Chapter 1. General Introduction

1.1. Current status of the substorm study

Substorm is one of the most spectacular phenomena in the terrestrial magnetosphere - ionosphere system. In a substorm, large energy stored in the magnetosphere is released explosively, which is manifested with the remarkable change and evolution in the ground magnetic field variation [e.g., *Akasofu et al.*, 1965; *Iijima and Nagata*, 1972; *Kokubun and Iijima*, 1975], auroral activity [e.g., *Akasofu*, 1964; *Elphinstone et al.*, 1996], and the magnetospheric plasma and field properties [e.g., *McPherron*, 1979; *Nishida*, 1998]. Those are called respectively as "polar magnetic substorm", "auroral substorm", and "magnetospheric substorm". The last term is also used to represent the substorm phenomenon as a whole including both the ionospheric and magnetospheric processes. In our study, however, we will use the "magnetospheric substorm" to represent only the magnetospheric processes literally because we would like to stress the importance of the ionospheric processes. Substorm study has a long history of about 40 years. A great amount of works has been done so far, and it is said that the global magnetospheric substorm problem has now largely been solved [*Baker et al.*, 1999]. But it seems that many essential problems still remain unsolved. One of such problems is the relationship between the magnetospheric substorm and the ionospheric, especially, auroral substorm.

1.1.1 Auroral substorm and associated magnetospheric process

(a) Development of the global auroral imager

Since the pioneering work by *Akasofu* [1964] using the ground-based all-sky camera observation, the global scale auroral substorm has been studied with various satellite-borne imagers. Table 1.1 lists the characteristics of the imagers and the satellites in the past and currently running projects. Though the spatial and temporal resolutions of the satellite-borne imagers are generally lower than those of the ground-based ones (typically less than 1 km, and 20 sec (1/60 sec) for the all-sky camera (all-sky TV)), great advantages of the satellite imager are its wide coverage of the Field Of View (FOV) including sunlit area for the UV filter, less distorted images, and being free from weather condition. A thorough review of the auroral substorm study was made by *Elphinstone et al.* [1996] up to the VIKING project [*Hultqvist*, 1987]. The VIKING introduced, for the first time, the two-dimensional CCD device for the imager target. This new device enabled us to get the snap-shot images with a higher time resolution compared with the previous scanning techniques in the DMSP (Hardy et al., 1979), ISIS-2 [*Anger, et al.*, 1973], HILAT [*Meng, et al.*, 1984], and DE-1 [*Frank et al.*, 1981] projects. It should be mentioned that the KYOKKO was the first satellite bringing an UV TV camera which could take such a snapshot with a high time resolution of 128 sec [*Kaneda*, 1979; *Kaneda, et al.*, 1981]. The AKEBONO was a successor of the VIKING in many points. It carried both the UV and visible range filtered CCD cameras [*Oguti, et al.*, 1990] whose

Table 1.1. List of the satellite global auroral imager in past experiments.

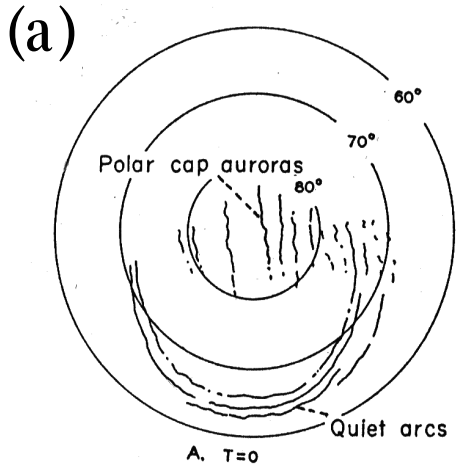
Satellite Name	Wave length (nm)	Technique	Time resolution	FOV spatial resolution	Apogee Perigee	Orbita period	Launch date	Nation PI	Reference
DMSP	400-1130	scanning mirror	1/orbit 101 min	2,960 km 3.7 km	830 km 830 km	101 min	?, 1973 ?	U.S.A.	Gussenhoven, et al, 1979
ISIS-2	557.7/391.4/630.0	spin scanning photometer	1/orbit 113.6 min	8 km - 32 km (0.4Þ)	1,400 km 1,400 km	113.6 min	Apr. 1, 1971	U.S.A. C.D.Anger G.G. Shepherd	Anger, et al., 1973s
KYOKKO (EXOS-A)	VUV115-165	TV camera	128 sec	176 x 198 (total 60Þ, 20 km)	3974.1 km 638.3 km	134.2 min	Feb. 4, 1978	Japan E.Kaneda	Kaneda, et al., 1981
DE-1	VUV(123-155)(12ch) VIS 2 sensor, 12ch, 557.7, 630.0, 391.4	rotating mirror	normal 12 min VIS 6 min max	0.23Þ x 0.25Þ (total 30Þ) 100 km at apogee	3.63 Re 570 km	6h 49m	Aug. 3, 1981	U.S.A. L.A.Frank	Frank, et al., 1981
HILAT	VUV(110-190)	scanning mirror	1/orbit 101 min	4 km x 20 km	830 km 830 km	101 min	Jun.. 27, 1983	U.S.A. C.-I. Meng	Meng, at al., 1984
VIKING	123.5-160 (OI, LBH), 134-180 (LBH, 149.3 NI)	2D-CCD 1sec exp	1/spin (20 sec) normal 60 sec	288 x 385 20Þx 25Þ 20 km x 20 km	14,000 km 822 km	270 min	Feb. 22, 1986	Sweden C.Anger (U.Calgary)	Anger, et al., 1987
Polar BEAR					1,000 km 1,000 km	105 min	Nov. 13, 1986	U.S.A. (APL)	
AKEBONO	UV (115-139) VIS (557.7)	2D-CCD 400ms exp	1/spin (8 sec) 8-8-8-40 sec	244 x 188 UV 36Þx 36Þ, 26 km VIS 30Þx 40Þ, 22 km	10,500 km 274 km	212 min	Feb. 22, 1989	Japan E.Kaneda (UV) T.Oguti (VIS)	Oguti, et al., 1990
FREJA	135-145 (LBH) 130.4 (atoic Oxygen)	2D-CCD 370ms exp	1/spin (6 sec)	144 x 192 25Þx 20Þ, 5 km	1,764 km 596 km	109 min I = 63Þ	Oct. 6, 1992	Sweden J.S.Murphree (U.Calgary)	Murphree, et al., 1994
POLAR	UVI (120-180 VUV, 247-337.1 NUV) VIS (2 VIS (391.4, 557.7, 630.0, 656.3, 732.0), 1 UV (123-149))	2D-CCD 18.4 or 36.8 sec integration	UVI 36.8 sec VIS 12 sec	UVI (8Þ, 35 km) 200 x 224 VIS (20Þx 20Þ, 10 km (VIS), 70 km (UV)) 256 x 256	7.9 Re 185 km	15h 38.1m	Feb. 24, 1996	U.S.A. G.K.Parks (UVI) L.A.Frank (VIS)	Torr, et al., 1995 Frank et al., 1995
MSX Midcourse Space Experiment	UVISI 5 spectrographic, 4 UV & visible				903.5 km 903.5 km	103 min	Apr. 24, 1996	U.S.A (APL)	
INTERBALL Auroral probe	UVAI (125-160) 2 cameras UFSIPS (oxygen 130.4, 135.6, 149.3)	2D-CCD 6.7 sec integration	1/spin (120 sec)	578 x 385 20Þx 25Þ, 11 km	3.0 Re 1,040 km	5h 47m	Aug. 29, 1996	Russia L.Cogger (UVAI) A. Kuzmin (UFSIPS)	http://www.iki.rssi.ru/interball.html
IMAGE	FUV WIC: 140-180 SI: 121.8 (Ly-a) & 135.6	CCD	2 min (spin period)	WIC: 17Þx17Þ, 256x256 SI: 15Þx15Þ, 128x128	7 Re 1,000 km	14.2 h I = 90Þ	Mar. 25, 2000	U.S.A. S.B. Mende (UC, Berkeley)	Burch, Space Sci. Rev., 91, 1, 2000

