

1993 CEDAR Workshop
Boulder, Colorado
June 21-26, 1993

Tutorial Lecture

by Gordon Rostoker
University of Alberta

Geomagnetic Substorms

Geomagnetic Substorms

**Gordon Rostoker
Canadian Network for Space Research
and Department of Physics
University of Alberta
Edmonton, Alberta CANADA T6G 2J1**

I would like to begin this lecture by noting that the term substorm means different things to different people. This single fact accounts for much of the confusion in the existing literature. It behooves us to go back into history to find out how this came to pass. This will allow you to read the prior literature with what I hope will prove to be an anti-confusion filter.

Over and above the problems of terminology which I will be drawing to your attention, there is one other simple reason why the term substorm did not have a universally acceptable definition. That is because there were many types of instruments used to monitor the solar-terrestrial interaction in the "old days", and each observer has ended up trying to identify the "substorm" within their data set. The data sets themselves featured observations of the aurora, magnetic field fluctuations, electric field variations, cosmic noise absorption as measured on the ground not to mention magnetic and electric field variations, energetic particle fluxes, and noise bursts in every portion of the em spectrum as measured in space. In each of these data sets the researchers sought to find the substorm - and they always found one, whatever it was.

While all these aforementioned parameters have been measured at one time or another, it is the magnetic field fluctuation that has become the universal signature that scientists have appealed to in deciding whether or not a substorm is in progress. So it is with magnetic signatures that I will start the story of the substorm, beginning by showing you how the concept evolved to its present confusing state. But I will leave you with no confusion in your minds at all - hopefully!

The term substorm seems to have arisen in the early 1960's, when Sidney Chapman and Syun Akasofu were working together up in the wilds of Alaska. They were aware of the only magnetic disturbance around that had a real name - the magnetic storm. FIGURE 1 shows a storm detected at a low latitude station. The canonical storm had a sudden commencement, an initial phase, a main phase and a recovery phase. By the time Chapman and Akasofu started their work, it was clear that the main phase was caused by the formation of a ring current circling the Earth, and the recovery phase reflected the slow decay of that ring current over a day or two. [Subsequently it became apparent that you could grow a ring current without a sudden storm commencement, and that you could go right into a main phase after a sudden storm commencement without an initial phase. That all became understandable when it became clear that the interplanetary magnetic field direction - particularly the direction of its north-south component - decided whether or there would be a main phase ring current. A ring current could grow and a disturbance could be termed a storm as long as you got a good dose of strong southward pointing IMF. If the storm was initiated by the impact of a shock on the outer boundary of the magnetosphere, there might be no initial phase if the IMF behind the shock was southward. As for the sudden storm commencement - well, you didn't need one of them either. An ssc simply indicated an increase in the mass density flux incident on the magnetosphere. You could get a perfectly respectable storm with no ssc - just a solid dose of southward IMF - which tells you that the only key property of a storm is the ring current enhancement. Finally, one is faced with the horrible reality that there is no threshold of ring current strength above which you can say there is a storm and below which you can say there is no storm. At least, not a threshold based on the physical properties of a storm. There are operational thresholds used in the forecasting business, but there is no physical basis for the choice of these thresholds. This is not a trivial problem - the fact is that ring currents can be very weak or very strong and ring current strengths are not quantized.] Ring currents

are the product of some process, and that process may be in progress almost all the time with only the size of the product pushing us to say whether a storm is in progress or not. Which means the term storm may refer to no particular special process which is not present when a storm is not deemed to be in progress. Unfortunately, the term "storm" has been usurped - because, when we talk about a substorm, I would much rather call that phenomenon a "storm". Which brings us to what Akasofu and Chapman discovered when they started to look at storms at high latitude stations. They looked at both auroral and magnetic field data, and on the basis of the auroral signatures Syun Akasofu defined the *auroral substorm* in 1964. FIGURE 2 shows the stages of a substorm as it was understood in those days. Keep in mind that this was all based on records from sparsely spaced allsky cameras whose fields of view most often did not overlap, and for which the film speeds permitted only bright discrete auroras to be detected. The old allsky cameras simply couldn't detect the diffuse aurora which are often observed in the equatorward portion of the evening sector auroral oval. Basically, this Figure asks you to break down a substorm into an *expansive phase* and a *recovery phase*. Now Akasofu and Chapman realized that the auroral substorms were accompanied by magnetic signatures of the type shown in FIGURE 3. By using latitude profiles such as that shown in FIGURE 4, one can see that these disturbances involve a westward latitudinally localized ionospheric electrojet. Since most researchers dealing with substorms were likely to have magnetometer data at their disposal rather than allsky camera data, they tended to look for substorm signatures in the network of magnetometers operated around the world. These they called *polar magnetic (or geomagnetic) substorms*. Whenever they saw the sharp onset of one of these magnetic bays (as they were called in the "old days"), they claimed they saw the onset of a substorm. In reality they had detected the onset of a substorm expansive phase, however people tended to be imprecise about what they were saying and simply said they saw a substorm. In the late 1960's the term "magnetospheric substorm" was coined to encompass all the various signatures observed by different types of detectors during an auroral substorm. However, as we shall see, the term substorm itself was not clearly defined, and that led to

confusion in the literature which dogs us to this very day. The problems started back in the 1960's, with the question of what form the equivalent current system for a substorm took. FIGURE 5 shows the two that were touted. The two cell model following the pattern proposed for geomagnetic bays by Silsbee and Vestine in the 1930's, was championed by Heppner and Sugiura. This model held that a substorm featured an eastward electrojet across dusk and a westward jet across dawn, converging towards midnight. In contrast, Akasofu Chapman and Meng believed that the substorm involved only one cell, with a westward electrojet crossing midnight from the morning into the evening sector. I got involved in this back in 1969 by noting that there was evidence for both types of equivalent current systems in the data. Witness the event of FIGURE 3 as seen at the lower latitude station of Tromso (FIGURE 6). Here you can see a lower frequency disturbance, which takes the form of a negative perturbation in the H-component characteristic of a westward electrojet. So, even at the end of the 1960's there was evidence that substorm disturbances involved two processes, one yielding large scale electrojets crossing the dawn and dusk meridians, and the other yielding a westward electrojet in the midnight sector.

The 1970's saw the two cell model set aside while those studying substorms considered only one of the components of the substorm disturbance - the so called substorm "current wedge", shown in FIGURE 7. What make this picture important is that it spoke to the question of what the real current system was that flowed during a substorm. Bob McPherron made this picture famous because it suggested that the wedge reflected a diversion of crosstail current. However, the actual idea of the geometrical form of the current wedge had been proposed long ago by Birkeland (FIGURE 8) and had been modelled at the end of the 1960's by Bonnevier, Bostrom and myself in Stockholm (J.G.R. 1970) .

[Bjorn Bonnevier had been assigned by Hannes Alfvén the problem of writing a computer program to model a three dimensional current system in the early 1960's, but by the time it was completed Alfvén had lost interest. When I came to Stockholm

in 1966, with the observations of substorm disturbances in my back pocket, it immediately became clear that the three dimensional current loop modelled by Bonnevier fitted those observations beautifully. But then, I didn't realize the importance of the lower frequency component of the activity at that time - only that it existed and ought not to be ignored for long. However, it was in fact ignored for a time while the current wedge had the exclusive attention of the substorm community.]

Syun Akasofu came back into the picture in 1978 when, working with Paul Perrault, he reached the conclusion that it was the low frequency component of the substorm disturbance that best tracked the input of energy from the solar wind. He called this component "directly driven activity" while the shorter polar magnetic substorms were attributed to unloading of energy stored in the magnetotail. At that point in history, a picture of the substorm disturbance was beginning to emerge - and the story was as follows.

There was general agreement that a substorm reflected an interval of enhanced energy input into the magnetosphere from the solar wind, typically because the interplanetary magnetic field acquired a more southward component. The merging of the IMF with the frontside field lines (FIGURE 9) led to magnetic flux being stored in the tail. The return of this stored flux to the dayside magnetosphere was thought to occur sporadically, with each burst leading to a polar magnetic substorm of the type shown in FIGURE 3. However, whenever there was an effort to correlate an individual polar magnetic substorm with an increase in southward IMF B_z , things didn't quite work out. In addition, the concept demanded that the substorm expansive phase reflect a conversion of magnetic flux in the tail to particle energy, however there was precious little evidence showing that a substorm expansive phase was accompanied by a reduction in tail lobe magnetic field. What had become clear to Perrault and Akasofu was that the directly driven system tracked the energy input from the interplanetary medium rather nicely. This was confirmed later on by Bob McPherron and Bob Clauer who linear prediction filtering techniques to demonstrate that at least 50% of the disturbance

associated with the energy input from the solar wind could be attributed to the directly driven process. The remaining question was, of course, to evaluate the role played by the substorm expansive phase activity in the overall energy budget of the magnetospheric substorm. And here we come to a real bind that many substorm researchers put themselves into - and a bind that still find themselves in.

Putting it bluntly, if one wants to study substorms, you have to look at the original data. There is often a tendency in our field to believe that one can make statistically based parameters and then believe that the parameters so created have some quantitative meaning in terms of the system under study. In space physics, it was felt to be much simpler to study an index of geomagnetic activity as a proxy measure of the magnetospheric activity levels, rather than to look at the individual magnetometer records which constitute the true reality. In the case of substorms, the chosen index was AE. Most everyone will go to a compendium of AE indices to establish whether or not there was a substorm and when it started. In order to put the technique in perspective, I should first of all make sure that you know exactly how the AE index is derived. The idea is as follows - FIGURE 10 shows a map of the northern hemisphere. The AE index is constructed from the north-south component of the perturbation field at a set of 12 observatories arrayed around the globe at average auroral zone latitudes. All the records are plotted on one plot (FIGURE 11) and the upper envelope (called AU) and lower envelope (called AL) are traced out with values of AU and AL being provided for each minute. AE is the sum of the absolute values of AL and AU. A sample set of traces of AU, AL and AE are shown in FIGURE 12. One generally says that there is a substorm if AE rises and, sometime later falls back to pre-substorm values. There is no particular threshold defined below which one says that a rise and subsequent fall does not constitute a substorm. Thus whether or not one believes a substorm has taken place differs from researcher to researcher depending on their personal thresholds. While that adds an element of subjectivity to the definition of a substorm using AE, there is a far more serious problem. This lies in the fact that AE

does not distinguish between the contributions of the directly driven and current wedge (expansive phase) contributions. Where this constitutes a problem is when an expansive phase takes place during large amplitude directly driven activity. The wedge contribution to the perturbation magnetic field usually maximizes near midnight while the driven system contribution peaks near dawn and dusk (for AL and AU respectively). FIGURE 13 shows that it is possible to have a wedge activated suddenly near midnight, but if its contribution to the north-south component of the field does not exceed the contribution of the driven system, you will never know that an expansive phase took place! It is rare that you can actually separate the contributions of the wedge and driven system, but FIGURE 14 shows a case where it was possible to do so. Driven system activity began to grow at ~ 0730 UT associated with a marked southward turning of the IMF and the onset of expansive phase activity was clearly identified at ~0818 UT as can be seen from the midnight sector magnetograms presented in FIGURE 15. Several things are worth pointing out here. First, it is clear the expansive phase activity commenced when the IMF became northward; thus the energy must have been stored prior to onset, because the solar-wind source was shut off just at the time of onset. Secondly, the midnight sector magnetograms show almost no precursory activity prior to expansive phase onset. This makes sense because the driven system electrojets both peter out close to midnight, where the wedge contribution maximizes. The stations contributing to AU and AL are far from midnight, near dawn and dusk where you would expect them to be.

