted that the summer/winter ratio was largest (about 2) at times when region 2 currents were absent and only region 1 was observed. The explanation for the seasonal dependence was that in winter the polar cap is dark, its ionospheric conductivity is low, and therefore the current flowing across it between the two sheets of region 1 is small.

The above explanation seems reasonable, but two comments may be appropriate. First, one notes that closer to equinox the dayside of both polar caps is sunlit, while the nightside is dark, and a larger current may thus flow across the dayside portions of the caps. Thus the dependence of region 1 currents on magnetic local time [*Iijima and Potemra*, 1976*a*, Figure 3*a*] may be enhanced by this effect or perhaps may even be largely caused by it.

Second, some of the current linking the region 1 sheets may be conducted through the auroral zone, where particle precipitation increases electrical conductivity [*Boström*, 1964; *Yasuhara and Akasofu*, 1977]. This linkage will add to the intensity of the westward electrojet and diminish that of the eastward one, in agreement with observations [*Rostoker*, 1980].

6. Among the many observations which have confirmed the existence of the cusp currents (item 6 in paragraph 1 above), perhaps the most clear-cut result is a correlation with the interplanetary B_y component, obtained by *Doyle et al.* [1981]. In 13 polar passes of S3-2 they found invariably that the cusp current was present on that side of noon from which open field lines approached the polar cap (e.g., with $B_y > 0$, northern dawn and southern dusk) and was absent on the opposite side.

7. As already noted, the pattern of polar currents is strongly related to that of the polar electric field, which is predominantly two-lobed [e.g., *Heppner*, 1977], with **E** from dawn to dusk across the polar cap (Figure 3*a*). However, it was found by *Burke et al.* [1979] that at times of northward interplanetary B_z the polar electric field may sometimes assume a four-lobed pattern, with sunward convection near the pole, flanked by lobes of tailward convection on the duskside and dawnside (Figure 3*b*). This pattern was seen in about 15% of the passes with $B_z > 0$, as confirmed by AE-C [*Spiro et al.*, 1979]. A similar result was claimed by *Horwitz and Akasofu* [1979] using ground magnetic data from the northern polar cap, taken mostly in July and August, when the cap is sunlit and **E** there is expected to drive appreciable currents.

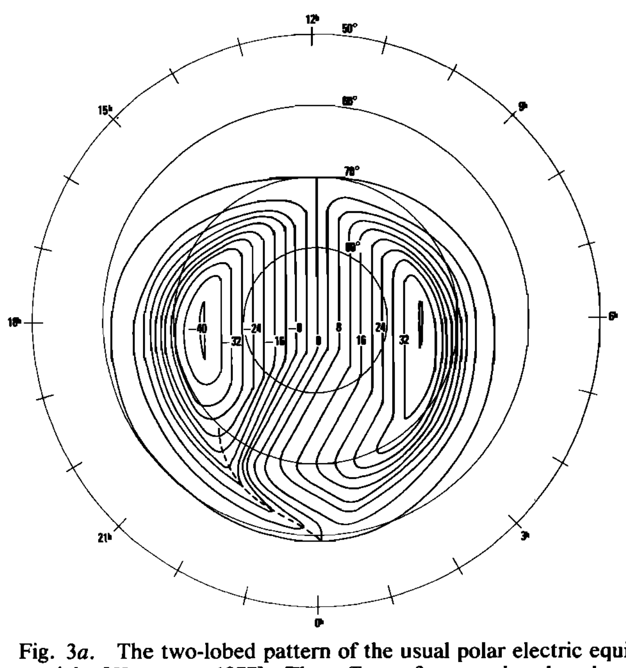

Fig. 3*a*. The two-lobed pattern of the usual polar electric equipotentials [*Heppner*, 1977]. The effect of corotation has been subtracted.

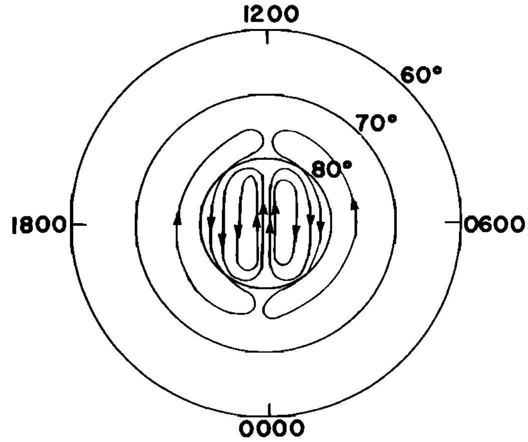

Fig. 3*b*. Four-lobed pattern of polar electric equipotentials associated with northward interplanetary B_z [*Burke et al.*, 1979].

DETAILED ANALYSIS

Region 1: The Lyons Diagram

One of the remarkable properties of field-aligned currents is their typical width of about 300 km. In contrast, the conducting layer in the ionosphere, which feeds them, is narrower than 50 km [*Boström*, 1964]. If the parallel conductivity σ_0 in the magnetosphere is well-nigh infinite, as simple ionospheric theory has it [c.f. *Boström*, 1964, Figure 3], the narrowest sheet (~1 gyroradius) would have sufficed to carry $j_{\parallel}$. In fact, observations (e.g., those of *Iijima and Potemra* [1976*a*]) suggest that on the average, $j_{\parallel}$ is limited to the order of 1 $\mu A/m^2$, though narrow spikes of much larger current density have also been noted [*Hardy et al.*, 1979].

Lyons [1980] suggested that the limitation on upward flowing $j_{\parallel}$ came from the magnetic mirror geometry by a process previously studied by *Knight* [1973] and also by *Alfvén and Fälthammar* [1963], *Persson* [1963], and *Chiu and Schulz* [1978]. Even in the presence of an appreciable $E_{\parallel}$, this restricts the upward current to a small multiple of what loss cone electrons carry when $E_{\parallel} = 0$. The process was further studied by *Stern* [1981], who noted that downward $j_{\parallel}$, which is observed to be comparable to upward $j_{\parallel}$, is similarly restricted by the ambipolar electric field in the ionosphere. The argument of all the above theories is that σ_0 is not appropriate here, since it is a local quantity and its use ignores the mirror geometry in which the flow occurs. Instead, it is argued that current-carrying particles (predominantly electrons) conserve their magnetic moment to the lowest approximation, so that the effect of $E_{\parallel}$ cannot be localized and is distributed over the entire field line.

Because $j_{\parallel}$ is impeded, there will exist a mismatch between the electric potential Φ in the source region and its value Φ_i in the ionosphere. In particular, Φ_i finds it difficult to follow Φ in the region where **E** reverses its direction. Lyons illustrated this by graphing Φ and Φ_i against the distance x measured across the sheet (his notation, not related to the geomagnetic x coordinate), and the result will be referred to here as a Lyons diagram. In Figure 4*a* (taken from Figure 2 of *Lyons* [1980]), **E** in the source regions is assumed to have constant values, so that Φ follows straight lines. The difference $\Phi - \Phi_i$ then represents the parallel

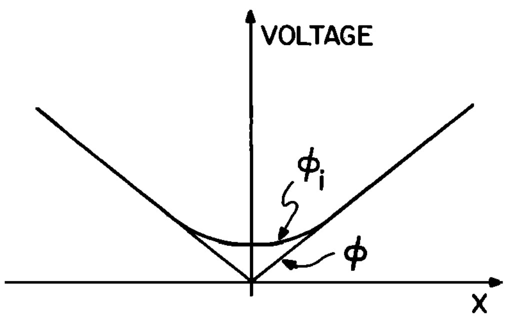

Fig. 4*a*. Voltage-distance relationship as proposed by *Lyons* [1980].

voltage drop, and the width of the region where it is appreciable is indeed comparable to that of the main $j_{\parallel}$ sheets. Lyons explained the existence of narrower structures, for example, auroral rays, by postulating narrower source regions [*Lyons*, 1981].