time resolution was very high (maximum 8 sec). The FREJA was a lower altitude satellite with a small inclination to study the details of the auroral activity instead of the global feature [*Murphree, et al.*, 1994]. The POLAR was one of the key satellites in the International Solar Terrestrial Physics (ISTP) program (others are the GEOTAIL, WIND, SOHO, and CLUSTER) [*Acuna, et al.*, 1995], and will be the most successful satellite for the global auroral substorm study. Because of its highly eccentric polar orbit with the perigee and apogee of 185 km and 8.0 Re, respectively, it can monitor the global auroral activity for up to 6 to 8 hours with the maximum time resolution of 18 seconds. [*Torr, et al.*, 1995]. Considering the typical substorm duration of 1 to 3 hours, it can follow a full evolution of one or more auroral substorms continuously. The IMAGE was the first satellite that can observe the global proton auroral activity simultaneously with the electron aurora with the two different kinds of UV imagers.

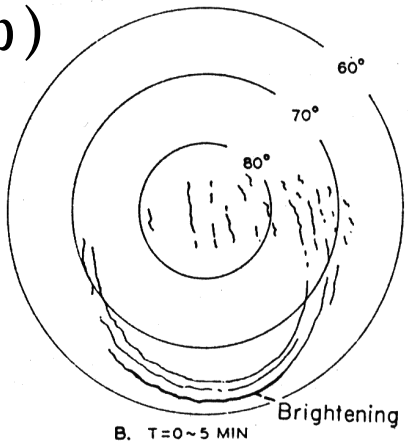
(b) Classical morphology

By the analysis of a large number of the all-sky camera observations during the International Geophysical Year (IGY) period (1957/58), it was found that the auroral activity appears in a specific circumpolar latitudinal zone (auroral zone) with an oval shape whose center is a little shifted to nighttime region from the geomagnetic pole [*Feldstein*, 1963], and the auroral oval changes its shape depending on the geomagnetic activity level [*Feldstein and Starkov*, 1967]. Using the similar data set, *Akasofu* [1964] constructed a phenomenological model for the auroral substorm evolution for the first time, which is shown in Figure 1.1. Many important concepts and terminology can be seen in this figure. During the quiet phase, there are several quiet and homogeneous arcs that are approximately parallel to the geomagnetic latitude circles there (Fig. 1.1(a)). The expansive phase starts with a sudden brightening of the southernmost arc a few thousand kilometers in length approximately centered at the midnight meridian (Fig. 1.1(b)), which is followed by a "bulge" around the midnight sector rapidly expanding poleward, westward and eastward (Fig. 1.1(c)). Within the bulge, "break-up" occurs such that a quiet curtain-like form appears to be disrupted and scattered all over the sky. The front of the bulge around the midnight meridian reaches its northernmost point (Fig. 1.1(d)) and within the bulge there are active bands. At the western edge of the expanding bulge, the folds of the arcs called the westward travelling surge (WTS) are formed and move rapidly westward because of the expansion of the bulge. The surges move along the preexisting bright bands (arcs) which are greatly displaced poleward after the passage of the surges. To the south of the bulge, isolated and cloud-like patches appear and drift westward and eastward in the evening and morning sectors, respectively. During the recovery phase, the northernmost active band starts to return southward (Fig. 1.1(e)). The size of the bulge is reduced, the speed of the WTS is also reduced, and the WTSs degenerate into a group of loops. During the later stage of the recovery phase, there appear quiet arcs slowly moving equatorward (Fig. 1.1(f)).

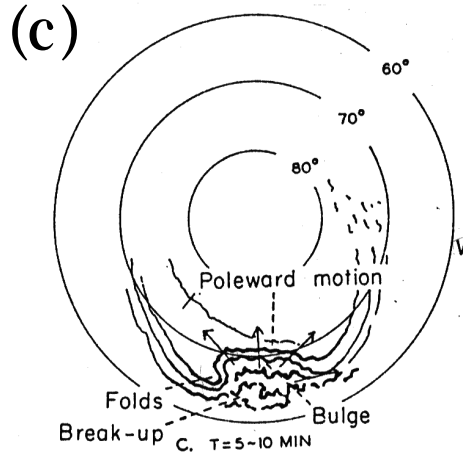
This initial morphology of the isolated auroral substorm has been modified and refined along new observational results obtained with various ground-based and satellite-borne optical instruments and other instruments, so far.