The moral of this story is clear - defining substorms using the AE, AU and AL indices is inappropriate if you want to study individual events in any way which will advance our physical knowledge of the substorm process. The index is useful for identifying periods of interest, and giving some measure of the general activity level. It also may have some operational uses for those who don't want anything more than a general measure of activity levels in near real time. The serious researcher should, however, be prepared to go back to the original data (viz. magnetograms) if they want to add to the knowledge base about substorms.

In concluding this discussion of the magnetic signatures of substorms, I would like to refer back to the auroral signatures and show the substorm in terms of its two components - directly driven activity and the storage-release process. FIGURE 16 shows the behavior of the oval during a period from the time of a southward turning of the IMF to the recovery of the oval after the IMF turns back to the north. The whole oval expansion and contraction follows the IMF southward turning and its subsequent return to pointing northward. The expansive phase activity is often triggered by the northward turning of the IMF (cf. FIGURES 14 and 15) although it may also be internally triggered.

Now that I have, I hope, convinced you that there are two components of substorm activity, each with its associated current system, I would like to show you some real data, which exemplifies what I have been talking about. I'll be using data from the CANOPUS array of stations which has been operating in Canada over the past few years (FIGURE 17). We were, with this array, able to catch the onset of the substorm expansive phase, using the allsky imager located at Gillam (FIGURE 18). The auroras at onset exhibit a quasi-periodic azimuthal structure as the activity spreads poleward. The actual poleward motion can be traced using meridian scanning photometers, and the record from Gillam is shown in FIGURE 19. Insofar as the electric currents are concerned, FIGURE 20 shows magnetograms from the CANOPUS array, with the Churchill line (from PINA to RANK) showing the poleward movement of the disturbed region. You can see, both from the scanning photometer and the magnetometer data, that the poleward edge of the disturbed region expanded rapidly poleward and then seemed to reach a latitude where it was satisfied to stick around for a while. During the period from 0356 UT to 0416 UT the electric current system experienced quasi-periodic intensifications. Note that the onset indicated in the photometer record (FIGURE 19) occurs very close to a region of hydrogen emission marked by 4681 Å luminosity. These emissions are caused by rather energetic protons (viz. tens of keV) precipitating into the upper atmosphere, and you will only find such energetic protons rather close to the Earth - certainly inside ~ 12 Re or so. Thus, the onset of the

Insofar as the physics of the substorm process is concerned, the problem is still open. There are several models for substorms, ranging from the time honored "near-earth neutral line paradigm" to a whole host of competitors. To discuss these models and to critique them properly would take more time than this lecture can provide. If you look further into this question, it may not take you long to realize that the question of substorm physics is still far from closed and you might consider applying your data sets and ingenuity to closing the book on this fascinating problem.

expansive phase occurs on quasi-dipolar field lines in a region which ought to map rather close to the equatoward edge of the nightside auroral oval. In fact, there is increasing evidence that the onset occurs on field lines threaded by Region 2 currents. What do I mean by this? FIGURE 21 shows an interesting way of viewing the magnetosphere-ionosphere interaction. One can view the solar wind - magnetosphere interaction as a cause of the development of regions of space charge. The velocity shear across the interface between the low latitude boundary layer and the central plasma sheet is equivalent to space charge which produces the dawn-to-dusk electric field associated with earthward convection of plasma. A consequence of the earthward convection of plasma is the development of a regime of shielding space charge closer to the earth. Part of this may be associated with velocity shear in the interface between earthward convecting plasma in the plasma sheet and the relatively stationary plasma inside the plasmapause. Part is due to the presence of the ring/crosstail current and the role it plays on the earthward convecting plasma. Figure 22 shows one of the possible circuits involving the discharge of this shielding space charge. The sudden disappearance (or at least sharp decrease) of the crosstail current near the inner edge of the plasma sheet may be associated with the rapid discharge of the shielding space charge and this may be the main physical process involved in the onset of a substorm expansive phase.

In this talk, I have tried to discuss the magnetic signatures of the magnetospheric substorm, and to convince you that:

1. You have to understand these magnetic signatures as the combined effect of the directly driven current system and the substorm current wedge. [The reader is referred to Rostoker et al., Space Sci. Rev., 1987 for a more detailed discussion of the directly driven and the storage release process responsible for the substorm current "wedge".]
2. You really ought to look at the original magnetic records rather than indices if you want to improve your knowledge of the substorm process.

Figure Captions

Figure 1 A magnetic storm as detected at a low latitude observatory. The disturbance detected at low latitudes is primarily that of a ring current, with the particles responsible for the ring current being energized by processes which appear to be associated with substorms.

Figure 2 The auroral substorm, shown as a sequence of arc brightenings and motions. The expansive phase occurs over the first hour in this figure, with the recovery phase following the expansive phase (after Akasofu, Planet Space Sci., 1964).

Figure 3 A sequence of four polar magnetic substorms detected at the high latitude ($> 70^{\circ}$ N) stations of Bjornoya and Isfjord. which both lie on a common magnetic meridian. The differences in the perturbations indicate that these disturbances are relatively localized (after Rostoker, JGR , 1969).

Figure 4 Latitude profile along a meridian crossing (at right angles) a westward electrojet. The polarity reversal in the Z-component perturbation marks the center of the ionospheric electrojet. Checking the polarity of the H and Z components can permit one to establish the position of the observing site relative to the center of the electrojet.

Figure 5 The two cell and one cell model for the substorm. History has shown that both patterns coexist, the single cell being representative of the substorm current wedge and the two cell pattern well describing the directly driven system.

Figure 6 A lower latitude magnetogram for the event shown in Figure 3, where the long period disturbance in the H-component is seen to be caused by the directly driven westward electrojet (after Rostoker, JGR , 1969).

Figure 7 The substorm current wedge (after McPherron, JGR, 1972).

Figure 8 The substorm current wedge (after Birkeland, 1908).

Figure 9 The flow of magnetic flux tubes during a period of magnetic field merging on the front side magnetopause and reconnection in the magnetotail .

Figure 10 The AE stations. These are located at average auroral oval latitudes. For contracted ovals during quiet times and expanded ovals during active times, the AE stations often underestimate the level of disturbance and miss certain expansive phase onsets.

Figure 11 The method by which AE, AU and AL are obtained. These indices represent the envelope of the disturbances in terms of the positive (AU) and negative (AL) maxima at any instant from any station in the array.

Figure 12 Typical AE, AU and AL indices. Most researchers would say that substorms took place between 0900-1200 UT and between 1900-2200 UT, although the actual identification would be subjective in nature as there is no commonly accepted threshold below which one says that a substorm did not take place.

Figure 13 Same as Figure 12, except a single H-component record from a midnight sector observatory is superposed on the record of the indices during the 0900-1200 UT substorm. This event would not register in AL or AE because the peak perturbation did not exceed the contribution to the index from the directly driven system. This contribution usually comes from a station in the dawn sector.

Figure 14 A case where the directly driven and current wedge contributions are separable because of good ground coverage over and above that provided by the AE stations. The directly driven system is dominant from ~0730-0815 UT, and a expansive phase took place thereafter (after Rostoker, JGR, 1983).

Figure 15 Magnetograms from the midnight sector stations of the University of Alberta array for the same day for which the indices are shown in Figure 14. No expansive phase activity is apparent in the midnight sector while the directly driven system is providing a significant contribution to AE from ~0730-0815 UT. Note that the driven system grows at a time of southward IMF, while the expansive phase occurs after the IMF has switched to being northward (after Rostoker, JGR, 1983).

Figure 16 The behavior of the auroral oval during a substorm. The equatorward expansion of the oval takes place during the growth phase the substorm, while the poleward contraction is a characteristic of the recovery phase. The auroral substorm, as described by Akasofu (see FIGURE 2) actually corresponds to the expansive phase of the substorm as exemplified by the development of the current "wedge". The expansive phase features poleward motion of the region of disturbed auroras over a longitudinally confined region, while the whole oval contracts poleward during the recovery of the directly driven system. Auroral surges, which appear as small auroral distortions in this figure, can easily cover the entire field of view of an allsky camera. Thus the original defining characteristics of a substorm did not incorporate the behavior of the directly driven system.

Figure 17 The CANOPUS stations. This station array has been in full operation since late 1989 and will provide ground based coverage during the Solar Terrestrial Energy Program (STEP).