However, Figure 4*a* is generally inconsistent with observations, since it only involves potential drops (and presumably currents) of one direction. A modified Lyons diagram with currents in both directions has been proposed by *Chiu et al.* [1981]. Their configuration predicts three current sheets, has $E_{\parallel}$ reversed in the center, and corresponds to none of the situations discussed in the present study. In place of both these configurations we propose here the one shown schematically (for the duskside ionosphere) in Figure 4*b*.

In both Figure 4*a* and Figure 4*b* the origin corresponds to the reversal of the source voltage, presumably (in an open magnetosphere) to the 'last open field line.' However, this 'source voltage' should be treated with caution. On open field lines ($x > 0$) the electric field indeed arises from a voltage source, and the currents flowing into the ionosphere and out of it are determined by the ability of the ionosphere to carry the current driven by this voltage. This view is supported by the seasonal variations of $j_{\parallel}$. A different (and perhaps higher) limit on the current may be imposed by the ability of the magnetosphere to carry field-aligned currents, but it appears unlikely that the solar wind source would ever become overloaded.

On closed field lines, on the other hand, the potential Φ is determined indirectly by the physical processes which govern the flow. If closed lines were completely passive electrically, their potential Φ would be determined by the fringing pattern of the polar electric field (itself determined not just by the ionosphere but also by the ring current). In that case no $j_{\parallel}$ is expected. The present study, however, assumes that a net source of unbalanced charge does exist on closed field lines, due to convective charge separation, and that this charge is neutralized by means of $j_{\parallel}$. Thus for $x < 0$ the magnetospheric generator may be viewed as a current source, and since a parallel current apparently requires a voltage drop [*Stern*, 1981, and references therein], Φ has to adjust itself to match $j_{\parallel}$.

It may be noted in this connection that the mere fact that charged particles are precipitated does not necessarily imply the existence of $j_{\parallel}$. If, for instance, magnetically confined electrons are scattered into the loss cone by plasma waves [*Kennel and Petschek*, 1966; *Lyons*, 1979], their precipitation will leave the plasma from which they came positively charged, causing it to draw electrons from the ionosphere. In a steady state (which ought to be quickly reached) the flux of electrons in each direction should balance, so that the net result is merely the replacement of energetic electrons by less energetic ones, with no attendant $j_{\parallel}$.

In Figure 4*b* the slope of Φ is steeper for $x < 0$, since the slope of Φ_i is observed to be steeper in that region. The reversal of **E** occurs at the lowest value of Φ_i and is marked R: it is well inside region 1. Schematically, region 1 now consists of three parts: part AB, on closed field lines; part BR, on open field lines equatorward of the reversal; and part RC, poleward of the reversal. We first analyze the latter two.

If the current flow is viewed in profile (Figure 5) and if, as a simplification, the ionosphere is viewed as an ohmic conductor (neglecting Σ_H), it becomes evident at once that ordinarily the reversal of $\mathbf{E}_{\perp}$ in the ionosphere must be somewhere in the interior of the sheet, for region 1 drains away currents from both region 2 and the cross-polar flow. If the two currents are equal, each will extend across half the thickness of the sheet and will be associated with an ionospheric electric field in the direction of its flow, and the reversal of this **E** then occurs in the middle. During polar night the cross-polar current diminishes, and the reversal shifts toward the poleward boundary of region 1, while during very quiet times in summer, region 2 may dry up, and the reversal then shifts equatorward. These correlations can, in principle, be tested experimentally.

A similar division of region 1 between two directions of $\mathbf{E}_{\perp}$ is also found in Figure 1 of *Sonnerup* [1980]. There, however, the reason is that region 1 is produced jointly by two dynamo processes, both of which, as noted earlier, are somewhat weak. Here it is proposed that the effect is a direct result of the existence of $E_{\parallel}$, which produces a mismatch between $\mathbf{E}_{\perp}$ at high altitudes and at low ones. At high altitudes (except for the narrow region AB in Figure 4*b*, which represents a further elaboration (see discussion below)), $\mathbf{E}_{\perp}$ has the same direction throughout the cross section of a region 1 current sheet, unlike $\mathbf{E}_{\perp}$ of *Sonnerup* [1980]. At low altitudes, on the other hand, two opposing directions of $\mathbf{E}_{\perp}$ divide the width of the sheet as drawn in Figure 5.

Section AB represents flow from the ionosphere to the outermost ring current and may be viewed as a region in which the outer part of the ring current takes back a fraction of the excess positive charge dumped into the ionosphere by

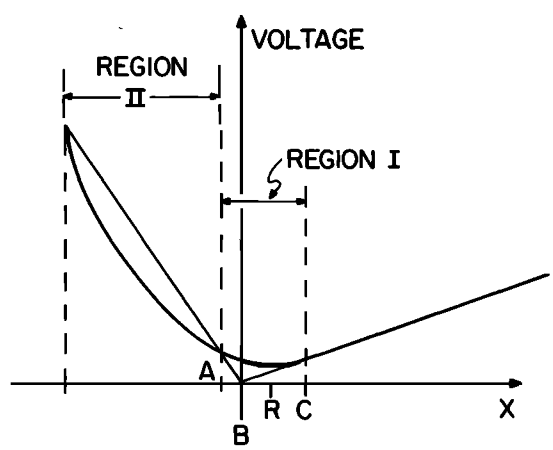

Fig. 4*b*. Voltage-distance relationship for twin sheets of Birkeland currents as proposed here.

its inner part. Most of that current flow is completed by a polarization current (appendix equation (A8)). A small fraction, however, goes to raise the local electric potential and increase the sunward convection velocity near the boundary. Energy is invested by this process as extra kinetic energy of convecting plasma and may be later withdrawn again through the dynamo process. All that, however, is a rather minor effect, since AB is small.

The Plasma Sheet

At local times near noon it is plausible that region 1 currents originate on open field lines. While such a source is consistent with the open model, it bears noting that experimental evidence on this matter is far from clear cut, as evidenced by Figure 6 of *Saflekos et al.* [1982], which on a few occasions places much of these currents equatorward of the boundary of trapped particles. Away from noon, however, it seems more likely that they first flow to the inner plasma sheet, then follow the plasma sheet to the flanks of the tail, and only from there head to their interplanetary source (Figure 6). The reasons are as follows.

In models of the open magnetosphere, closed field lines may belong either to the inner magnetosphere or to the tail. The former lines contain the stably trapped particles of the ring current, and the flow lines of currents associated with them circle the earth. The latter ones are threaded by a current which crosses the plasma sheet from dawn to dusk.

The magnetic flux embraced by the second group is readily estimated. Assuming that $B_z = 1$ nT for $(-x) = 20$–70 R_E [cf. *Behannon,* 1970] and estimating the width of the sheet at 40 R_E gives 2000 nT $R_E{}^2$. In the range $(-x) = 10$–20 the B_z component is much stronger, and we estimate that the additional flux from this range about equals the preceding value. This means that the polar area threaded by the ionospheric 'footprint' of the plasma sheet is about 220 deg^2, which comes to about one fifth of the area of the polar cap.

Now the boundary of region 2 seems to follow fairly closely the boundary of trapped electrons described by *Burrows and McDiarmid* [1972], which may be viewed as the boundary between the two groups of field lines. Since

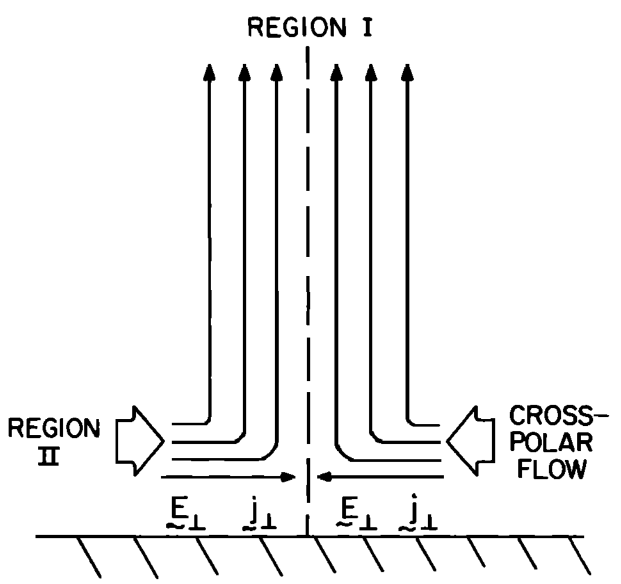

Fig. 5. The convergence of $j_\perp$ (and hence of $E_\perp$) in the middle of region 1. Hall currents are neglected here.