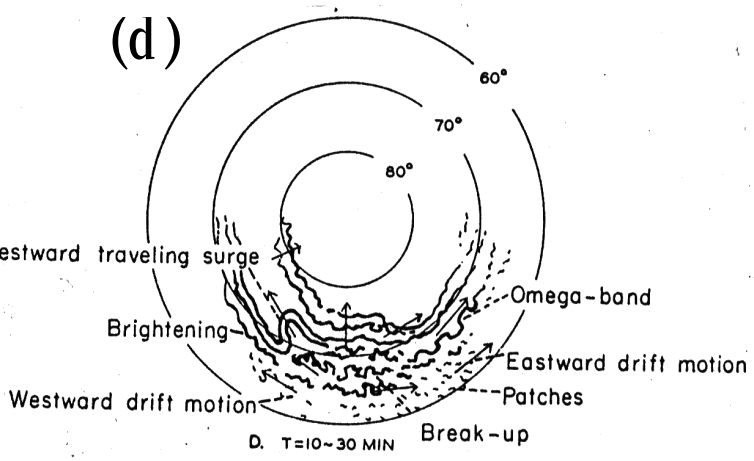
THE DISTRIBUTION OF THE AURORAS DURING THE QUIET PHASE.



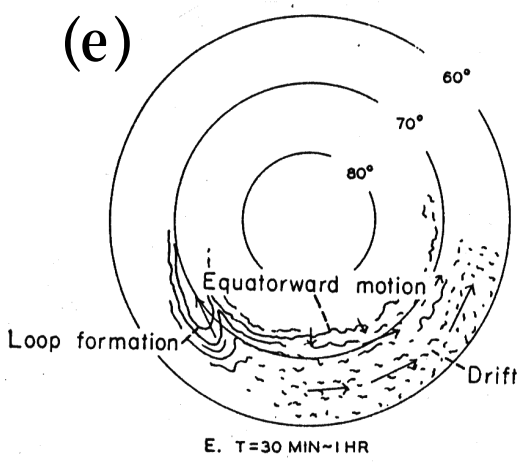
AURORAS DURING THE EXPANSIVE PHASE (STAGE I).



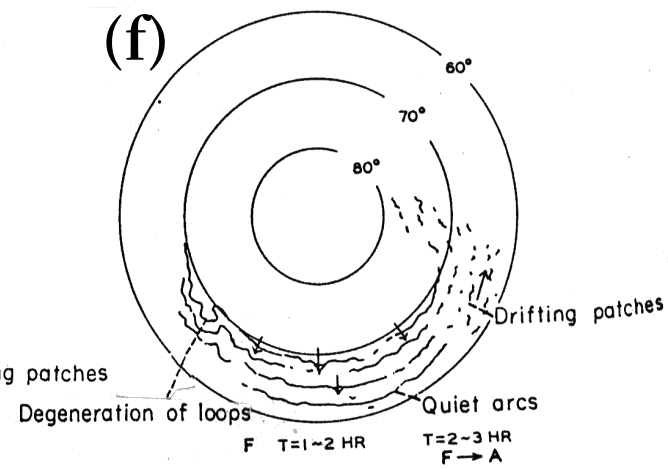
AURORAS DURING THE EXPANSIVE PHASE (STAGE II).



AURORAS DURING THE EXPANSIVE PHASE (STAGE III).



AURORAS DURING THE RECOVERY PHASE (STAGE I).



AURORAS DURING THE RECOVERY PHASE (STAGE II).

(c) Diffuse aurora, onset arc, and onset location

The ISIS-2 scanning photometer first revealed the existence of the uniform diffuse auroral belt which is located at lower latitudes of the discrete auroral arcs and bands [*Lui and Anger*, 1973a]. Relatively uniform precipitation of the low energy electrons is responsible to the diffuse auroras, and the discrete auroras are caused by the highly structured and intense electron precipitation [*Lui et al.,* 1977]. These two characteristics of the precipitating electrons are categorized by *Winningham et al.* [1975] as the CPS (central plasma sheet) and BPS (boundary plasma sheet), respectively. Protons precipitate in both the discrete and the diffuse auroral region but do not contribute significantly to the total energy flux, except near the equatorward boundary of the diffuse aurora [*Lui et al.*, 1977]. Such kind of diffuse auroras are identified also from ground-based observations, and are most clearly seen equatorward of the WTSs [*Lui et al.,* 1973b]. The bright arcs along which the WTS travels are located around the boundary between the CPS and the BPS regions near the poleward termination of the durably trapped energetic electrons [*Lui et al.*, 1978] which are categorized as VA (Van Allen belt) by *Winningham et al.* [1975], and the poleward termination location is called the trapping boundary [cf. *Frank and Ackerson,* 1972]. *Lui et al.* [1978] presumed that the location is around where the magnetic field lines begin to be distorted from the dipolelike to the taillike configuration. *Persson et al.* [1994a; 1994b] reported that southward drifting arcs are observed poleward of the initial expanding bulge, and the initial brightening arc (hereinafter we call this arc as "onset arc") should be located near the trapping boundary.

Since the proton auroral Balmer-line emissions (486.1 nm (Hβ) and 656.3 nm (Hα)) are emitted from excited neutral hydrogen atoms which are originated from the precipitated protons through charge exchange processes, the ionospheric emission region spread and becomes diffusive due to the free motion of the hydrogen atoms from the earth magnetic field [*Davidson*, 1965; *Lorentzen*, 2000]. The proton precipitation also produces the other emissions which appear in the case of the electron precipitation [*Eather*, 1968]. Hence the electron diffuse auroral region usually co-exists with the proton diffuse auroral region optically, though the emission intensity ratios between the two kind of emissions are theoretically estimated and should be within some ranges if the excitation is purely due to the proton precipitation. If the observed ratios are beyond the estimated range, then the contribution of the electron precipitation should be important. Distribution of the diffuse proton and electron precipitation regions depends on the local time and the magnetic activity level or the substorm phase. In the quiet phase, in the evening to midnight region, the proton auroral region is located at lower latitudes of the electron discrete aurora and the electron diffuse precipitations are absent in the proton auroral region, whereas after midnight, the proton and electron precipitation regions overlap together [e.g. *Ono et al.*, 1987; *Fukunishi*, 1975b; *Creutzberg et al.*, 1988]. Such a quiet time distribution could be partly understood by the magnetic drift motion of the energetic protons and electrons in the near earth magnetosphere.