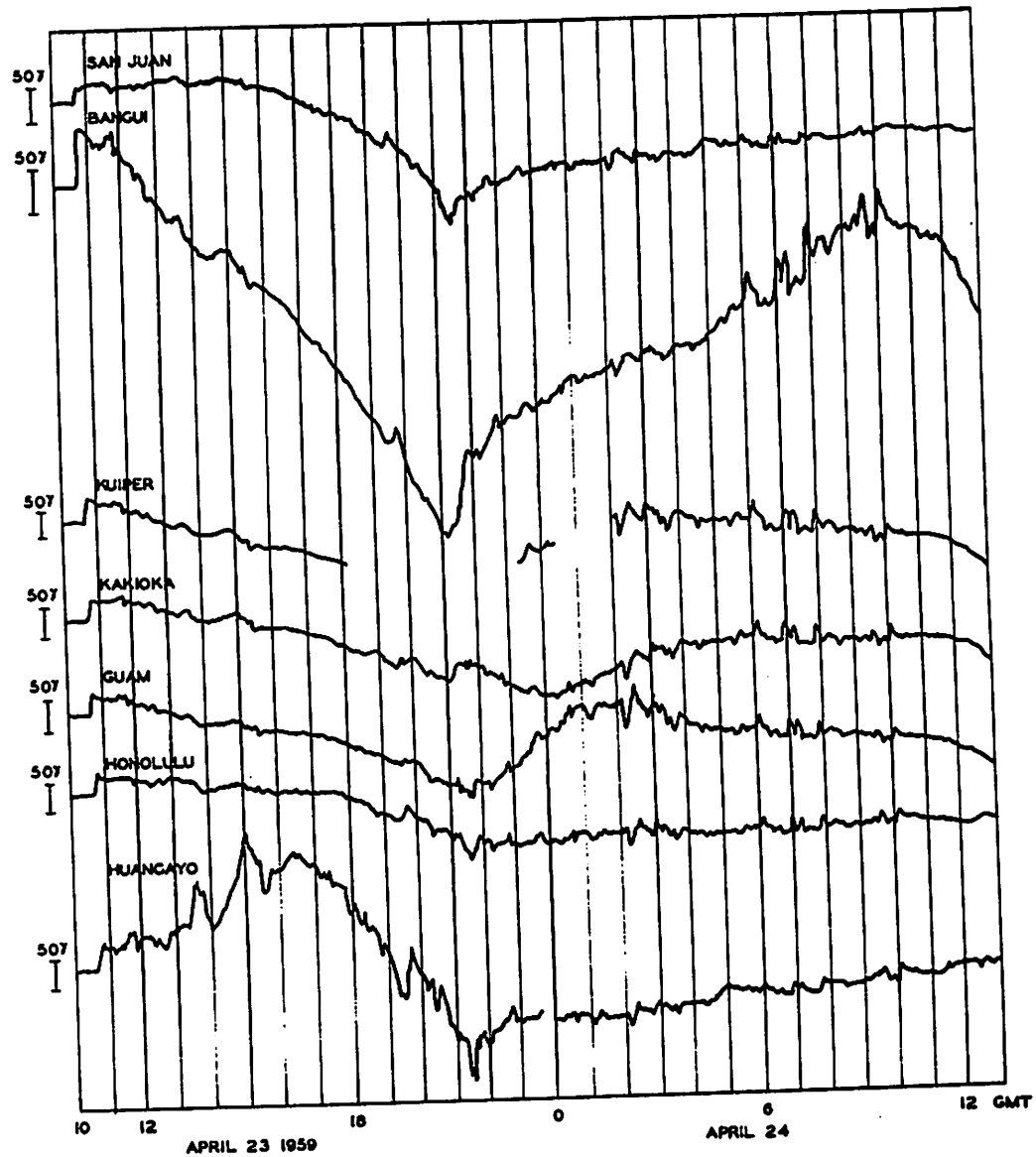
Figure 18 Allsky image recording of an expansive phase onset on February 1, 1990 recorded at Gillam (GI in FIGURE 17).

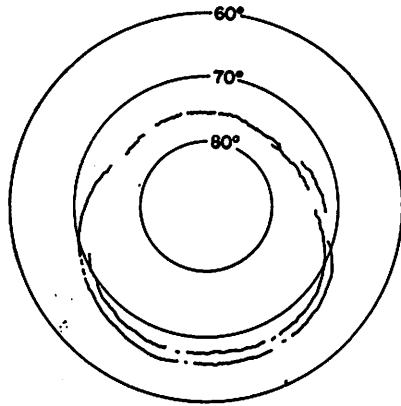
Figure 19 Meridian scanning photometer record of the substorm on February 1, 1990 taken at Gillam (GI in FIGURE 17) . The substorm took place in two phases, with the initial disturbance involving poleward motion starting at ~0356 UT and stalling until ~0416 UT after which time explosive poleward motion caused the oval to contract.

Figure 20 Magnetograms for the event of February 1, 1990. The poleward motion is clearly evident, even in the single (H) component records from the stations along the meridian through Fort Churchill.

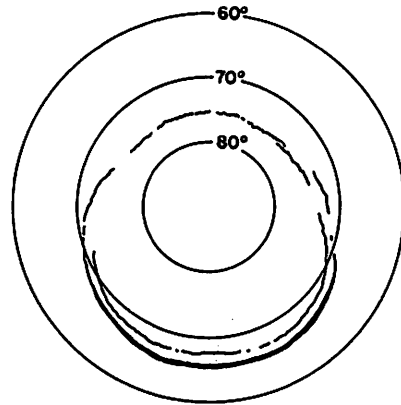
Figure 21 A space charge viewpoint of the magnetosphere. Energy input from the solar wind leads to vorticity at the interface between the low latitude boundary layer and the central plasma sheet. The inner magnetosphere is shielded from this primary convection electric field by a space charge buildup near the inner edge of the plasma sheet.

Figure 22 A possible scenario for substorm initiation, in which onset represents a discharge of accumulated space charge near the inner edge of the crosstail current.

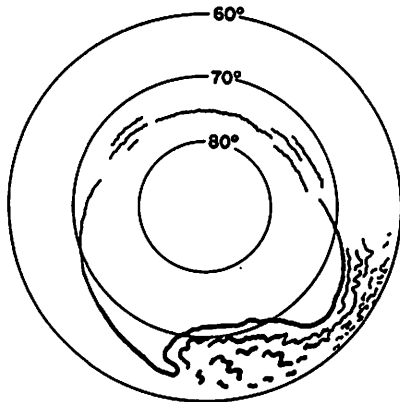




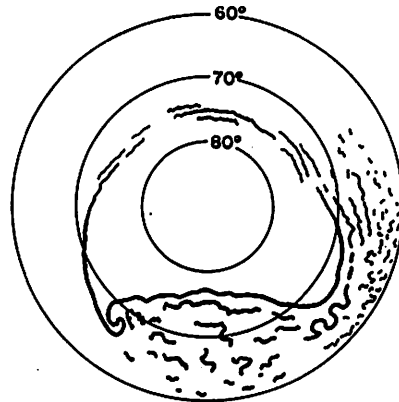
A. T=0



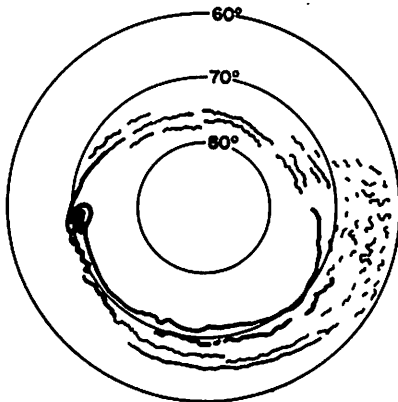
B. T=0~5 MIN



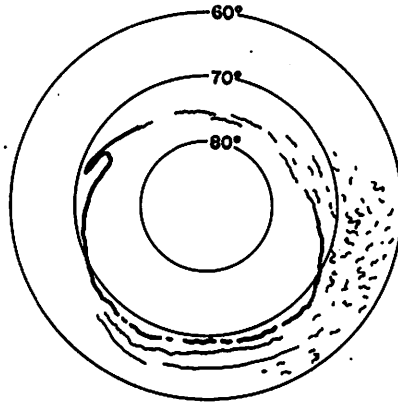
C. T=5-10 MIN



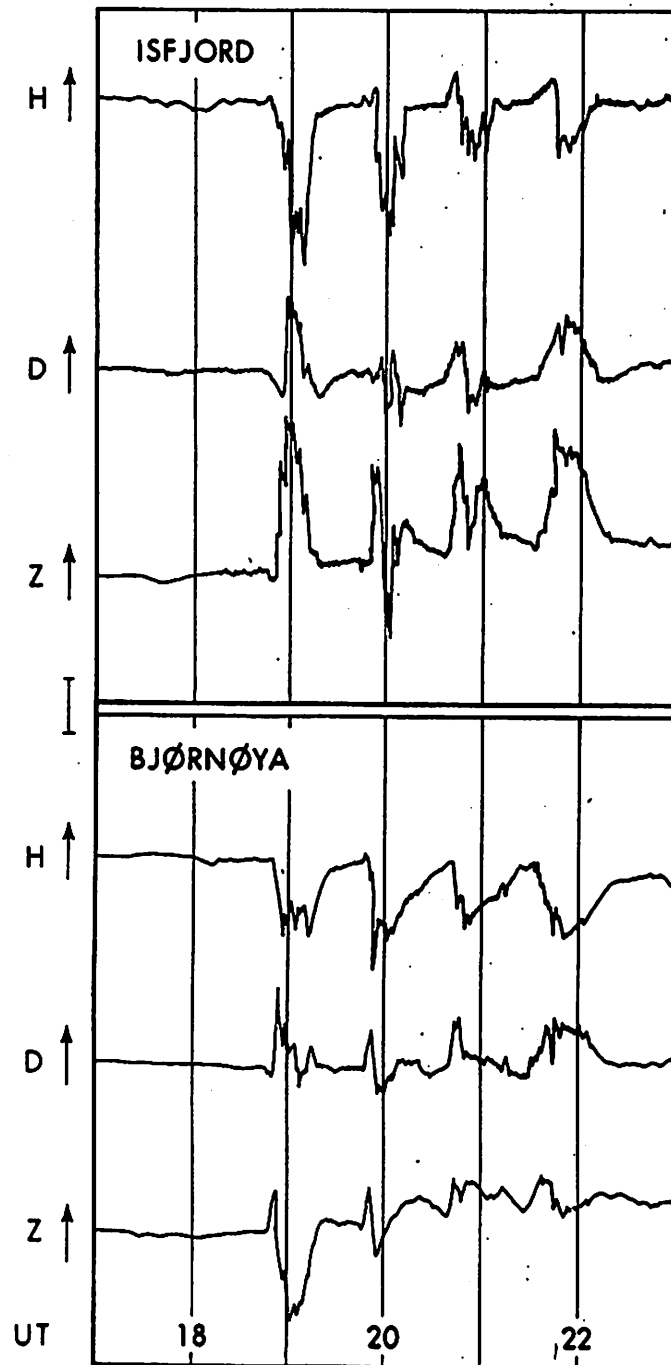
D. T=10-30 MIN

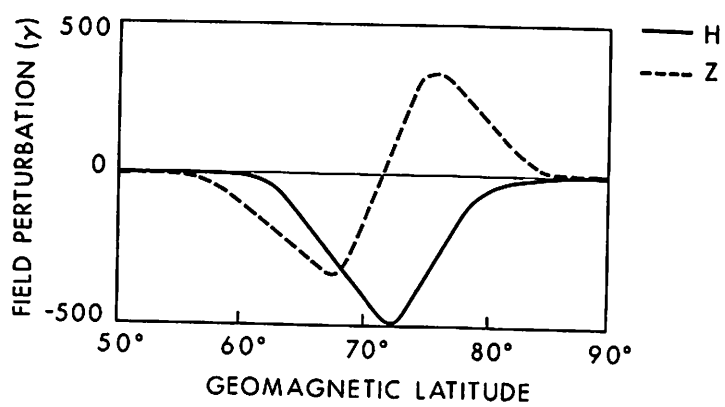
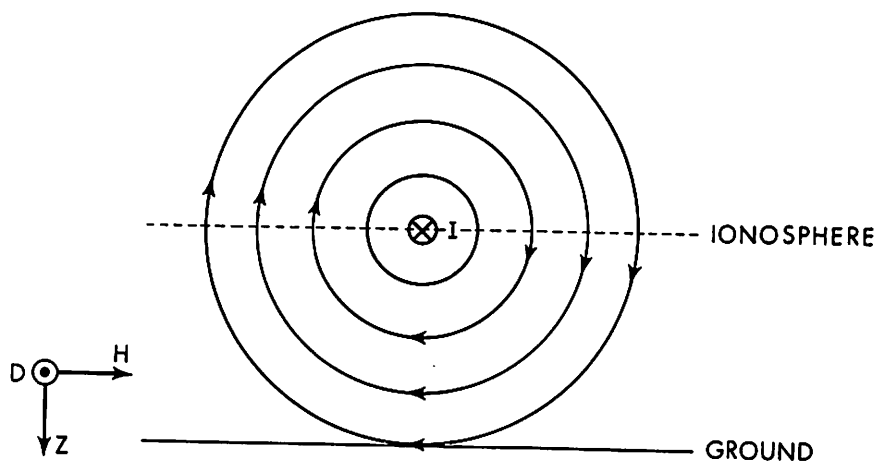