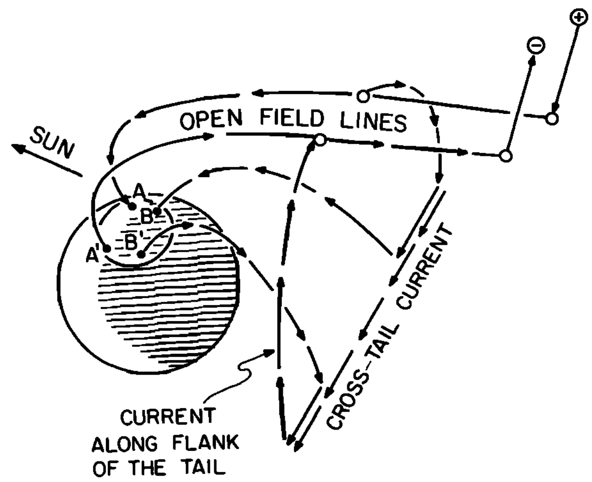

Fig. 6. The way in which open field lines supply Birkeland currents to points (A, A') on the dayside of the polar cap and to points (B, B') on its nightside. Arrows denote the directions of currents and are not related to the direction of the magnetic field.

regions 1 and 2 are generally contiguous, this suggests that on the nightside the currents of region 1 must originate in the plasma sheet: if they came from open field lines, a gap should exist between the two regions, with an area of the order of 200 deg^2.

This conclusion also appears to be plausible on physical grounds. Open field lines starting on the nightside of the polar cap stretch for a long distance, perhaps 200 R_E or more, before reaching the solar wind, and they extend through the rarefied high-latitude lobes. It is quite likely that in that case the electric field in the solar wind becomes essentially decoupled from the ionosphere [*Stern,* 1981, section B4*a*]. There also exists, however, an electric field **E** across the plasma sheet. This electric field is communicated to the tail plasma by the cross-tail current; that is, it has an independent link to the interplanetary source, not involving the ionosphere (Figure 6). Field lines threading the plasma sheet will communicate this **E** to the ionosphere, and here 'decoupling' is less likely, since the plasma is denser and distances are shorter. With the electric field extending to the ionosphere, currents will be driven there (to neutralize region 2 or to connect regions of different potential across the polar cap and along the auroral oval), and as a result, field-aligned currents will inevitably be produced [*Vasyliunas,* 1968]. This is how the plasma sheet comes to act as a source for Birkeland currents.

The overall effect of all the preceding is the same as would occur if part of the cross-tail current were diverted through the ionosphere. It has already been proposed [*McPherron et al.,* 1973] that such a diversion takes place during substorms. Here it is suggested that it exists at all times, although its extent may quite possibly increase during substorms.

Finally, it should be noted that wherever region 1 Birkeland currents originate in the plasma sheet, all field lines involved are closed. Thus if the reversal of **E** well above the ionosphere is observed to be clearly poleward of region 1 currents, this would agree with the source being in the plasma sheet.

Region 2: The Branching Ratio

Suppose that unbalanced electrical charge is generated at location **r** in the magnetosphere at a rate of $Q(\mathbf{r}, t)$ C m^{-3} s^{-1}

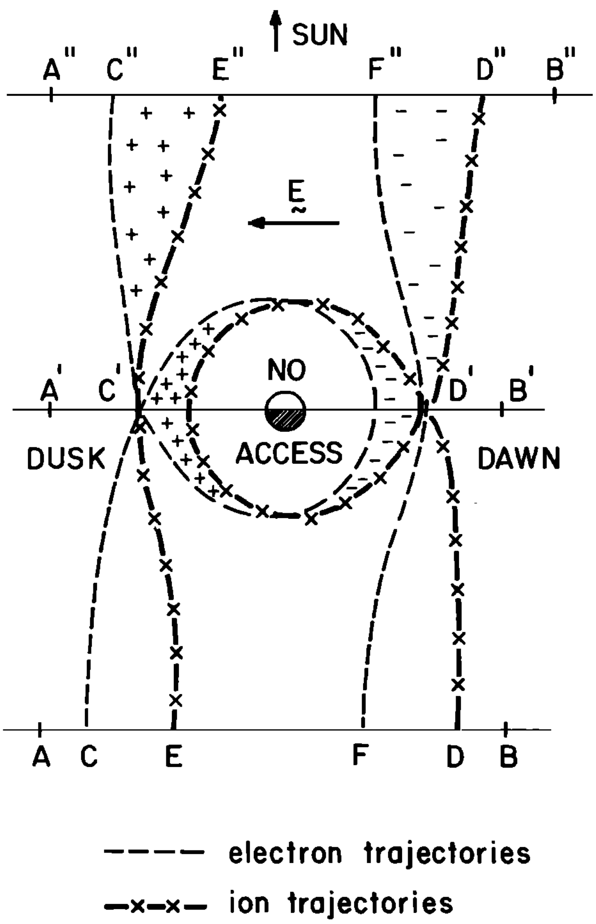

Fig. 7. A schematic view, with no day-night asymmetry, of particle convection past the earth, adapted from Figure 4 of *Schield et al.* [1969].

and that positive and negative charges (the latter denoted by a negative Q) appear in adjacent regions of space, threaded by different field lines. This space charge can then be removed in one of two ways:

1. The charge may be removed through the ionosphere, where a surface conductivity $\Sigma \sim 1$ mho (quiet times) exists ($\Sigma = \int \sigma_P \, dh$; Hall conductivity is not taken into account). The loss of electric energy in this case is assumed to be irreversible.

2. The charge may build up locally, creating a perturbation electric field which in this section (only) will be denoted simply by $\mathbf{E}$ (this is in addition to the externally imposed electric field, which is thereby lowered). If this happens, the polarization current $\mathbf{j}_p$ due to $d\mathbf{E}/dt$ [e.g., *Longmire*, 1963, p. 50] then drains almost all of the charge away. The work performed in driving $\mathbf{j}_p$ is invested as an increase in the bulk velocity $\mathbf{u} = \mathbf{E} \times \mathbf{B}/B^2$ of the plasma, and it may be recovered later by means of the dynamo effect: thus the magnetospheric plasma acts as a flywheel, storing energy in the form of bulk motion. (The ionosphere can also act as a flywheel, but this will not be addressed here.) Here $\mathbf{E}$ is understood to be in a frame at rest relative to the earth, but $d\mathbf{E}/dt$ is its rate of change as sensed by a particle moving with the plasma.

Schield et al. [1969] tacitly assumed that the first process was the dominant one, so that any space charge separated in the magnetosphere is discharged mainly through the ionosphere. Here two somewhat crude models were used to derive approximate branching ratios between the two currents and to check whether the assumption was justified. In the initial model the magnetic field was assumed to be constant, while later all steps were repeated assuming a cylindrical magnetic field, in order to assess the effect of diverging magnetic field lines. The calculations themselves are given in the appendix.

The results suggest that when a free electric charge Q appears in the equatorial magnetosphere and Q is given the choice between following the conductivity of the ionosphere or the 'conductivity' of the polarization current, with commonly observed time scales, it prefers the fomer. The most fallible assumption, among those underlying the calculation, is that of perfectly conducting field lines. Delays associated with the finite transit time of particles were also neglected.