Samson et al. [1992a] claimed that the onset arc is formed within a region of intense proton precipitation well equatorward of the nightside open-closed field line boundary. *Samson et al.* [1992b] also showed a case in which the onset arc is located equatorward of the region of the field line resonance. The field line resonance was observed before the onset by the CANOPUS

magnetometer array [*Rostoker*, 1995] and the Goose Bay HF radar [*Greenwald et al.*, 1995]. This result implies that the onset location should be situated in the closed and more dipolar field line region in the magnetosphere. *Deehr* [1994] showed that the onset arc is located 10 to 300 km poleward of the proton precipitation in the 33 onset cases, and the proton emission is in the diffuse auroral region which have a significant electron-induced component (electron characteristic energy > 20 keV). *Montbriand* [1971] and *Fukunishi* [1975b] also depicted the onset arc around the poleward boundary or just poleward side of the proton auroral region in their figures for the proton and electron auroral substorms. *Vallance Jones et al.* [1982] showed that for the premidnight substorms, the onset arc is located around the poleward edge of the proton auroral region where the gradient of the proton emission intensity is rather small. *Oguti* [1973] showed that on many occasions, a discrete arc splits from the higher latitude arc, moves toward the lower latitudes, and eventually makes contact with the poleward boundary of the proton auroral region, and the breakup occurs at or nearest point of the contact. Since the electrons responsible to the higher latitude arc should come from the plasma sheet, the onset location should be well equatorward of the open-close boundary. *Oguti* [1973] presumed that the onset of the auroral breakup should be related to the close approach, or contact, of the inner boundary of the plasma sheet to the proton trapping region near the Earth. From these observations, it seems to be confirmed now that the onset arc is located poleward of the proton diffuse emission region, around the boundary of the CPS and BPS type electron precipitation regions, around the trapping boundary of the energetic electrons, and well equatorward from the open-close boundary, hence, the onset region should be very close to the Earth in the closed field line region in the nightside plasma sheet, although the source mechanism for the onset arc itself is still debatable [e.g. *Lyons and Samson*, 1992; *Lui and Murphree*, 1998] and is a very important issue to be clarified.

Elphinstone et al. [1995a] summarized that one typical type of the auroral brightening occurs on the middle of the main UV oval poleward of the energetic proton isotropic boundary (IB) and very close to the energetic (> 100 keV) electron IB. The IB is defined as the lowest latitude of equal flux of the precipitating and trapped particles observed by the low altitude satellites, e.g. NOAA [e.g. *Sergeev et al.*, 1993a; *Sergeev and Gvozdevsky*, 1995]. Below the IB latitude, the trapped particle flux becomes dominant. For this definition, the IB is usually located at lower latitudes of the trapping boundary. The IB depends on the particle energy and is considered to signify the transition of the particle motion in the magnetosphere from the non-adiabatic to adiabatic in the tailward and earthward sides, respectively, owing to the relationship between the magnetic field line curvature and the particle gyroradius. *Yahnin et al.* [1997] showed that the discrete auroral arcs are always situated polewards from (or very close to) the IB of > 30 keV electrons, whereas the IB of the > 30 keV protons is often seen inside the diffuse aurora, which means that the discrete arcs are located in the isotropic region both for the energetic electrons and protons. *Newell et al.* [1996] categorized the nightside precipitation characteristics observed by the DMSP satellites. *Kauristie et al.* [1999] compared the some boundaries defined by *Newell et al.* [1996] with the VIKING UV oval location, and found that the maximum UV intensity is usually located poleward of the transition boundary from unstructured to structured precipitation which should be very close to the

CPS-BPS boundary in *Winningham et al.* [1975]. This result, considering together with the results in *Elphinstone et al.* [1995a], indicates that the auroral brightening occurs in the structured precipitation region. *Elphinstone et al.* [1991] mapped the UV oval into the magnetosphere using the *Tsyganenko* [1987] long magnetospheric model, and found that the mapped location is very close to the region of the model cross-tail current density maximum in the near earth region, and the oval poleward boundary is located equatorward of the model open-closed field line boundary. *Pulkkinen et al.* [1991; 1992; 1995; 1998] mapped the auroral brightening region into the nightside tail using the time-evolving magnetic field model based on the *Tsyganenko* [1989a] model (modified T89 model), and found that the mapped locations are around 5 to 15 Re from the earth where the tail field configuration is very stretched and the equatorial magnetic field intensity becomes very small, and even for the 1 keV electrons, their motions could become chaotic, judging from the specific parameter $\kappa = (R_{min}/\rho_{max})^{1/2} = (B_n/B_0)(L/\rho_0)^{1/2}$, where R_{min} and ρ_{max} are minimum curvature radius and maximum Larmor radius at the equatorial plane, respectively, and B_n, B_0, L, ρ_0 are the magnetic field component perpendicular to the neutral sheet, the lobe field strength, the thickness of the field-reversal region, and the Larmor radius of particle, respectively [*Büchner and Zelenyi*, 1987; 1989]. All these results also support the idea of the near earth substorm onset possibly around 10 Re from the earth.

(d) Premidnight preference of the onset region

By examining a large number of photographs taken from the DMSP satellite, *Akasofu* [1974] found that the initial brightening tends to occur more frequently in the late evening sector than in the midnight sector, while the well expanding bulge is approximately centered around the midnight meridian. This kind of premidnight preference of the onset region has been confirmed in other studies. From an analysis of 67 substorms observed by the DE-1 satellite, onsets are found to occur predominantly within the limited magnetic local time (MLT) interval 2200-2400 hours [*Frank and Craven*, 1988]. The average location of the peak intensity for 80 onsets observed by the VIKING satellite was 22.9±1.2 hour MLT and 65.9±3.5 deg CGMLAT (Corrected GeoMagnetic LATitude) and the spatial extent was about 1 hour local time (about 680 km at 65.9 deg CGMLAT) [*Elphinstone et al.*, 1995a] or even more localized being less than 500 km [*Murphree et al.*, 1991]. At geosynchronous orbit, the occurrence frequency of the energetic electron flux recovery at the onset time has also a maximum in the premidnight region [*Nagai*, 1982a], the plasma injection initially occurs in a relatively narrow sector skewed toward the evening side of midnight and expands both eastward and westward later [*Arnoldy and Moore*, 1983], and the magnetic field dipolarization also occurs in a narrow local time sector centered at 2330 MLT and expands longitudinally afterward [*Nagai*, 1991]. Event frequency of the dispersionless ion injection observed by the AMPTE/CCE satellite showed a larger occurrence in the premidnight sector [*Lopez et al.*, 1990]. The GEOTAIL observations showed that the magnetic reconnection in the near earth tail also takes place around the premidnight region of (X, Y) ~ (-19, 6) Re [*Miyashita et al.*, 1999; 2000] or (-20, 5) Re [*Machida et al.*, 1999] in the GSM coordinates a few minutes before the onset, and the center of the plasmoids and the near earth neutral line (NENL) is likely to be located at Y ~

3 Re in GSM in the near earth region [*Ieda et al.*, 1998]. Hence the premidnight preference of the onset region is an essential feature of substorm to be explained.