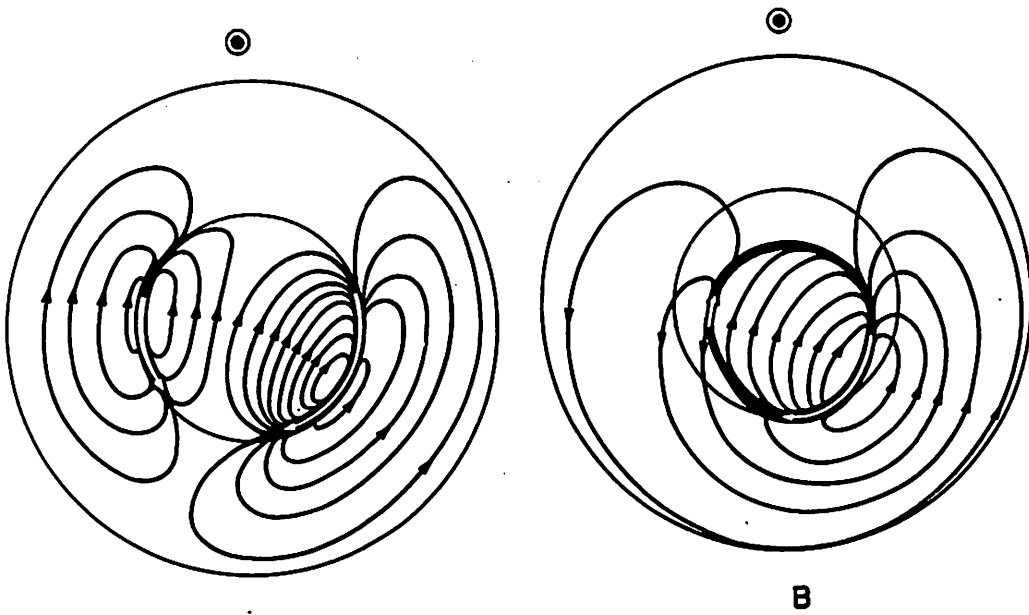
E. T=30 MIN-1 HR

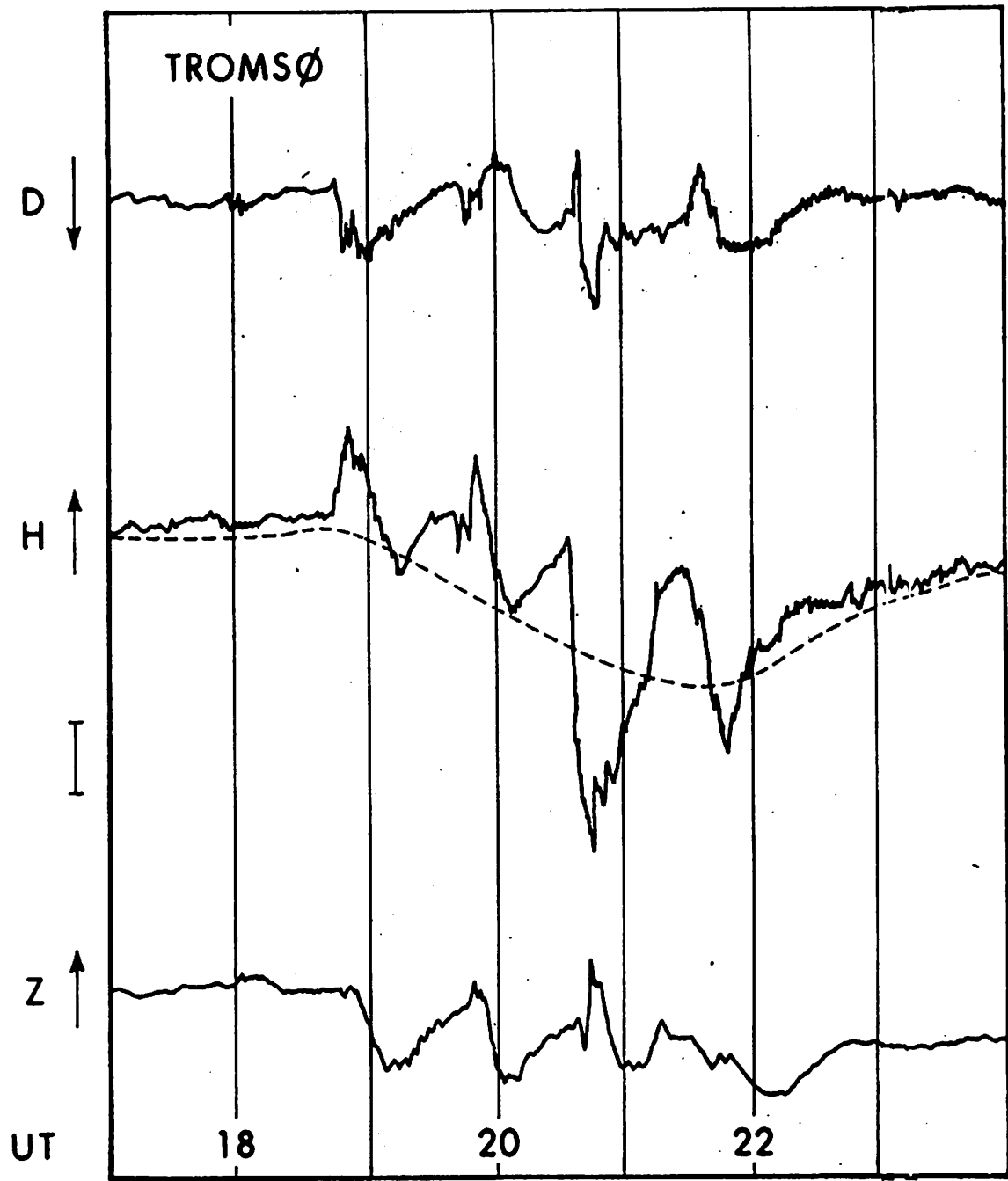


F. T=1-2 HR

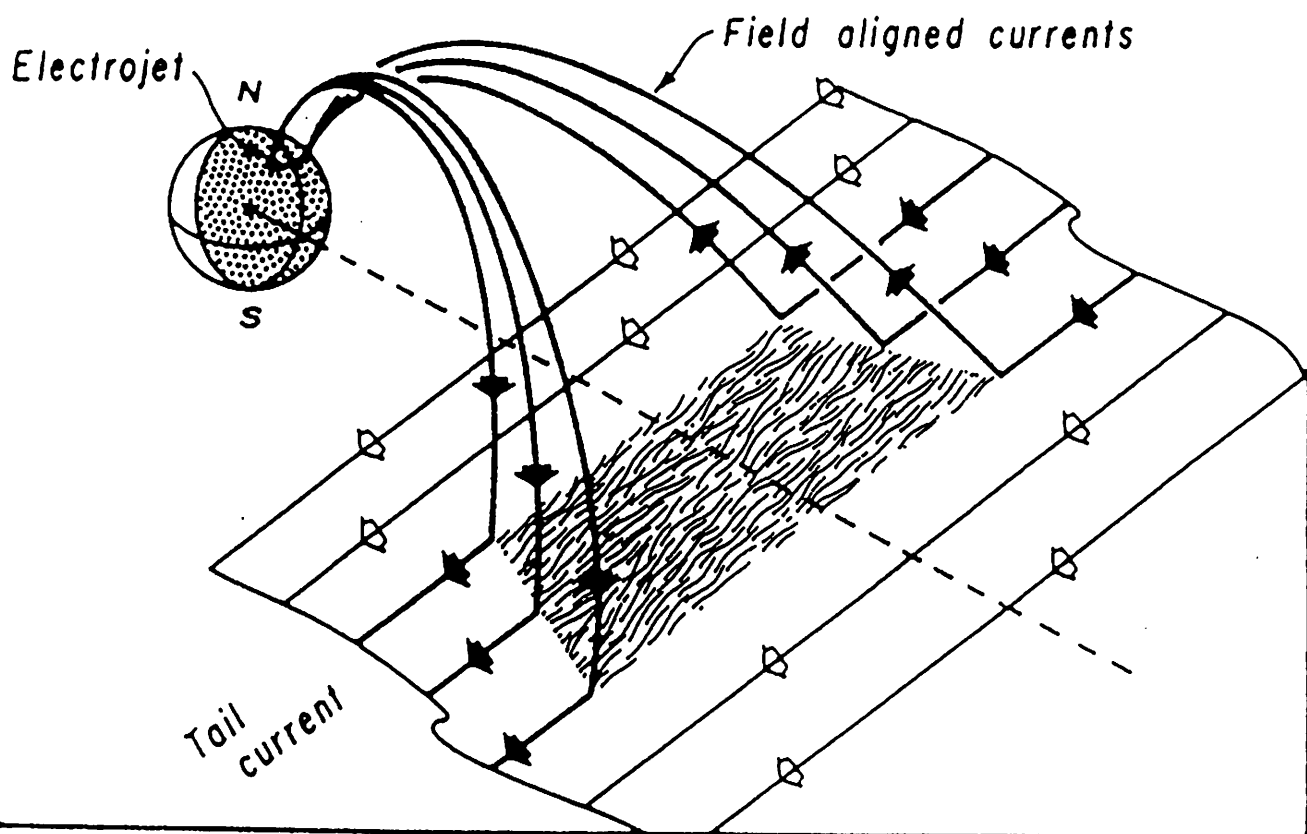


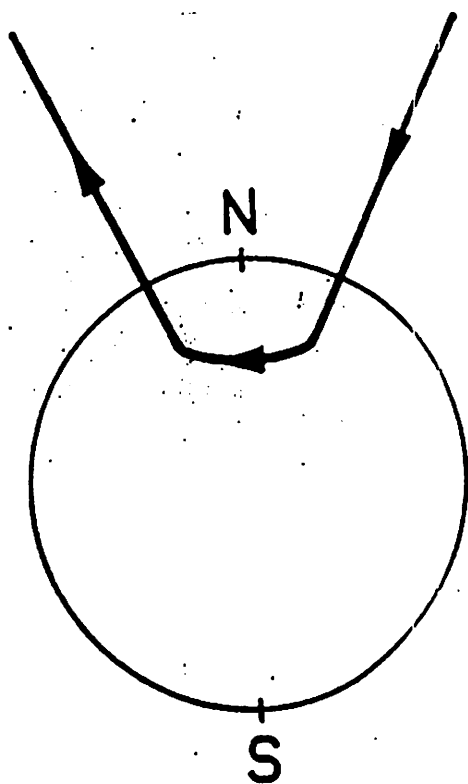