In order to incorporate such restrictions it is useful to divide the charged particles which compose Q into three groups, according to their magnetic moments, and view each group separately. First there exist particles of the loss cone, which can flow to the ionosphere unimpeded; for these particles the assumption of an extremely high parallel conductivity is indeed valid. Second, there exist particles just outside the loss cone, covering a solid angle perhaps 10–20 times larger. Assuming that such particles have energies of a few keV, they will generally precipitate in response to a moderate $E_{\parallel}$, and their flow can be treated (approximately) by using a 'kinetic Ohm's law' [*Chiu et al.*, 1981, appendix; *Fridman and Lemaire*, 1980; *Knight*, 1973; *Stern*, 1981, equation (109)], which would somewhat modify the above calculation. Finally, however, there exist particles with relatively large pitch angles, which cannot be precipitated by any 'reasonable' $E_{\parallel}$. If the input distribution function is not too anisotropic, these particles will constitute the great majority, and their charge can be neutralized along their guiding field lines only if either they undergo scattering into the loss cone, or ionospheric particles of opposite sign arrive and are scattered into large pitch angles. If both these processes are slow compared to t_1, the time scale of charge removal through the ionosphere (less than 1 s), even the expanded loss cone will quickly become depleted, and charge flow to the ionosphere will then slow down to a trickle. It is then quite possible that large polarization currents will indeed flow.

Secondary Charge Separation

The loss of separated charges produces further imbalances. As a qualitative approximation, let the day-night symmetry of the magnetosphere be neglected, and let only particles of one particular set of (μ, J) be considered. Suppose that each second, N ions and N electrons with those values of (μ, J) are injected along an equatorial cross section AB of the nightside magnetosphere, as shown in Figure 7, adapted from Figure 4 of *Schield et al.* [1969]. As is done there, corotation is neglected.

Because day-night symmetry is assumed, the particles ultimately wind up (not simultaneously) on a mirror image line A″B″ on the dayside. In each species the particles are divided into two groups, those passing on the dawnside of the earth and those passing on the duskside, and a limiting trajectory separates those two groups; in Figure 7 the limiting trajectories are DD″ for ions and CC″ for electrons. Each such limiting trajectory also encloses a 'forbidden region' to which neither of its adjoining groups has access. It is because these forbidden regions do not exactly overlap that Alfvén layers are created, accessible to one species but not to the other: a positively charged layer near dusk and a negatively charged layer near dawn, as drawn. Figure 7 also includes the ion trajectory EE″ which passes the electron cusp C′ and the electron trajectory FF″ which passes the ion cusp D′.

The excess charge, of course, is mostly removed either by precipitation, by neutralization by ionospheric particles, or by the polarization current. Of these, the last named process was shown to play a minor role: to simplify matters, neutralization will also be first neglected, and only precipitation will be taken into account.

The charge separation effect is most pronounced in the dawn-dusk cross section A′B′. If somehow none of the particles were lost in Alfvén layers, then as particles advanced sunward from A′B′ to A″B″, ions and electrons would undergo 'reunification,' until at A″B″, thanks to the assumed day-night symmetry, the plasma would be reconstituted (in a steady state flow) exactly as it started from AB; that is, it would again be neutral. The ions from ED would reappear along E″D″, and the electrons from CF along C″F″.

Actually, however, at least some of these particles are lost in transit. The ions arriving at E″D″ are fewer in number than those that started from ED, because their trajectories have passed the positive Alfvén layer (if no electrons are drawn up, it can be argued that all these ions must be lost). Similarly, fewer electrons arrive at C″F″ than had started along CF. The depleted regions do not overlap: in particular, in C″E″C′ a deficiency of electrons would exist, and in F″D″D′ a deficiency of ions. Thus additional ions are precipitated on the duskside, and additional electrons on the dawnside, and in both cases the precipitation extends further sunward than in the simple model of *Schield et al.* [1969].

If instead of particle loss there occurs in the Alfvén layer neutralization by ionospheric particles, the situation is more complicated, because the neutralizing particles may have different (μ, J). For simplicity, let it be assumed that for such particles, (μ, J) are the same, so that neutralizing particles follow the same trajectories as those of Figure 7. In the primary Alfvén layer each neutralizing particle materializes inside its forbidden zone, and as such particles drift to the opposite side of the earth (through the midnight sector, if corotation is ignored), they enter the region marked 'no access.' Charge neutrality then cannot be maintained, and they are expected to be precipitated or neutralized (however, neutralizing particles materializing in the secondary regions remain there).

The net effect of secondary charge separation thus seems to be a modification of the primary process and does not change the main division, namely, that excess positive charge occurs on the duskside of the noon-midnight meridian and negative charge on the dawnside. Precipitation extends the space charge region sunward, neutralization may extend it toward midnight, and the fact that electrons are more likely to precipitate than ions may lead to interesting asymmetries. What does not seem to happen is what intuition might have suggested, namely, the reversal of the sign of the space charge. Intuitively, it might be argued that since in the charge separation phase between AB and A′B′, electrons are lost on the dawnside and ions on the duskside, during the charge reunification phase between A′B′ and A″B″ there should exist a deficiency of electrons at dawn and one of ions at dusk, and therefore particles of opposite sign should be precipitated there. As the preceding analysis shows, this does not seem to happen.

Time-Dependent Charge Separation

In the actual magnetosphere this process is expected to be modified by the day-night asymmetry, but an even greater effect may be caused by time variations. As an example, assume that an initial low level of convection is followed by a sudden particle injection along AB. Injected ions now will not only divide their numbers in unequal proportion between the dawn and dusk flanks but will also advance at different speeds: faster at dusk, where magnetic and electric drifts reinforce each other, and slower at dawn, where they are opposed. The opposite holds for electrons, and one thus expects that in an impulsive injection event the greater part of the front of either species (ions at dusk, electrons at dawn) is precipitated or neutralized. All the preceding mechanisms require that region 2 currents are peaked on the nightside, in agreement with observations.

Energy Flow

The currents of region 2, which flow into the ionosphere or out of it, link up to their sources by means of the currents of region 1. This means that the energy required by any of the processes involved is drawn from that source.

To illustrate this, suppose that an ion-electron pair starts out together in the plasma sheet (somewhere along AB of Figure 7) and moves earthward. As magnetic drifts separate the particles, they both gain energy W (for equatorial particles this is evident from the constancy of W/B). After space charge develops, they may be precipitated, or else particles of opposite sign may be drawn up from the ionosphere, and in each case, further energy may be imparted by $E_{\parallel}$. Additional energy is expended by continuing this current through the resistive ionosphere, and finally, the flow of charge closes as part of region 1 and may involve further energization by $E_{\parallel}$.

Each of these eight steps (four for ions, four for electrons) requires a supply of energy. The ultimate outcome of the sequence is that one elementary charge has been transferred from the positive source of region 1 at dawn and its negative source at dusk: the energy released by this process is shared by all eight steps, including the convective energization of drifting particles in the magnetosphere. The charge separation process and its associated region 2 currents thus enable the partial ring current, which energizes particles near midnight, to be continued toward its sources in the solar wind.

Cusp Currents

As noted by *Svalgaard* [1973], ionospheric currents in the polar cap depend on the interplanetary B_y component, and one would expect that such dependence would lead to a similar dependence of $j_{\parallel}$. Indeed, *Doyle et al.* [1981] have claimed that at any given time the cusp currents of Figure 1 exist only on one side of the noon-midnight axis, specifically, on the 'entry side' from which interplanetary field lines approach the polar cap under observation (when $B_y > 0$, field lines enter from the northern dawnside and from the southern duskside).

Earlier, *Heppner* [1972] had observed from OGO 6 that in the polar cap, $E_{\perp}$, too, tended to have an asymmetry

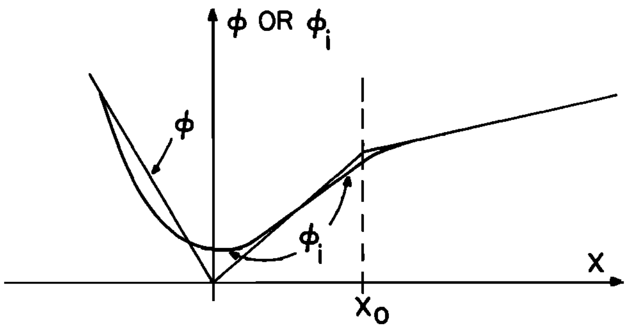

Fig. 8*a*. Lyons diagram for the 'entry side' of the polar cap, assumed to occur near dusk.

correlated with B_y. In particular, he noted that the cross-polar electric field was on the average stronger on the entry side of the cap than on the side opposite to it. The explanation commonly given [*Russell and Atkinson*, 1973; *Stern*, 1973] is that the earth solar wind linkage along open field lines is stronger on lines where the linkage is shorter, producing there a stronger **E**.