(e) Double oval configuration during the recovery phase

Akasofu [1974] also found that during the expansive phase, there appears often a dark space between the poleward advancing discrete arc and the poleward boundary of the diffuse aurora in lower latitudes. In most cases, the dark space is caused by the lack of auroral electrons of small pitch-angles, while the flux of large pitch-angle electrons is high. This kind of auroral configuration is later called the "double oval" distribution [*Elphinstone and Hearn*, 1992; *Elphinstone et al.*, 1995b; *Elphinstone et al.*, 1995c] from the VIKING data, which is clearly seen during the recovery phase. The "double oval" distribution is also recognized in the DE-1 data [e.g. *Craven and Frank*, 1987], in the AKEBONO data [*Yamamoto et al.*, 1993], and in the POLAR data [e.g. in Fig. 1 of *Cummer et al.* [2000]], hence, is considered to be a very essential auroral feature during the recovery phase. Many studies showed that the poleward most discrete arc system of the "double oval" distribution is located just equatorward of the velocity-dispersed ion structures (VDIS) in the low altitude plasma sheet boundary layer (PSBL) [e.g. *Fukunishi et al.*, 1993; *Yamamoto et al.*, 1993; *Elphinstone et al.*, 1995b], and it is located in the upward field-aligned current (FAC) region adjacent to the downward FAC region in the VDIS region on higher latitude side [*Fukunishi et al.*, 1993; *Yamamoto et al.*, 1993; *Burke et al.*, 1994]. Since the VDIS ion beams are considered to be supplied from a reconnection process at an X-type neutral line in the tail [e.g. *Saito et al.*, 1992], it is inferred that the intense most poleward arc should be closely related with the reconnection process in the tail which should become active during the recovery phase. It should be also noted that in almost all cases, the most poleward discrete arc system connects with the WTS at its western end, which suggests some relationship between the WTS and the reconnection process in the tail. One important characteristic of the VDIS region to be noted is a soft electron precipitation and resultant diffuse auroral emissions in the red line (630.0 nm) [e.g. *Feldstein and Galperin*, 1993]. The poleward boundary of the diffuse red line emissions observed with the ground-based meridian scanning photometer is considered to be very close to the polar cap boundary [*Blanchard et al.*, 1995; *Blanchard et al.*, 1997], and a good "tracer" for the mapping into the magnetospheric PSBL region [*Galperin and Feldstein*, 1996]. As for the lower latitude diffuse auroral region in the "double oval" configuration, complex pulsating auroral activities can usually be seen within it especially in the dawn sectors [*Oguti*, 1981]. Since the pulsating auroras are considered to be created by the precipitation of energetic electrons injected in the near earth region and drifting eastward, the diffuse auroral region should be related with the precipitation of the injected electrons during the expansion and recovery phases. In some cases in the late expansion phase, a horn-like structure develops in the dusk-side from the WTS at the poleward side of the pre-existing oval forming the other type of the "double oval" configuration in the dusk-side [*Henderson et al.*, 1994; Plate 2(d) in *Fujii et al.*, 1994]. Its shape, orientation and the timing of the appearance suggest that it is caused by an activation of the low latitude boundary layer (LLBL) after a contact with the disturbance in the central plasma sheet [*Henderson et al.*, 1994] and the WTS should be located

very close to the contact point in such cases. There are some cases such that the "double oval" configuration already exists when the onset starts within the lower latitude diffuse UV oval because the recovery phase of the preceding substorm has not yet completed or a substantial activity still continues. In such cases, it takes about 3 to 10 minutes for the bulge to expand from the onset region to the poleward most arc location in the "double oval" configuration [*Elphinstone et al.*, 1995a]. This kind of situation suggests that the near earth onset region should be separated from the tailward reconnection region.

(f) North-south structured aurora

Akasofu [1974] also noted that an intense WTS leaves behind considerable disturbance in the diffuse aurora. The diffuse aurora often develops a torch-like structure that appears as a series of arcs being aligned in the north-south direction. Such a north-south (N-S) structure was first stressed by *Montbriand* [1971], who considered that the N-S structure develops from the poleward most active band in the bulge, while *Akasofu* [1974] suggested that it is embedded in the diffuse auroral region. The N-S structured auroral form appears at the time of the initial expansion around a very close vicinity of the WTS [*Nakamura et al.*, 1993], at the time of the re-activation of the bulge expansion around the center of the bulge [*Rostoker et al.*, 1987a; *Henderson et al.*, 1994; *Henderson et al.*, 1998], and after a full expansion of the bulge in the late expansion and recovery phases also around the center of the bulge [*Sergeev et al.*, 1999; *Sergeev et al.*, 2000]. In the former two cases, its appearance is clearly related with the activation of the most poleward discrete arc system in the bulge, and in the last case, it is called the "auroral streamer" extending from the poleward boundary to the equatorward one in the bulge [*Elphinstone et al.*, 1996]. In the first case, the N-S auroras quickly develop into the pulsating and/or diffuse auroras [*Nakamura et al.*, 1993]. This kind of transition processes were analyzed in detail by *Oguti* [1981] using the high-sensitivity auroral TV observation. The geosynchronous injection features and bursty bulk flow (BBF) features [*Angelopoulos et al.*, 1994] in the near tail are also observed in association with the N-S auroral formation [e.g. *Henderson et al.*, 1998; *Sergeev et al.*, 1999; *Sergeev et al.*, 2000]. Hence it can be inferred that the N-S structured aurora should be closely related with the temporal, bursty plasma transport from a reconnection region in the near tail into the near earth region, and should indicate an interrelationship between the most poleward arc system and the lower latitude diffuse auroral region in the "double oval" configuration, although the formation mechanism is still debatable [e.g. *Nakamura et al.*, 1993; *Liu and Rostoker*, 1993; *Chen and Wolf*, 1993; *Sergeev et al.*, 2000].