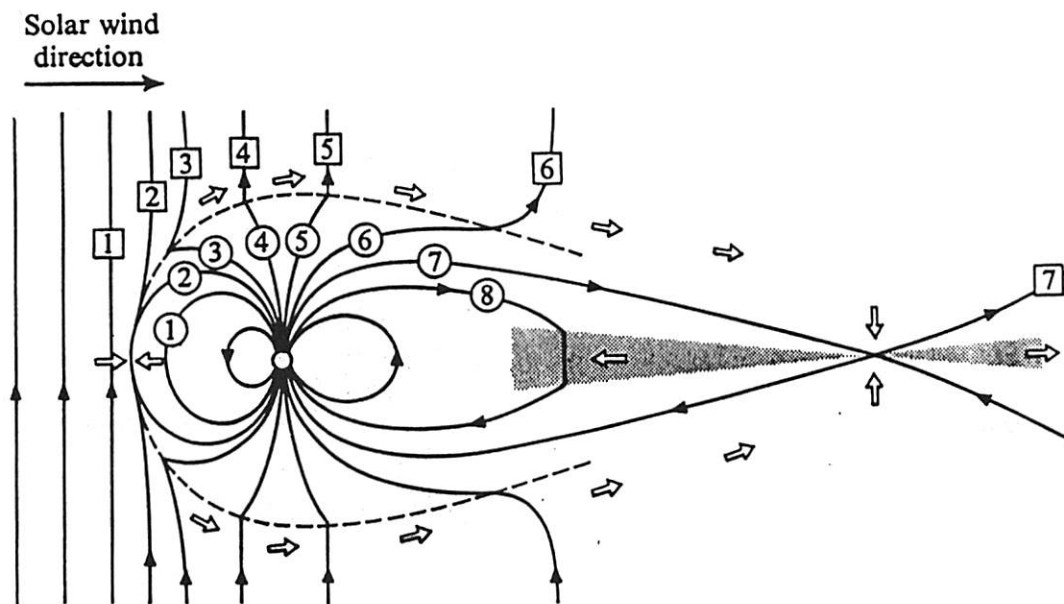




Expansion Phase Short Circuit of Tail Current







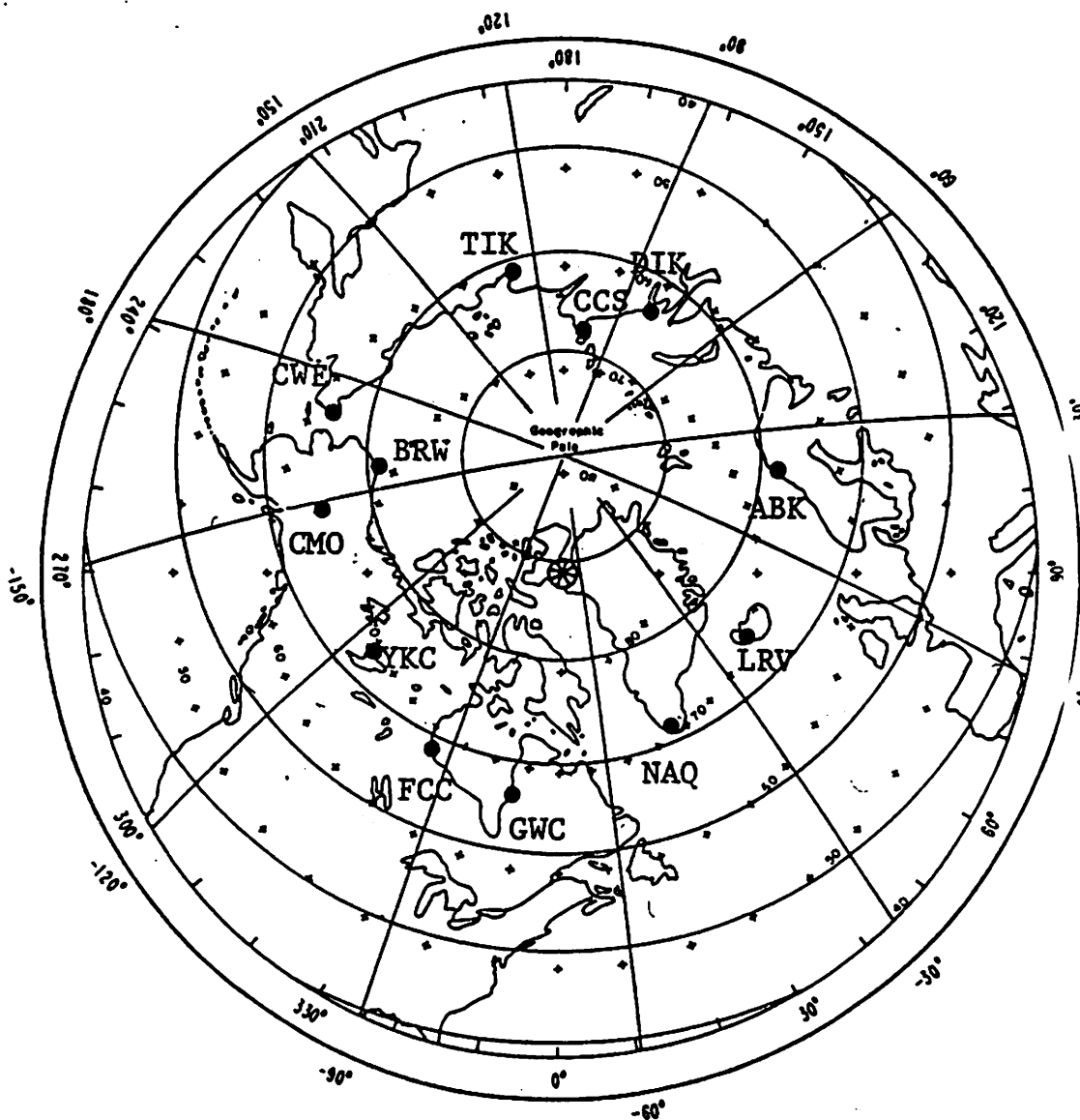
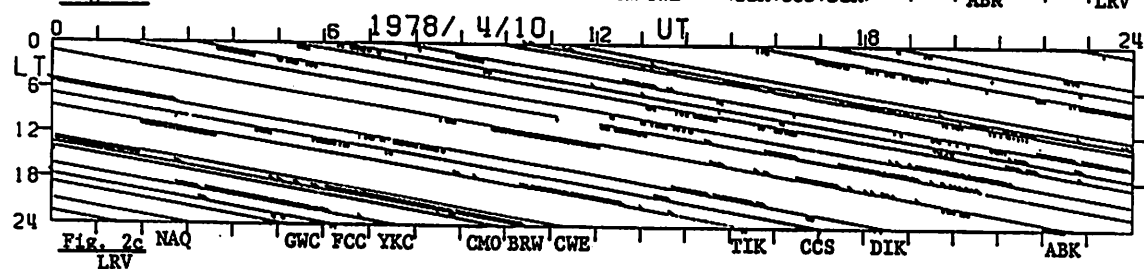