In the terminology of the present study, the source electric field on open field lines is nonuniform. Thus the slope of the polar potential Φ on the entry side, shown in the Lyons diagram of Figure 8*a*, is larger close to the reversal than it is further away from it. As the diagram suggests, $\Phi - \Phi_i$ in the region where the slope changes can drive a current in a direction opposite to that of region 1. This additional current is identified with the cusp current.

A different way of describing the same effect appears in Figure 8*b* and illustrates an argument of *Vasyliunas* [1979], who predicted Birkeland currents due to this effect, though these were not identified as cusp currents. Let the strengthening of **E** in the ionosphere, near the entry side, be approximated by allowing $E_\perp$ to assume one of two constant values, a smaller value E_1 for $x > x_0$ and a larger one E_2 for $x < x_0$. If constant ionospheric conductivity is assumed, then $j_\perp$ is proportional to $E_\perp$, and it is thus larger for $x < x_0$ than it is for $x > x_0$. The equation of continuity then requires an inflow of $j_\parallel$ near $x = x_0$. Strictly speaking, Figure 8*a* assumes two constant values of E in the source region, while Figure 8*b* assumes them to be constant in the ionosphere. The qualitative effect, however, is the same in both cases.

It is interesting to note that these currents exist only close to noon, even though the asymmetry of **E** is reported to exist over much larger parts of the polar cap. Because current flow across the polar cap is clearly involved (as shown by Figure 8*b*), the reason could be related to the higher ionospheric conductivity near noon, where the solar illumination of the ionosphere is more intense and where cusp particles also may enhance the conductivity. The control of B_y-related Birkeland currents by solar illumination can be indirectly inferred from the strong seasonal dependence by B_y-related ionospheric currents [*Svalgaard*, 1973, Figure 8], which presumably are fed by such Birkeland currents and which are much weaker in winter.

Four-Lobed Fields

Whenever the interplanetary magnetic field has a strong northward slant, if its field lines are continued until they intercept the earth's surface, then those of them that enter the magnetic poles have a polarity opposite to that of geomagnetic field lines there. It follows that if at such times, interplanetary lines merge with terrestrial ones, they must bend around sharply. It is proposed here that at times when a four-lobed polar electric field (or current pattern) is observed, the magnetosphere is in fact temporarily closed.

A closed magnetosphere may have plasma flow induced in it by a viscouslike interaction at the magnetopause [*Axford and Hines*, 1961]. Plasma then flows nightward near the boundary, and a return flow exists closer to the noon-midnight axis, as in Figure 2, which schematically represents motion in the equatorial plane. The velocity **v** of such a flow is associated with an electric field $\mathbf{E} = -\mathbf{v} \times \mathbf{B}$, and **E** maps into the ionosphere and generates currents there: this is the dynamo mechanism, already described. What follows concerns only the flow outside the corotating part of the magnetosphere.

The resulting potential distribution along the dawn-dusk cross section AA′ of the equatorial plane (Figure 2) resembles the one discussed earlier in connection with the boundary layer dynamo. We assume solar geomagnetic coordinates, with x pointing sunward and y toward dusk. Let that cross section intercept the magnetopause at (A, A′) and the corotating region at (C, C′), and let B and B′ be the points where v_x reverses; E_y then reverses there, too. One thus expects the largest potential difference along the line AA′ to be the one between B and B′. Furthermore, if the magnetopause is an equipotential, A and A′ should have the same potential. If that holds, then when these potentials are mapped into the ionosphere (Figure 9), the potential along the line bb′, the footprint of BB′, varies smoothly, and only the usual two-celled potential pattern is expected.

Instead of considering potentials it may be more meaningful to discuss current flows. If the magnetopause were a perfect conductor, then the dynamo currents AB and B′A′ would be smooth continuations of each other, A and A′ would have (as before) the same potential, and no currents Aa or A′a′ would exist. It is not at all certain, however, that the magnetopause has this property. If the magnetopause is closed, it is expected to be a tangential discontinuity of the magnetic field: modes of particle motion along such a discontinuity exist [*Alpers*, 1969, 1971], but it is not known, for instance, how well they can couple to the interior of the magnetosphere. In order to explain the four-lobed patterns it is indeed necessary that the magnetopause not be efficient in transmitting current, so that current flows along aA and A′a′ will in fact exist.

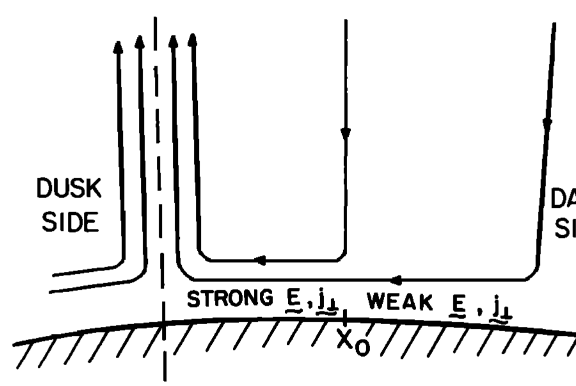

Fig. 8*b*. A cross section of the cross-polar current flow, similar to the one in Figure 5, under conditions similar to those of Figure 8*a*.

If such currents were infinitely narrow, then their ionospheric effects would still effectively cancel. However, as noted earlier, the capacity of the magnetosphere to carry a steady $j_\parallel$ is limited to the order of 1 μA/m^2, so that any Birkeland currents generated are by necessity spread out over an appreciable width. This suggests that the closure currents denoted by aA and A'a' have a finite width and that their centers are separated by a small distance. The currents connecting them will then produce in the ionosphere a small two-lobed pattern of $\mathbf{j}_\perp$. Associated with it will be a small two-lobed pattern of electric equipotentials, opposed to the regular pattern in which it is embedded. The resulting pattern resembles the diagram of *Burke et al.* [1979] (Figure 4*b* here), in which the central lobes are indeed rather small. As expected, the entire polar cap in this case is also very small, since (as noted above) the magnetic flux threaded by the boundary flow is quite limited.

CONCLUSION

The main conclusions of this study have already been listed in the initial section and will not be repeated. Together they constitute a reasonable qualitative explanation for most properties of quiet-time Birkeland currents, excluding the Harang filament. For the first time a consistent theory has been given for the relation between the convection reversal and the position of Birkeland current sheets, for the four-lobed patterns observed with northward B_z, and for the cusp currents.

Underlying this theory is the tight association between $j_\parallel$ and $E_\parallel$ for currents that flow both up and down. This association is implicit in the use of Lyons diagrams and (at least for region 1) in the relatively large width of Birkeland current sheets. It is hoped that future observations will test and confirm the many aspects of this theory.

APPENDIX: CALCULATION OF THE BRANCHING RATIO

Slab Model

Assume that a homogeneous plasma with density ρ kg/m^3 is confined between two insulating plates at $x = 0$ and at $x = x_0$. A homogeneous magnetic field

$$\mathbf{B} = B_0\hat{\mathbf{z}} \quad \text{(A1)}$$

is present, and the plasma is confined to the half plane $z > 0$

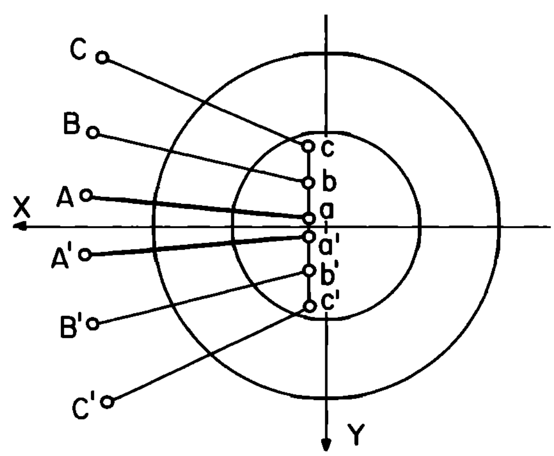

Fig. 9. Schematic map of the polar cap with the footprint of the equatorial dawn-dusk cross section of the closed magnetosphere of Figure 2.