(g) Growth phase

(g.1) Auroral features

In the initial morphology of *Akasofu* [1964] shown in Fig. 1.1, there were three distinct phases, i.e., the quiet phase, expansive phase, and the recovery phase. From very beginning of the substorm study, it was noticed that there is other distinct stage before the expansive phase onset, which is manifested in the polar magnetic disturbances [e.g. *Iijima and Nagata*, 1968], in the geosynchronous magnetic field configuration change [e.g., *Cummings et al.*, 1968] and the

energetic electron flux change [*Lezniak and Winckler*, 1970]. This stage, called the growth phase [*McPherron*, 1970], seems now widely accepted to be certainly existed though this concept has been very controversial [ref. *Akasofu and Snyder*, 1972; *Kamide and Matsushita*, 1978; *McPherron*, 1994]. The growth phase features are most clearly seen in the substorm which starts after a southward turning of the interplanetary magnetic field (IMF) preceded by a prolonged northward IMF period [e.g. *Lyons et al.* 1997; *Shue et al.*, 2000], and its duration is about 64 to 72 min as a median value in such cases [*Lyons et al.*, 1997]. During the growth phase, both the electron and proton auroral ovals move equatorward, hence, the auroral void region in the poleward side (often assumed as the polar cap region) is expanding, which is observed both from the ground [e.g. *Feldstein*, 1970; *Vorobjev et al.*, 1975; *Fukunishi*, 1975b; *Horwitz and Akasofu*, 1977; *Deehr*, 1994; *Samson et al.*, 1992a; *Lyons et al.*, 1997] and from the satellites [e.g. *Murphree et al.*, 1991; *Brittnacher et al.*, 1999]. The equatorward expanding motion is gradual and not concurrent at all local times with a highest speed around the nightside onset meridian (e.g. 100-250 m/s in *Fukunishi* [1975b]; 140-330 m/s in *Murphree et al.* [1991]; 80-140 m/s in *Brittnacher et al.*[1999]) and much slower one around the noon [*Brittnacher et al.*, 1999]. Very often, the poleward boundary of the auroral oval moves equatorward more rapidly than the equatorward boundary, resulting in the decrease of the oval latitudinal width [*Brittnacher et al.*, 1999]. As the oval moves equatorward and becomes thin, the oval emission intensity gradually increases toward the onset, which can be seen at all local times except around the noon [*Brittnacher et al.*, 1999; *Deehr*, 1994]. The "onset arc" is usually embedded in the rapidly moving poleward part of the oval and approaches to the slower equatorward part, which is more clearly seen in the ground-based high spatial resolution observations [e.g. *Oguti*, 1973; Fig. 3 in *Lyons et al.* [1997]].

It is known that about 10 min after a southward turning of the IMF preceded by a prolonged northward IMF period, sun-aligned transpolar arcs (so-called the theta auroras [*Frank et al.*, 1982]) are often formed in the "polar cap" area and start to drift duskward (dawnward) across the "polar cap" under the positive (negative) Y-component of the IMF (IMF-By) condition in the northern hemisphere and conversely in the southern hemisphere [*Chang et al.*, 1998], then they are gradually swept away from the "polar cap" area. This sweeping process proceeds concurrently with the above mentioned equatorward motion, thinning, and intensification of the auroral oval during the growth phase, which is typically seen in the Plate 2 of *Brittnacher et al.* [1999] for the case of the magnetic cloud event on January 10, 1997. It should be noted that the auroral activity in that event is initiated by an arrival of an interplanetary shock wave and associated magnetospheric compression during the northward IMF period. Development of the auroral activity associated with the shock is shown in detail in *Spann et al.* [1998] and *Zhou and Tsurutani* [1999].

In the dayside, there exist characteristic regions of the cusp auroral emissions on the higher latitude side of the emission region of the plasma sheet origin . The former region is dominated by the red line (630.0 nm) emission and the latter is by the diffuse green line (557.7 nm) emission [*Sandholt et al.*, 1998a; 1998b]. When the IMF turns from northward to southward, the cusp auroral region shifts its location discretely from the poleward side to more equatorward side by about 100 to 200 km in 5 to 10 min, and both the cusp and plasma sheet auroral regions gradually migrate

equatorward during the growth phase. The equatorward migration of the cusp auroral region proceeds in stepwise such that individual auroral intensification occurs intermittently a little equatorward of the preexisting region, and the individual form moves poleward while the entire region moves equatorward [*Sandholt et al.*, 1998c; *Vorobjev et al.*, 1975]. The interval of each individual intensification is about 4 to 5 min [*Sandholt et al.*, 1998c]. It was suggested that the latitudinal shift of the cusp auroral region corresponding to the IMF polarity change should be associated with the change of the reconnection site from the higher latitude tail magnetopause during the northward IMF period to the lower latitude dayside magnetopause during the southward IMF period [cf. *Maezawa*, 1976]. The intermittent intensification and subsequent poleward motion of the individual auroral form and gradual equatorward migration of the entire region indicates that the lower latitude reconnection process should occur intermittently, tailward transportation of the reconnected flux occurs intermittently, and the dayside magnetosphere erosion proceeds intermittently [*Sandholt et al.*, 1998a; 1998b; 1998c].

(g.2) Dayside process

When the IMF turns southward, the IMF and the earth magnetic field can reconnect more efficiently at the dayside magnetopause, the coupling efficiency between the solar wind and the magnetosphere increases, and larger energy can enter the magnetosphere. The dayside reconnection process has been studied theoretically [e.g., *Sonnerup*, 1979; *Hoshino and Nishida*, 1983] and actually observed by satellites [e.g., *Nakamura et al.*, 1998]. The dawn-to-dusk electric field induced by the reconnection process should produce the accelerated plasma flows out from the reconnection region in the interface layer around the magnetopause. This plasma motion with the reconnected magnetic field lines could work as a dynamo to drive the field-aligned currents flowing into the ionosphere at the morning side and away from the ionosphere at the afternoon side around the dayside cusp region. The current closure of this field-aligned current system on the ionospheric side is the Pedersen current. Because of the resistive nature of the ionosphere, the power of the dynamo is dissipated, and the flowing motion of the plasma should be decelerated. Hence the closure of the current system on the dynamo side was considered to be an inertia current [*Maezawa*, 1974; *Nishida*, 1978]. On the other hand, the global MHD simulation by *Tanaka* [1995] showed that the diamagnetic current due to the plasma pressure gradient directed toward the reconnection region should be a main contributor of the dynamo current. This closed current system reduces the magnetic flux in the dayside magnetosphere and enhances in the tailward side one, and as a result, the dayside magnetopause position tends to move closer to the earth, i.e. the erosion of the dayside magnetosphere occurs [e.g., *Maezawa*, 1974; *Holzer and Slavin*, 1978], and the tail lobe radius increases [*Maezawa*, 1975], and as a result, the tail flaring angle increases [*Fairfield*, 1985]. Such a change of the magnetopause shape responding to the variation of the IMF-Bz was shown in *Sibeck et al.* [1991]. Corresponding to the change in the dayside magnetosphere, the cusp latitude in the ionosphere shifts equatorward [e.g., *Burch*, 1973; *Kamide, et al.*, 1976] and the dayside auroral oval moves equatorward. These magnetic flux changes should be compensated if the sufficient magnetic flux transport is maintained inside the magnetosphere from the tail region toward the dayside

reconnection region. Hence, these observations suggest that such kind of a steady magnetic flux circulation should not be realized at least during the growth phase, and the flux transported from the dayside should be accumulated in the tail region.