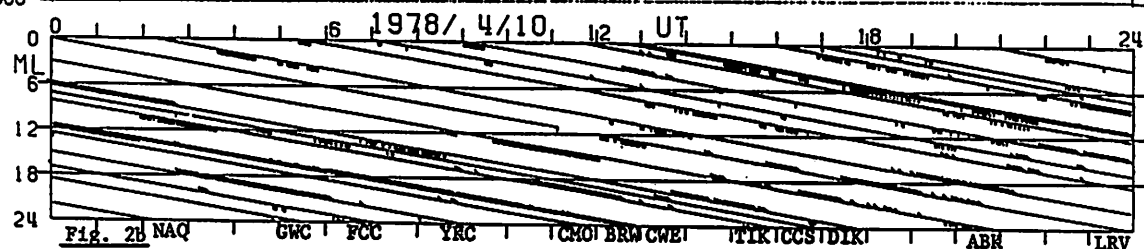
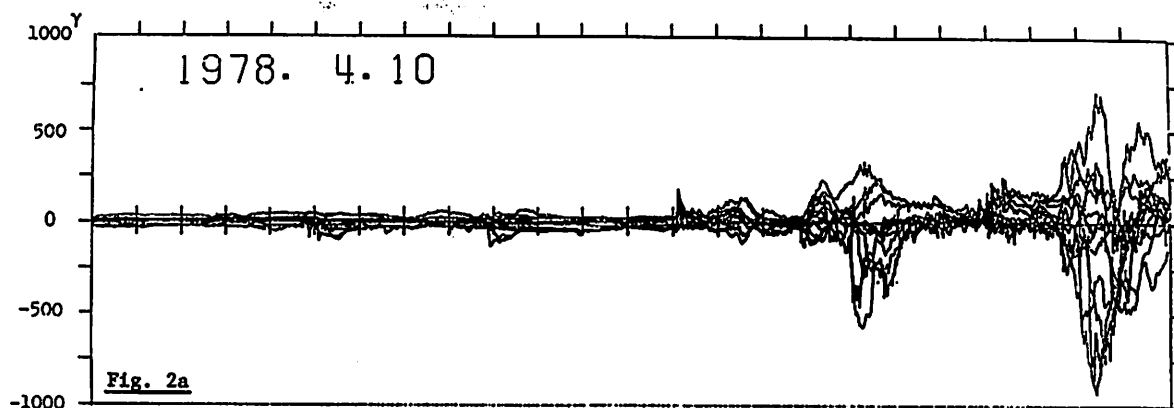
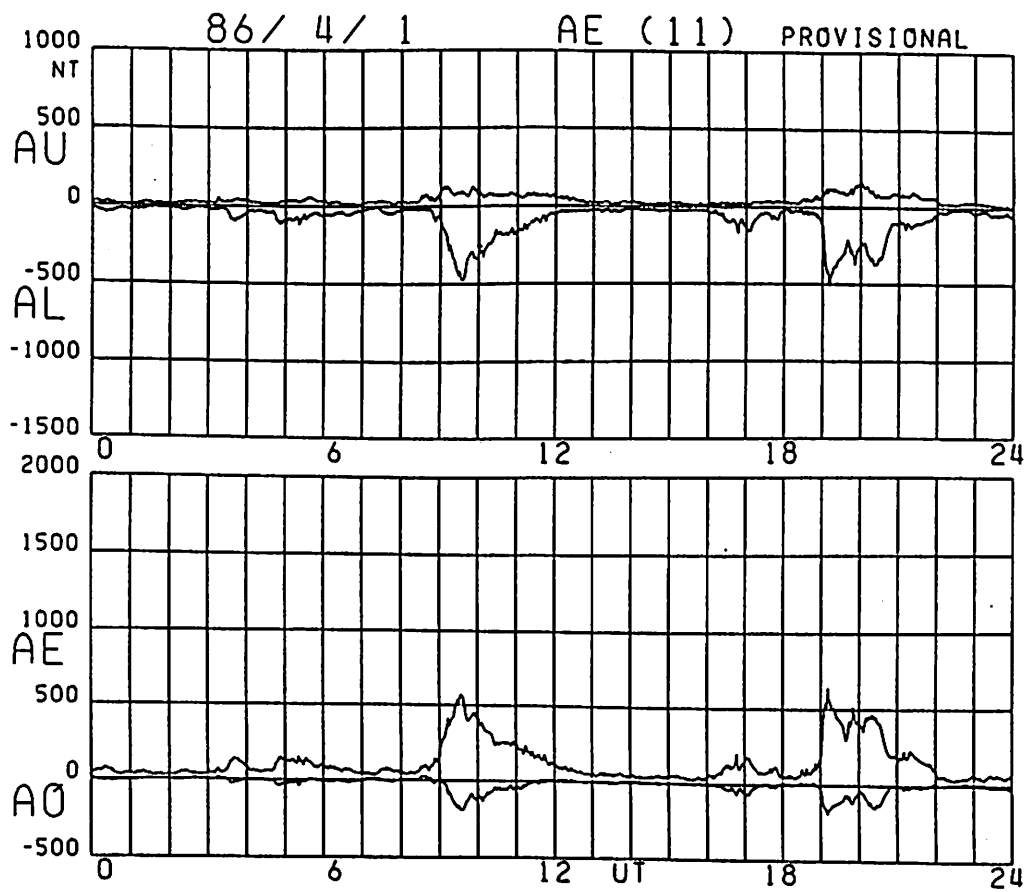
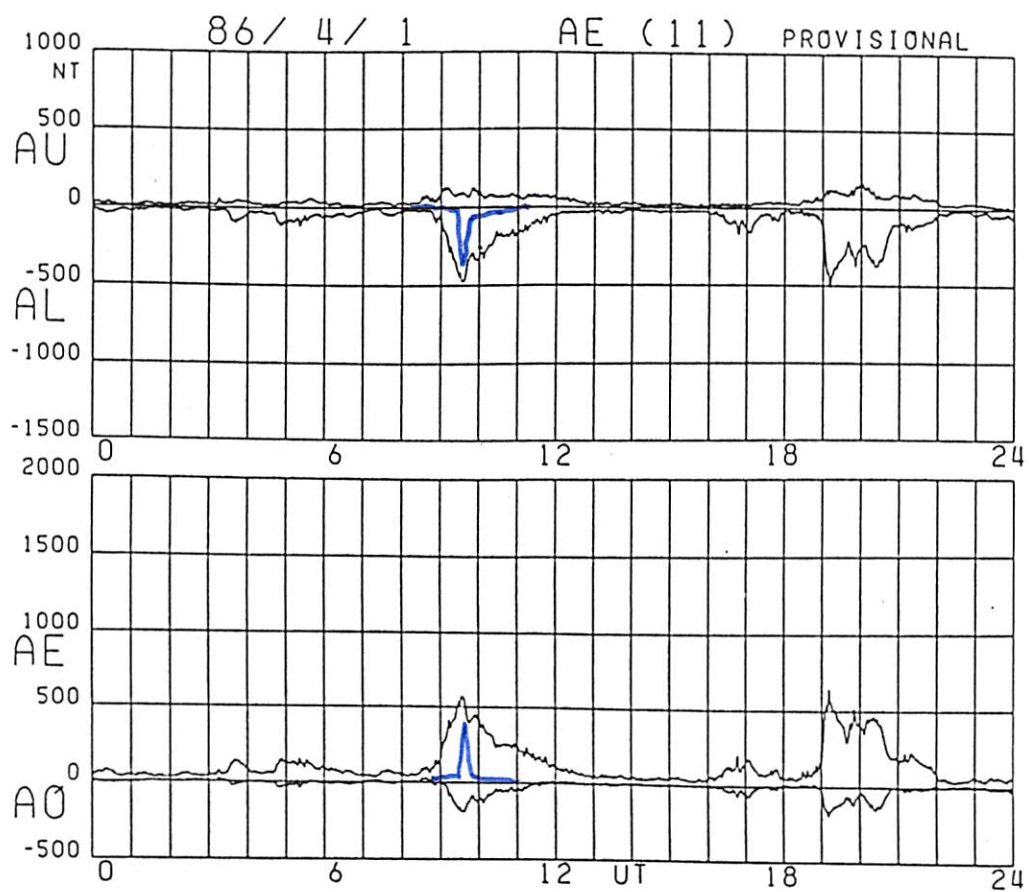
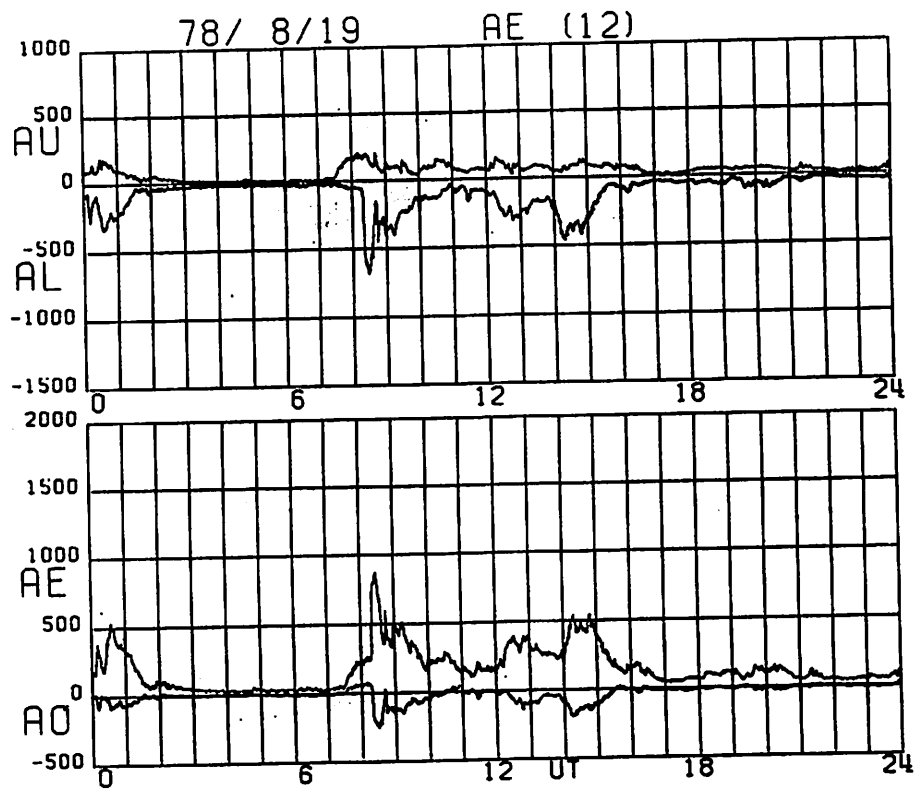