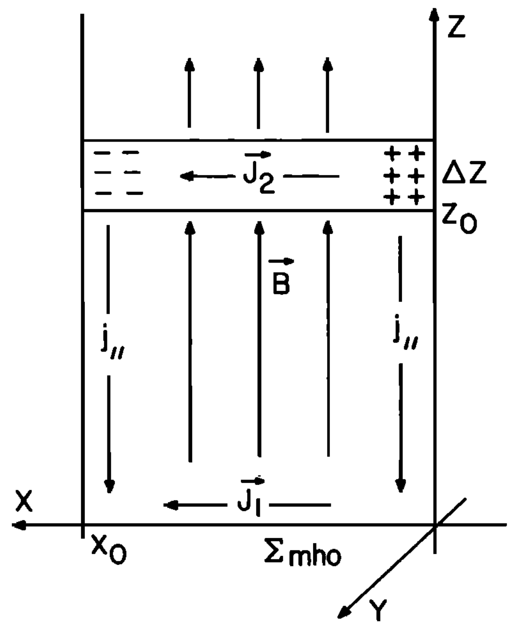

Fig. 10. The branching of J_1 and J_2 in the neutralization of magnetospheric space charge: the slab model.

by a conducting plate at $z = 0$, a plate which has a surface conductivity Σ per meter length in the y direction (Figure 10). Thus if an electric field $\mathbf{E} = E(x)\hat{\mathbf{x}}$ is imposed upon the plate, its surface current density will be

$$\mathbf{J}_1 = \Sigma_0 E(x)\hat{\mathbf{x}} \text{ A/m} \quad \text{(A2)}$$

that is, $\mathbf{J}_1$ A for each strip 1 m wide in the y direction.

The space occupied by the plasma is divided into two regions. For $0 < z < z_0$ the conductivity is assumed to be infinite along **B** and zero orthogonally to B. For $z_0 < z < z_0 + \Delta z$, in addition, free charge is generated spontaneously at a rate

$$Q(x, t) = Q_0 \sin \omega t \cos kx \quad \text{(A3)}$$

where the dimension of Q is coulombs per second per square meter, charge generation rate per unit area in the xy plane, since all of Δz has already been summed over. (The region with $z > z_0 + \Delta z$ is not assumed to be involved in any current flow.) The above form can be superposed to a Fourier integral, to analyze different time dependences, but here we assume only that it lasts a time

$$t_2 = \pi/\omega \quad \text{(A4)}$$

and then ceases. The charge Q then drains away either as a polarization current $\mathbf{j}_2$ in the layer or as a conduction current in the boundary: to derive the ratio between the currents, we note that current continuity requires, at $z = 0$,

$$\partial J_1/\partial x = -j_\parallel \quad \text{(A5)}$$

where the parallel current density $j_\parallel$ is taken as positive in the $+z$ direction. In the layer Δz, if the total integrated surface current density is $J_2(x, t)$ and we assume that the buildup of space charge absorbs a negligible fraction of Q, the same relation gives

$$\partial J_2/\partial x - j_\parallel = Q(x, t) \quad \text{(A6)}$$

Adding,

$$\partial J_1/\partial x + \partial J_2/\partial x = Q(x, t) \quad \text{(A7)}$$

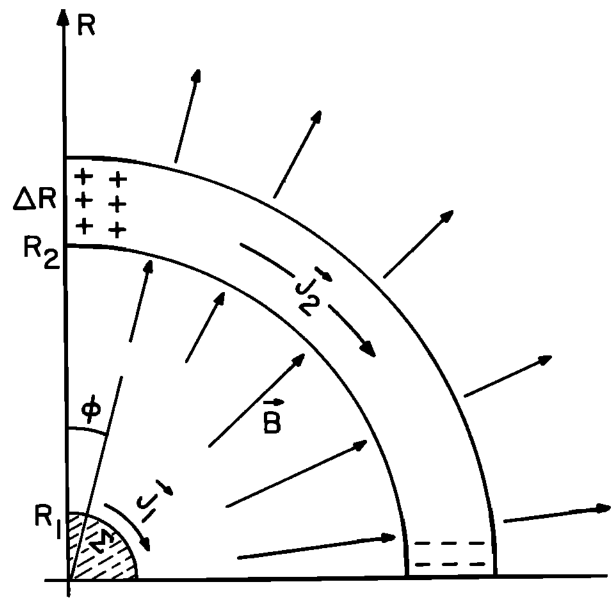

Fig. 11. The cylindrical analogue of Figure 10.

The polarization current $\mathbf{j}_2$ in SI units is

$$\mathbf{j}_2 = (c/V_A)^2\varepsilon_0\, d\mathbf{E}/dt \quad \text{(A8)}$$

where V_A is the Alfvén velocity

$$V_A = cB[\varepsilon_0/\rho]^{1/2} \quad \text{(A9)}$$

Integrating over Δz,

$$\mathbf{J}_2 = \Delta z(c/V_A)^2\varepsilon_0\, d\mathbf{E}/dt \quad \text{(A10)}$$

The conduction current satisfies

$$\mathbf{J}_1 = \Sigma\mathbf{E} \quad \text{(A11)}$$

The integrated Pedersen conductivity Σ (strictly, Σ_P) for the quiet (sunlit) and auroral ionosphere has been estimated at 0.56 mho and 36 mhos, respectively, by *Boström* [1964] and at 1 mho and 5–10 mhos by *Yasuhara and Akasofu* [1977]. Let the electric potential now be

$$\Phi(x, t) = \Phi_0(t)\cos kx \quad \text{(A12)}$$

Then

$$E = E_x(x, t) = \Phi_0(t)k\sin kx \quad \text{(A13)}$$

which means that E is largest in the middle, at the boundary between opposed charges. From this, (A10) and (A11) can be evaluated, and after substitution in (A7) and cancelation of $\cos kx$ one finds

$$a\, d\Phi_0/dt + \Sigma\Phi_0 = b\sin\omega t \quad \text{(A14)}$$

where

$$a = \Delta z(c/V_A)^2\varepsilon_0 \quad \text{(A15)}$$

$$b = Q_0/k^2 \quad \text{(A16)}$$

A solution which vanishes at $t = 0$ is

$$\Phi_0 = U\sin\omega t + V\cos\omega t - V\exp(-\Sigma t/a) \quad \text{(A17)}$$

where the last term is contributed by the homogeneous part of the equation. That term decays with a time constant

$$t_1 = a/\Sigma = \Delta z\rho/(\Sigma B^2) \quad \text{(A18)}$$

and if $\Sigma = 1$ mho, $B = 100$ nT $= 10^{-7}$ MKS, $n = 10^6$ m^{-3}, ion mass $m_i = 1.67 \times 10^{-27}$ kg, $\rho = nm_i$, and $\Delta z = 2 \times 10^7$ m (about 3 R_E), then $t_1 \sim 3$ s. Substituting (A17) in (A14) and equating sines and cosines then gives

$$U = \Sigma b/M \qquad V = -a\omega b/M \quad \text{(A19)}$$

where

$$M = \Sigma^2 + a^2\omega^2 \quad \text{(A20)}$$

With this,

$$J_2 = a\, dE/dt = a[\omega(U\cos\omega t - V\sin\omega t) + (V/t_1) \cdot \exp(-t/t_1)]k\sin kx \quad \text{(A21)}$$

$$J_1 = \Sigma E = \Sigma[U\sin\omega t + V\cos\omega T - V\exp(-t/t_1)]k\sin kx \quad \text{(A22)}$$

We seek to derive the ratio of the averages of J_2 and J_1, when the charge input pulse of (A3) lasts only the time t_2 of (A4), after which everything decays with a time constant t_1. Let us assume that t_2 is considerably longer than t_1, which also means that U is much larger than V. Then

$$\langle J_2\rangle t_2 = \int_0^{t_2} J_2\, dt = -ak\sin kx\, V[1 + \exp(-t_2/t_1)] \quad \text{(A23)}$$

and since t_2/t_1 is large, the exponential will be neglected. Also,

$$\langle J_1\rangle = \int_0^{t_2} J_1\, dt \sim \Sigma k\sin kx(2t_2/\pi)U \quad \text{(A24)}$$

Thus

$$|\langle J_2\rangle/\langle J_1\rangle| \sim (\pi t_1/t_2)^2/2 \quad \text{(A25)}$$

which is much less than unity. Therefore if the present approximation is any guide, the magnetosphere makes a poor flywheel for time intervals much larger than 3 s.