The resistive ionosphere responds to the field-aligned current requirement from the dynamo region, and creates a suitable electric potential distribution to make the Pedersen current closure. Though the field-aligned current location is very localized around the cusp region, the created potential distribution spreads globally in a very short time, which is evidenced by a good correlation (correlation coefficient of 0.9) and simultaneity (time difference of less than 25 s) of the DP 2 type magnetic fluctuations [*Nishida*, 1968] between the auroral and equatorial latitudes [*Kikuchi*, et al., 1996]. *Kikuchi et al.* [1996] proposed an electric field transmission process between the ionosphere and the earth surface with a speed of light, accompanying a transverse electromagnetic (TEM) mode. This kind of quick global change of the ionospheric electric potential pattern responding to the IMF change was also evidenced by the SuperDARN HF radar observations [*Ruohoniemi and Greenwald*, 1998] and the analysis by the assimilative mapping of ionospheric electrodynamics (AMIE) technique [*Ridley, et al.*, 1998]. The SuperDARN radar observed the ExB drift motion of the field-aligned electron density irregularities in the F region [*Greenwald, et al.*, 1995] and the AMIE technique mainly depends on the ground magnetometer observations [*Richmond and Kamide*, 1988]. *Ruohoniemi and Greenwald* [1998] showed that the reconfiguration time of the global convection pattern from the positive IMF-Bz type to the negative type was only 2-4 min, and *Ridley, et al.* [1998] showed that the reconfigured pattern does not show any remarkable motion from the beginning until the time of the full development after about 12-13 min, on average, which implies a very rapid communication time in the ionosphere. Both authors suggest that the electric field projected around the cusp region should be propagated with the fast mode Alfvèn wave to the entire ionosphere in a few seconds. Although there are some different views about the ionospheric response [e.g., *Lockwood and Cowley*, 1999], and this is still a critical issue, these observations clearly show that the characteristic reconfiguration time in the ionosphere is much shorter than that expected by the ionospheric plasma convection speed from dayside to nightside (74 min for 20 deg latitudes with 0.5 km/s convection) or the solar wind speed from dayside to nightside (8 min for 30 Re with 400 km/s solar wind) with the magnetospheric convection from the tail magnetopause to the equatorial plane (16 min for 15 Re with 100 km/s).

The globally spread ionospheric electric potential should be transmitted from the ionosphere into the entire magnetosphere beyond the localized dayside dynamo region by the shear Alfvèn wave propagating along the field lines and the fast mode wave propagating across the field lines. The propagation time of the wave with the typical velocity of 1000 km/s [*Moore, et al.*, 1987] for the 20 Re distance, for example, is about 2 min, which is much faster than the convection time estimated above. Hence the ionospheric reconfiguration information should be transmitted into the nightside tail much faster than that brought by the solar wind tailward flow, hence, well before the reconnected magnetic flux is transported from the dayside and accumulated in the tail region. As a result, the magnetospheric reconfiguration should start a few minutes after the ionospheric reconfiguration. Such a quick correspondence between the start of the ionospheric potential change

and that of the magnetospheric reconfiguration was reported by many authors using the ground magnetometer data and the geosynchronous satellite data [e.g., *Kokubun and McPherron*, 1981; *Nagai*, 1982b; *McPherron and Manka*, 1985].

(g.3) Nightside process

It is generally believed that the increase of the tail flaring angle causes a larger normal dynamic pressure exerted by the solar wind on the tail lobe and a resulting increase of the lobe magnetic field pressure for balancing it. The increase of the lobe magnetic pressure causes a compression and thinning of the plasma sheet which results in the increase of the plasma pressure within it and the magnetic field configuration change toward more taillike. This more stressed configuration is sustained by a re-arranged lobe magnetopause - cross-tail current closure circuit system such that the intensity of the current increases and the earthward edge of the current system moves more earthward [e.g. *Coroniti and Kennel*, 1972; *Lui*, 1991b; *Baker et al.*, 1996; *McPherron*, 1979; 1994].

(g.3.1) Magnetospheric pressure change

The increase of the magnetospheric total pressure is actually observed during the growth phase [e.g. *Nishida and Nagayama*, 1973a; *Caan et al.*, 1975; *Fairfield et al.*, 1981; *Nagai et al.*, 1997; *Petrukovich et al.*, 1999]. *Nishida and Nagayama* [1973a] suggested that the increase of the lobe magnetic field propagates tailward in the range of 6.6 to 80 Re. *Caan et al.* [1975] analyzed 20 tail lobe events at 9-19 Re observed by the OGO 5 satellite, and showed that average increase of the lobe magnetic pressure is about 27 % of the initial quiet value. They estimated that the total energy stored during the growth phase is about 10^{15} J, which is comparable to the total energy released and dissipated in a typical substorm [cf. *Baumjohann*, 1986]. *Nagai et al.* [1997] analyzed 19 plasma sheet events observed by the GEOTAIL at 20-50 Re, and found that the increase of the plasma pressure is caused mainly by the increase of the plasma number density, while the plasma temperature does not show any significant changes. They also mentioned that the pressure increase starts immediately after the southward turning of the IMF. *Petrukovich et al.* [1999] compared the pressure variations observed in the plasma sheet by the GEOTAIL and in the tail lobe by the INTERBLL-TAIL satellite, and showed that these two values are nearly identical during almost whole period of the growth phase. In their one event, in the first part of the growth phase (about 45 min duration), the increase of the plasma pressure is caused mainly by the increase of the plasma number density as indicated by *Nagai et al.* [1997], while, in the second part (about 27 min duration), that is clearly caused mainly by the increase of the plasma temperature, which occurred together with the increase of the earthward plasma flow. Such a transition of the pressure increase can also be seen in the event of *Mukai et al.* [1998] where the plasma temperature increases from about 5 min before the onset together with the tailward flow observed at 17 Re down the tail. These results imply that there exist at least two kinds of processes to increase the plasma sheet plasma pressure. One is a more general one which operates during the whole growth phase and causes mainly the increase of the plasma number density, and the other is less common which should be

associated with the enhanced plasma flow and causes mainly the increase of the plasma temperature. *Fairfield et al.* [1981] also compared the lobe and plasma sheet observations by the IMP satellites and showed a good correspondence between them at near geosynchronous to 40 Re. All these observations showed that the enhanced magnetospheric pressure starts to decrease everywhere in the near and mid-tail regions, immediately after the expansion phase onset.