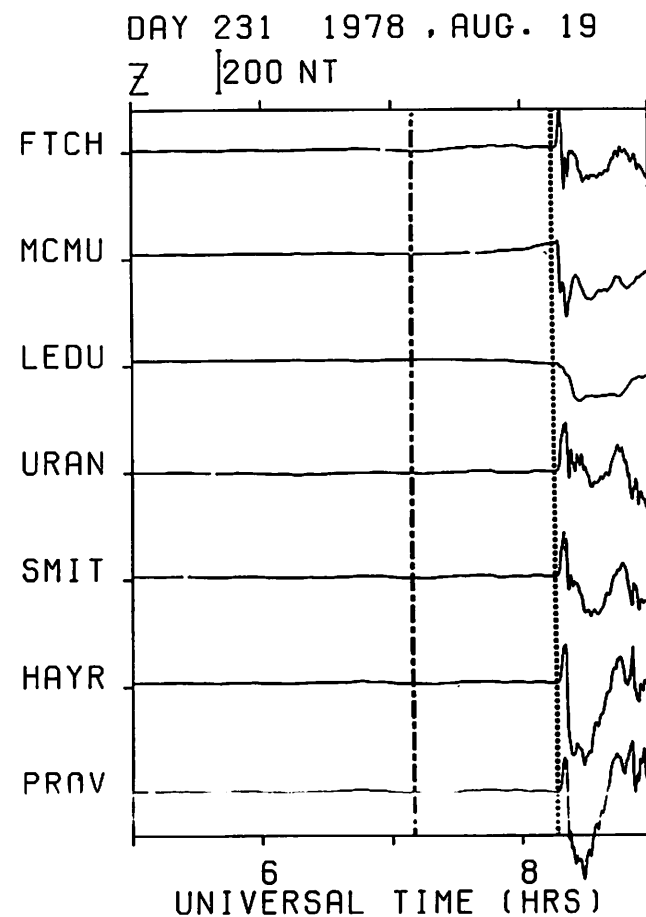
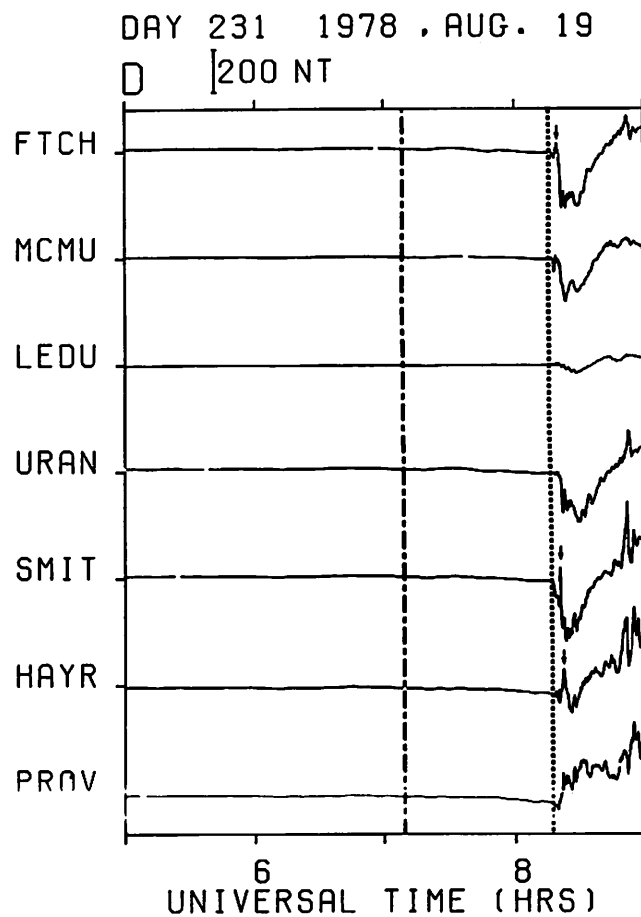
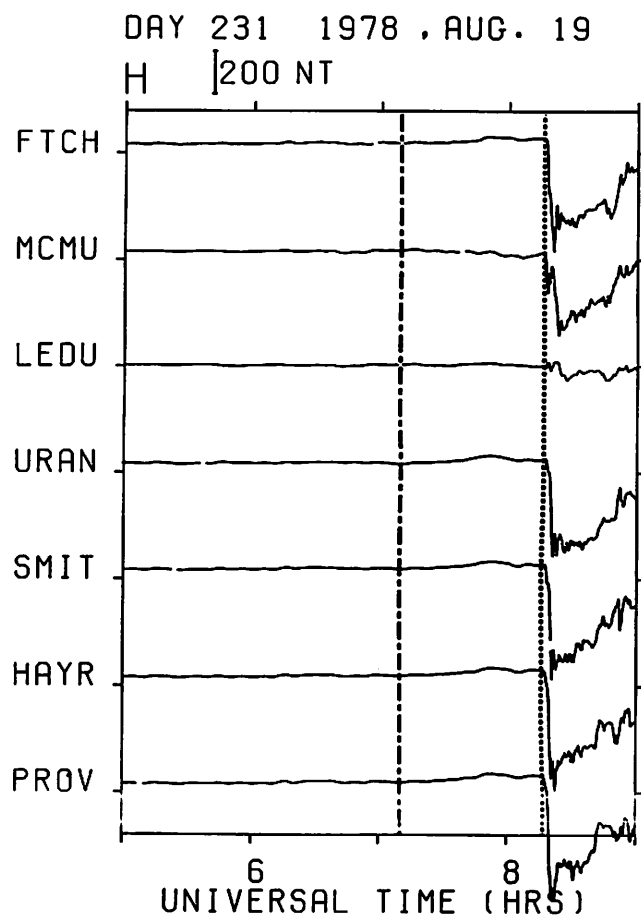
Fig. 10 Distribution of AE(12) stations.

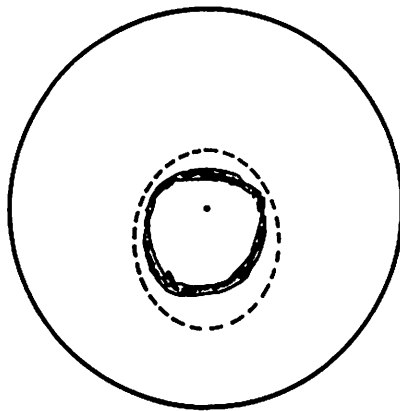




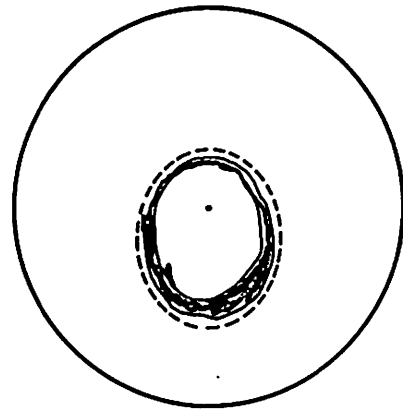




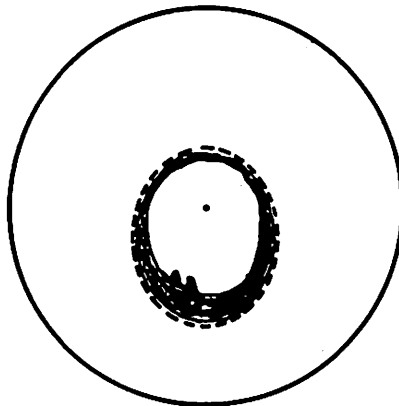




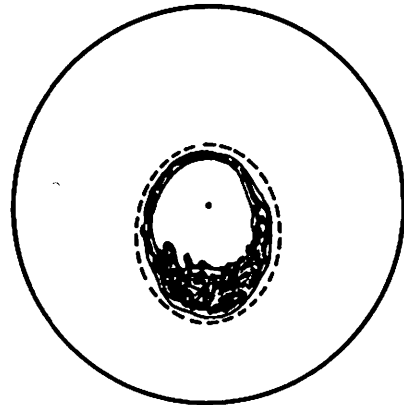
IMF after being northward for some time has just turned southward



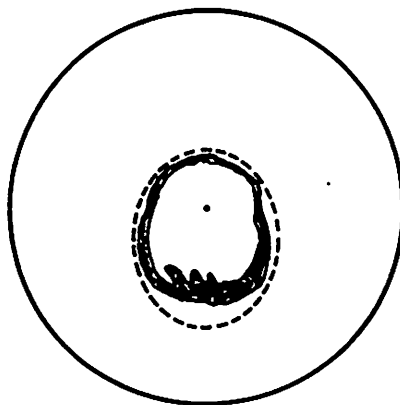
Oval expands equatorward under influence of southward IMF



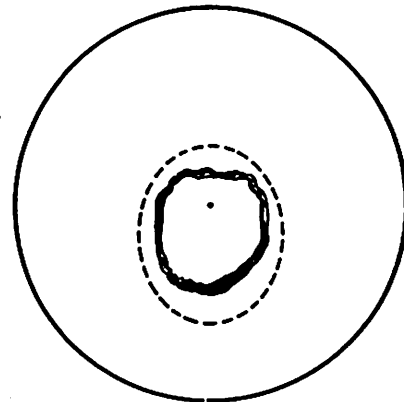
IMF has just turned northward



Poleward edge of oval expands poleward and features multiple surges. Bright structures seen within diffuse evening auroras



Surge action at poleward edge of oval continues as diffuse auroras recede poleward