There exists, however, a qualitative difference between this model and the magnetosphere, because here field lines are parallel, while in the actual magnetosphere they diverge as one proceeds away from the ionosphere. To incorporate such a divergence, at least qualitatively, we repeat the calculation for a cylindrical model.

Cylindrical Model

Let the plasma be confined to a quadrant of a cylindrical geometry (R, ϕ, z), with

$$\mathbf{B} = B_0(R_1/R)\hat{\mathbf{R}} \quad \text{(A26)}$$

The plasma is bounded at $R = R_1$ by a conducting surface with surface conductivity Σ, and charge is generated in a layer $R_2 < R < R_2 + \Delta R$ at a rate

$$Q = Q_0\sin\omega t\cos 2\phi \quad \text{(A27)}$$

(Figure 11). As before, the charge-generating process only lasts a time t_2, given by (A4), after which everything decays. Equation (A11) may still be used, but (A12) is replaced by

$$\Phi = \Phi_0(t)\cos 2\phi \quad \text{(A28)}$$

from which

$$\mathbf{E} = \Phi_0(t)(2/R) \sin 2\phi \, \hat{\boldsymbol{\phi}} \quad (A29)$$

Retracing (A5)–(A7),

$$(1/R_1)\, \partial J_1/\partial\phi = -j_{\parallel} \quad (A30)$$

$$(1/R_2)\, \partial J_2/\partial\phi - j_{\parallel} = Q \quad (A31)$$

$$(1/R_2)\, \partial J_2/\partial\phi + (1/R_1)\, \partial J_1/\partial\phi = Q_0 \sin \omega t \cos 2\phi \quad (A32)$$

Replacing Δz in (A15) by ΔR and retracing previous steps gives, in analogy with (A14),

$$(4/R_2^2)a\, d\Phi_0/dt + (4/R_1^2)\Sigma\Phi_0 = Q_0 \sin \omega t \quad (A33)$$

The rest is completely analogous to the solution of (A14), except that the ratio t_1 between the two coefficients on the left, which was a/Σ, now becomes

$$t_1 = (R_1/R_2)^2(a/\Sigma) \quad (A34)$$

Thus the 'flywheel effect' gets worse: one factor of R_1/R_2 may be traced to the weakening of E with distance, and another to the divergence of $j_{\parallel}$. In the actual magnetosphere, B decreases with distance not as r^{-1} (as in A26) but more as r^{-3}, and this is likely to reduce the effect even more than is derived here.

REFERENCES

Akasofu, S.-I., Interaction between a magnetized plasma flow and a strongly magnetized celestial body with an ionized atmosphere: Energetics of the magnetosphere, *Annu. Rev. Astron. Astrophys.*, *20*, 117–138, 1982.

Akasofu, S.-I., et al., Power transmission from the solar wind–magnetosphere dynamo to the magnetosphere and to the ionosphere: Analysis of the IMS Alaska meridian chain data, *Planet. Space Sci.*, *29*, 721–730, 1981.

Alfvén, H., A theory of magnetic storms and the aurorae, *K. Sven. Vetenskapsakad. Handl., Ser. 3, 18(3)*, 1939. (Partially reprinted with comments by A. Dessler and W. Wilcox in *Eos Trans. AGU*, *51*, 180–194, 1970.)

Alfvén, H., and C.-G. Fälthammar, *Cosmical Electrodynamics*, 2nd ed., Oxford University Press, New York, 1963.

Alpers, W., Steady state charge neutral models of the magnetopause, *Astrophys. Space Sci.*, *5*, 425–437, 1969.

Alpers, W., On the equilibrium of an exact charge neutral magnetopause, *Astrophys. Space Sci.*, *11*, 471–474, 1971.

Axford, W. I., and C. O. Hines, A unifying theory of high-latitude geophysical phenomena and geomagnetic storms, *Can. J. Phys.*, *39*, 1433–1464, 1961.

Behannon, K. W., Geometry of the geomagnetic tail, *J. Geophys. Res.*, *75*, 743–753, 1970.

Boström, R., A model of the auroral electrojets, *J. Geophys. Res.*, *69*, 4983–4999, 1964.

Boström, R., Currents in the ionosphere and magnetosphere, *Ann. Geophys.*, *24*, 681–694, 1968.

Burke, W. J., et al., Polar cap electric field structure with northward interplanetary magnetic field, *Geophys. Res. Lett.*, *6*, 21–24, 1979.

Burrows, J. R., and I. B. McDiarmid, Trapped particle boundary regions, in *Critical Problems of Magnetospheric Physics*, edited by E. R. Dyer, pp. 83–106, Inter-Union Committee on Solar-Terrestrial Physics, National Academy of Sciences, Washington, D. C., 1972.

Bythrow, P. F., et al., Observational evidence for a boundary layer source of dayside region 1 field-aligned currents, *J. Geophys. Res.*, *86*, 5577–5589, 1981.

Chapman, S., and J. Bartels, *Geomagnetism*, Oxford University Press, New York, 1940.

Chiu, Y. T., and M. Schulz, Self-consistent particle and parallel electrostatic field distributions in the magnetosphere-ionosphere auroral region, *J. Geophys. Res.*, *83*, 629–642, 1978.

Chiu, Y. T., A. L. Newman, and J. M. Cornwall, On the structures and mapping of auroral electrostatic potentials, *J. Geophys. Res.*, *86*, 10,029–10,037, 1981.

Doyle, M. A., F. J. Rich, W. J. Burke, and M. Smiddy, Field-aligned currents and electric fields observed in the region of the dayside cusp, *J. Geophys. Res.*, *86*, 5656–5664, 1981.

Eastman, T. E., E. W. Hones, Jr., S. J. Bame, and J. R. Asbridge, The magnetospheric boundary layer: Site of plasma, momentum and energy transfer from the magnetosheath into the magnetosphere, *Geophys. Res. Lett.*, *3*, 685–688, 1976.

Frank, L. A., and K. L. Ackerson, Several recent findings concerning the dynamics of the earth's magnetotail, *Space Sci. Rev.*, *23*, 375–392, 1979.

Fridman, M., and J. Lemaire, Relationship between auroral electrons fluxes and field-aligned electric potential difference, *J. Geophys. Res.*, *85*, 644–670, 1980.

Fujii, R., T. Iijima, T. A. Potemra, and M. Sugiura, Seasonal dependence of large-scale Birkeland currents, *Geophys. Res. Lett.*, *8*, 1103–1106, 1981.

Fukushima, N., Equivalence in ground geomagnetic effect of Chapman-Vestine's and Birkeland-Alfvén's current systems for polar magnetic storms, *Rep. Ionos. Space Res. Jpn.*, *23*, 219–227, 1969.

Fukushima, N., Generalized theorem for no ground magnetic effect of vertical current connected with Pedersen currents in the uniform-conductivity ionosphere, *Rep. Ionos. Space Res. Jpn.*, *30*, 35–40, 1976.

Hardy, D. A., F. J. Rich, and W. J. Burke, Auroral dynamics during a substorm near the dawn-dusk meridian (abstract), *Eos 60*, 358, 1979.

Helmer, J. C., Theory of forbidden zones in the flow of magnetized plasma, *Phys. Fluids*, *6*, 723–728, 1963.

Heppner, J. P., Polar cap electric field distributions related to the interplanetary magnetic field direction, *J. Geophys. Res.*, *77*, 4877–4887, 1972.

Heppner, J. P., Empirical models of high-latitude electric fields, *J. Geophys. Res.*, *82*, 1115–1125, 1977.