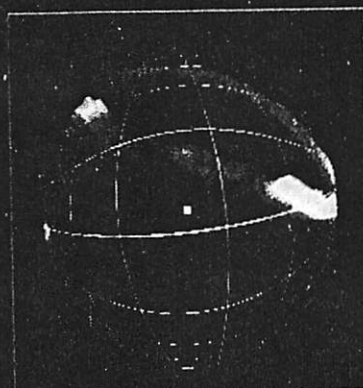


Return to quiet oval about 1 hr after IMF northward turning

5577
 10^3 kR
 10^2 kR
 10^1 kR
 10^0 kR



90: 32 3:54:49 UT
 5577 EXP= 1.7 s



90: 32 3:55:44 UT
 5577 EXP= 1.7 s



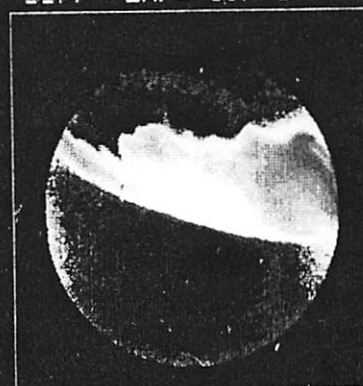
90: 32 3:56:39 UT
 5577 EXP= 1.7 s



90: 32 3:57:34 UT
 5577 EXP= 1.7 s



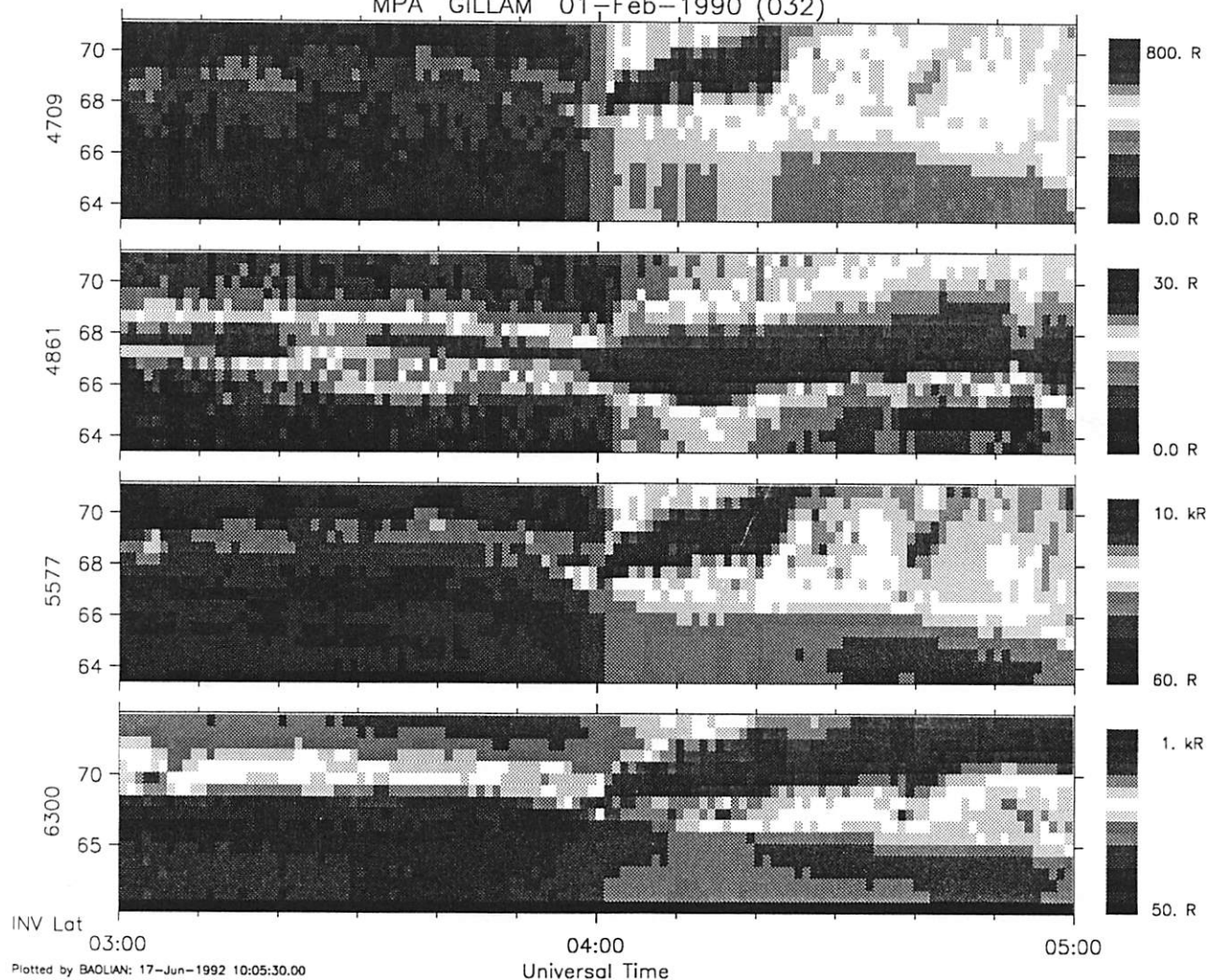
90: 32 3:58:29 UT
 5577 EXP= 1.7 s



90: 32 3:59:24 UT
 5577 EXP= 1.7 s

GILLAM ASI

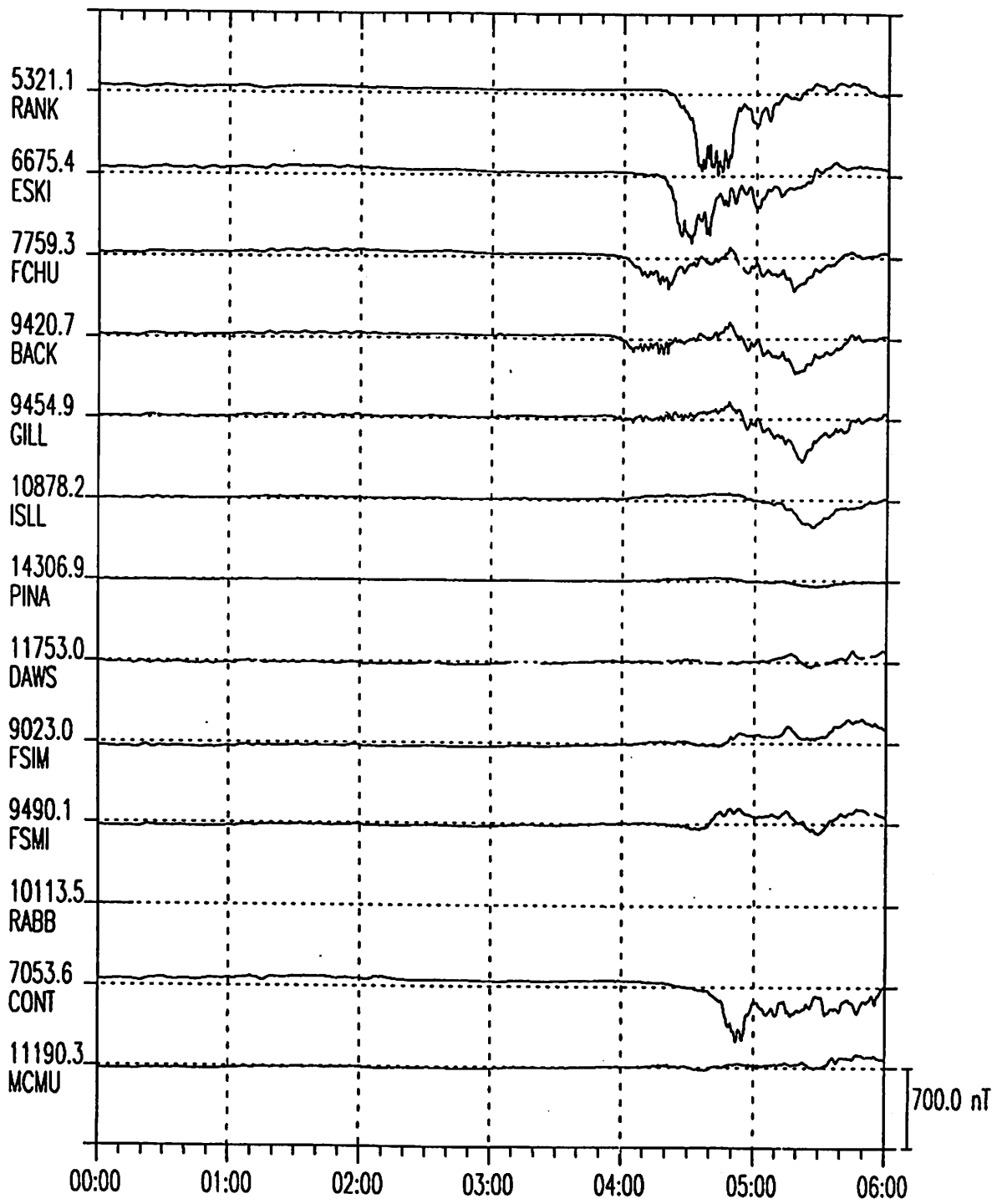
MPA GILLAM 01-Feb-1990 (032)



Year:1990 Day:032 X

BETA Data

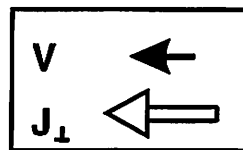
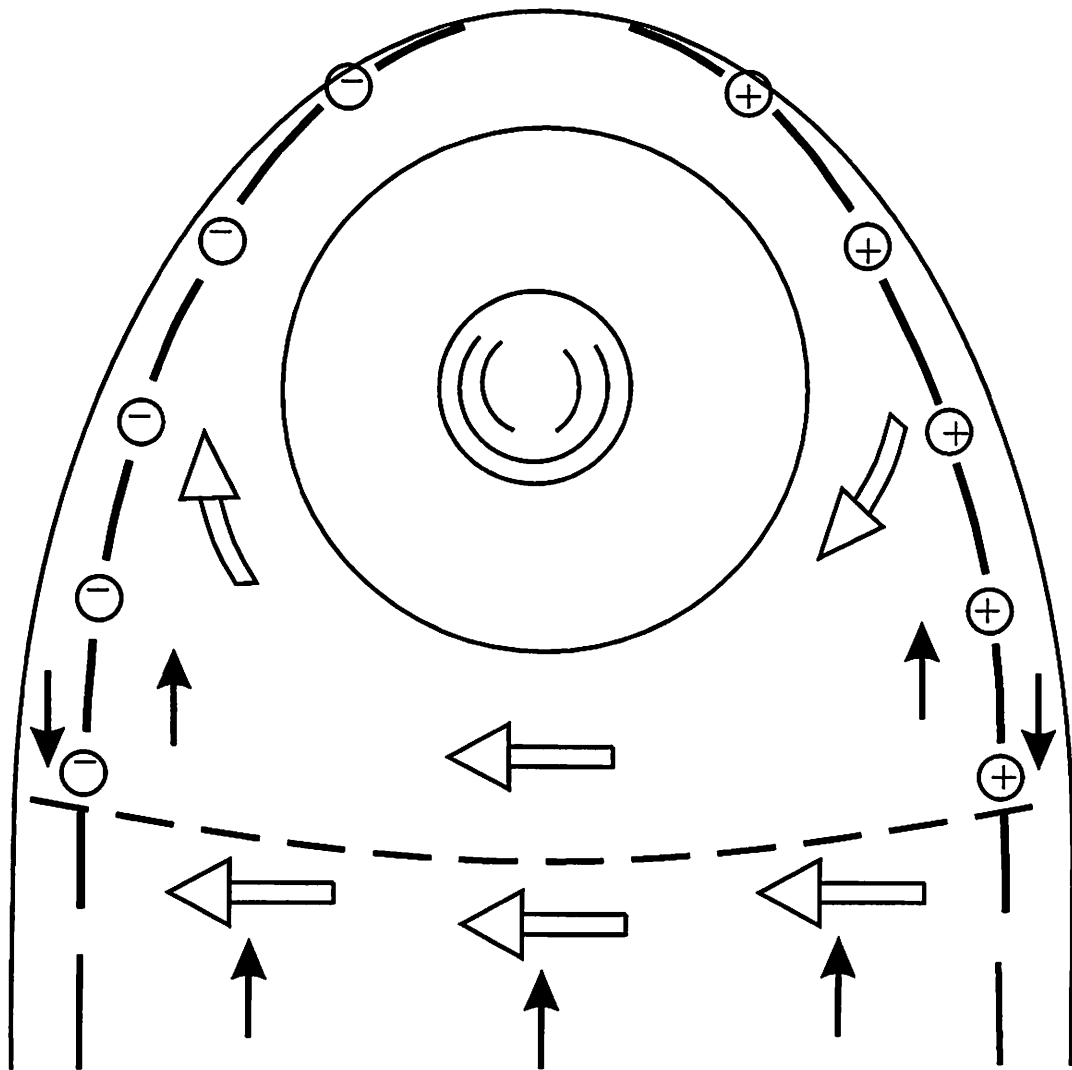
Start Time 1990/02/01 00:00:00 UT



Plotted: 21-AUG-1990 15:40:52.04 UT

Plot Decimation = 2

20



$$\begin{aligned}
 \rho &= \epsilon_0 \nabla \cdot \mathbf{E} = -\epsilon_0 \nabla \cdot \mathbf{v} \times \mathbf{B} \\
 &= -\epsilon_0 \mathbf{B} \cdot \nabla \times \mathbf{v} + \epsilon_0 \mathbf{v} \cdot \nabla \times \mathbf{B} \\
 &= -\epsilon_0 \mathbf{B} \cdot \boldsymbol{\Omega} + \frac{1}{c^2} \mathbf{J}_\perp \cdot \mathbf{v}
 \end{aligned}$$