Hines, C. O., The energization of plasma in the magnetosphere: Hydromagnetic and particle-drift approaches, *Planet. Space Sci.*, *10*, 239–246, 1963.

Horwitz, J. L., and S.-I. Akasofu, On the relationship of the polar cap current system to the north-south component of the interplanetary magnetic field, *J. Geophys. Res.*, *84*, 2567–2572, 1979.

Hughes, T. J., and G. Rostoker, A comprehensive model current system for high-latitude magnetic activity, I, The steady state system, *Geophys. J. R. Astron. Soc.*, *58*, 525–569, 1979.

Iijima, T., and T. A. Potemra, The amplitude distribution of field-aligned currents at northern high latitudes observed by Triad, *J. Geophys. Res.*, *81*, 2165–2174, 1976*a*.

Iijima, T., and T. A. Potemra, Field-aligned currents in the dayside cusp observed by Triad, *J. Geophys. Res.*, *81*, 5971–5979, 1976*b*.

Iijima, T., N. Fukushima, and R. Fujii, Transverse and parallel geomagnetic perturbations over the polar regions observed by MAGSAT, *Geophys. Res. Lett.*, *9*, 369–372, 1982.

Inhester, B., W. Baumjohann, R. A. Greenwald, and E. Nielsen, Joint two-dimensional observations of ground magnetic and ionospheric electric fields associated with auroral zone currents, 3, Auroral zone currents during the passage of a westward traveling surge, *J. Geophys.*, *49*, 155–162, 1981.

Kamide, Y., and G. Rostoker, The spatial relationship of field-aligned currents and auroral electrojets to the distribution of nightside auroras, *J. Geophys. Res.*, *82*, 5589–5608, 1977.

Karlson, E. T., Motion of charged particles in an inhomogeneous magnetic field, *Phys. Fluids*, *5*, 476–486, 1962.

Karlson, E. T., Streaming of a plasma through a magnetic dipole field, *Phys. Fluids*, *6*, 708–722, 1963.

Kennel, C. F., and H. E. Petschek, Limit on stably trapped particle fluxes, *J. Geophys. Res.*, *71*, 1–28, 1966.

Knight, S., Parallel electric fields, *Planet. Space Sci.*, *21*, 741–750, 1973.

Lemaire, J., The magnetospheric boundary layer: A stopper region for a gusty solar wind, in *Quantitative Modeling of Magnetospheric Processes, Geophys. Monogr. Ser.*, vol. 22, edited by W. P. Olson, pp. 412–422, AGU, Washington, D. C., 1979.

Longmire, C. L., *Elementary Plasma Physics*, 296 pp., Interscience, New York, 1963.

Lyons, L. R., Plasma processes in the earth's radiation belts, in *Solar System Plasma Physics*, vol. III, edited by C. F. Kennel, L. J. Lanzerotti, and E. N. Parker, pp. 137–163, North-Holland, Amsterdam, 1979.

Lyons, L. R., Generation of large-scale regions of auroral currents,

electric potentials, and precipitation by the divergence of the convection electric field, *J. Geophys. Res.*, *85*, 17–24, 1980.
Lyons, L. R., Discrete aurora as the direct result of an inferred, high-altitude generating potential distribution, *J. Geophys. Res.*, *86*, 1–8, 1981.
McPherron, R. L., C. T. Russell, and M. P. Aubry, Satellite studies of magnetospheric substorms on August 15, 1968, 9, Phenomenological model for substorms, *J. Geophys. Res.*, *78*, 3131–3149, 1973.
Nisbet, J. S., M. J. Miller, and L. A. Carpenter, Currents and electric fields in the ionosphere due to field-aligned auroral currents, *J. Geophys. Res.*, *83*, 2647–2657, 1978.
Persson, H., Electric field along a magnetic line of force in a low-density plasma, *Phys. Fluids*, *6*, 1756–1759, 1963.
Potemra, T. A., Current systems in the earth's magnetosphere, *Rev. Geophys. Space Phys.*, *17*, 640–656, 1979.
Rostoker, G., Current flow in the magnetosphere during magnetospheric substorms, *J. Geophys. Res.*, *79*, 1994–1998, 1974.
Rostoker, G., The auroral electrojets, in *Dynamics of the Magnetosphere*, edited by S.-I. Akasofu, pp. 201–211, D. Reidel, Hingham, Mass., 1980.
Rostoker, G., and R. Boström, A mechanism for driving the gross Birkeland current configuration in the auroral oval, *J. Geophys. Res.*, *81*, 235–244, 1976.
Rostoker, G., and T. J. Hughes, A comprehensive model current system for high-latitude magnetic activity, II, The substorm component, *Geophys. J. R. Astron. Soc.*, *58*, 571–581, 1979.
Russell, C. T., and G. Atkinson, Comments on 'Polar cap electric field distributions related to interplanetary magnetic field direction' by J. P. Heppner, *J. Geophys. Res.*, *78*, 4001–4002, 1973.
Saflekos, N. A., R. E. Sheehan, and R. L. Carovillano, Global nature of field-aligned currents and their relation to auroral phenomena, *Rev. Geophys. Space Phys.*, *20*, 709–734, 1982.
Schield, M. A., J. W. Freeman, and A. J. Dessler, A source for field-aligned currents at auroral latitudes, *J. Geophys. Res.*, *74*, 247–256, 1969.
Shuman, B. M., et al., Field-aligned current, convective electric field, and auroral particle measurements during a major magnetic storm, *J. Geophys. Res.*, *86*, 5561–5575, 1981.
Smiddy, M., et al., Effects of high-latitude conductivity on observed convection electric fields and Birkeland currents, *J. Geophys. Res.*, *85*, 6811–6818, 1980.
Sonnerup, B. U. Ö., Theory of the low-latitude boundary layer, *J. Geophys. Res.*, *85*, 2017–2026, 1980.
Spiro, R. W., P. H. Reiff, J. L. Burch, and R. A. Heelis, High latitude plasma convection and the IMF (abstract), *Eos Trans. AGU*, *60*, 915, 1979.
Stern, D. P., A study of the electric field in an open magnetospheric model, *J. Geophys. Res.*, *78*, 7292–7305, 1973.
Stern, D. P., Large-scale electric fields in the earth's magnetosphere, *Rev. Geophys. Space Phys.*, *15*, 156–194, 1977.
Stern, D. P., Energetics of the magnetosphere, *NASA Tech. Memo.*, *82039*, 1980.
Stern, D. P., One-dimensional models of quasi-neutral parallel electric fields, *J. Geophys. Res.*, *86*, 5839–5860, 1981.
Sugiura, M., Identification of the polar cap boundary and the auroral belt in the high-latitude magnetosphere: A model for field-aligned currents, *J. Geophys. Res.*, *80*, 2057–2068, 1975.
Sugiura, M., and T. A. Potemra, Seasonal variation in the field-aligned current as observed by Triad (abstract), *Eos Trans. AGU*, *59*, 1170, 1978.
Svalgaard, L., Polar cap magnetic variations and their relationship with the interplanetary magnetic sector structure, *J. Geophys. Res.*, *78*, 2064–2078, 1973.
Vasyliunas, V. M., Discussion of paper by Harold E. Taylor and Edward W. Hones, Jr., 'Adiabatic motion of auroral particles in a model of the electric and magnetic fields surrounding the earth,' *J. Geophys. Res.*, *73*, 5805–5807, 1968.
Vasyliunas, V. M., Interaction between the magnetospheric boundary layers and the ionosphere, Proceedings of the Magnetospheric Boundary Layers Conference, Alpbach, 11–15 June 1979, *Eur. Space Agency Spec. Publ., ESA SP-148*, 387–393, 1979.
Williams, L. P., *Michael Faraday*, 531 pp., Basic Books, New York, 1965.
Yasuhara, F., and S.-I. Akasofu, Field-aligned currents and ionospheric electric fields, *J. Geophys. Res.*, *82*, 1279–1284, 1977.
Zmuda, A. J., and J. C. Armstrong, The diurnal flow pattern of field-aligned currents, *J. Geophys. Res.*, *79*, 4611–4619, 1974.

(Received July 13, 1982;
accepted October 18, 1982.)