(g.3.2) Plasma sheet thinning

The thinning of the plasma sheet and the taillike reconfiguration of the field lines during the growth phase are also reported by many investigators [e.g. *McPherron, et al.*, 1973; *Nishida and Fujii*, 1976; *Fairfield et al.*, 1981; *Kokubun and McPherron*, 1981; *McPherron et al.*, 1987; *Sergeev et al.*, 1993b; *Sanny et al.*, 1994; *Pulkkinen et al.*, 1994a]. The timing of the remarkable thinning depends on the distance from the earth and the local time separation from the initial onset meridian. In near earth region (< ~15 Re), the thinning often starts before the onset during the growth phase and subsequent plasma sheet recovery or the thickening occurs immediately after the onset, while in the further tailward region, the remarkable thinning tends to occur after the onset, and the subsequent thickening often occurs during the recovery phase of the substorm [e.g. *Nishida and Fujii*, 1976; *Fairfield et al.*, 1981; *Pytte et al.*, 1976a; 1978]. In the further mid-tail region, the thinning occurs after the onset while the magnetospheric total pressure is decreasing. This indicates that the main cause of that thinning is not the compression due to the enhanced tail lobe magnetic pressure. The formation of the near earth neutral line is a most possible explanation for that thinning [e.g. *Nishida and Nagayama*, 1973b].

As for the thinning in the nearer earth region during the growth phase, the thickness of the current sheet was estimated by the two-point observations by the ISEE 1 and 2 satellites [*McPherron et al.*, 1987; *Sanny et al.*, 1994; *Sergeev et al.*, 1993b]. *Sanny et al.* [1994] showed that the thickness in the 13 Re tailward region decreased exponentially from about 5 Re to 1 Re with a time constant of about 14 min before the onset, and the further thinning was observed immediately after the onset with the estimated current sheet thickness of about 500 km. *Pulkkinen et al.* [1994a] applied their modified T89 magnetic field model to the same event of *McPherron et al.* [1987] and *Sanny et al.* [1994], and could reproduce the thinning of the current sheet at the ISEE locations. They suggested that the chaotization of thermal electrons due to the small value of κ parameter as mentioned in the section 1.1.1(c) can be expected around the ISEE locations at the end of the growth phase. Electron energy spectra observed by the ISEE 1 showed that the electron pressure distribution at the beginning of the growth phase was anisotropic such that the parallel pressure was greater than the perpendicular one. The pressure distribution became isotropic just before and after the onset, which supports their expectation. *Sergeev et al.* [1993b] analyzed another event observed at about 11 Re from the earth, and showed that the current sheet thickness decreased from about 3000 km to about 800 km during the growth phase, which was comparable to the gyroradius of thermal protons in the field just outside of the current sheet. They found a very small (~1 nT) normal component and a large shear east-west component of the magnetic field in the current sheet. About 1 min before the local onset, they found another type of neutral sheet crossing ("turbulent

crossing"), which was characterized by a further small current sheet thickness less than 600 km and a strong earthward and duskward plasma flow with a positive normal magnetic field component. One minute later, the current disruption [cf. *Lui*, 1996] and dipolarization features appeared. *Mitchell et al.* [1990] analyzed the same event in detail focusing on the characteristics of the electrons and ions during the growth phase. They indicated that during the early growth phase, the excess parallel pressure of the adiabatic electrons near 1 keV can provide a significant fraction of the cross-tail current in a thin current sheet within a thicker plasma sheet, while late in the growth phase, the Speiser type motion [*Speiser*, 1965] of the non-adiabatic ions in the thin current and thin plasma sheet becomes the major current carrier. This transition should occur very rapidly within the one minute. They estimated the current sheet thickness at the late stage as about 400 to 800 km. It should be noted that they did not observe the initial onset but the local onset as *Sergeev et al.* [1993b] showed, which occurred with the eastward expansion of the substorm current wedge [cf. *McPherron et al.*, 1973] or the current disruption region. However their event seems to contain very essential information about the cause of the current disruption, i.e. initial onset processes. *Lu et al.* [1999] took a similar approach as *Pulkkinen et al.* [1994a] using the modified *Tsyganenko* [1996] (T96) model. They analyzed other three events observed by the GOES and GEOTAIL satellites and showed that the current sheet half thickness around 7.5 Re from the earth at the end of the growth phase should be about 190-1340 km.

(g.3.3) Taillike reconfiguration

The taillike reconfiguration of the magnetic field lines in the plasma sheet is another aspect of the thinning of the current/plasma sheet. In the near earth region, the normal component gradually decreases and the radial component increases, which results in a stretched configuration [e.g. *Kokubun and McPherron*, 1981; *Nagai*, 1982b]. As depicted by *Kistler et al.* [1993], the field line rooted at a fixed point on the earth has its foot in the neutral plane at further tailward point in such a stretched configuration. Conversely, the field line rooted at a fixed point in the neutral plane has its foot on the earth surface at lower latitude. *Sergeev et al.* [1990] estimated that about 80 % of the equatorward expansion of the auroral oval can be explained by such a magnetic field configuration change and remaining 20 % is due to the earthward motion of the source region.

Such a tailward movement of the field lines should be caused by the dusk-to-dawn directed electric field E_i which is induced as the plasma pressure increases and the magnetic field intensity decreases in the plasma sheet, the magnetic field intensity increases in the tail lobe, and the dawn-to-dusk cross tail current density increases during the growth phase. Hence, the E_i should be produced internally and closely related with the diamagnetic characteristics of the plasma sheet plasma. From the frozen-in relation, $E_i + VxB = 0$, plasma should move tailward with the field lines, because both E_i and VxB should become zero in the frame moving with the plasma. On the other hand, the enhanced dawn-to-dusk convection electric field E_c is projected from outside during the growth phase, possibly transmitted by the Alfvèn wave from the ionosphere as already mentioned. Under such an external electric field, plasma should move earthward across the field lines to produce the inductive electric field E_i' to maintain the frozen-in relation, $E_c + E_i' = 0$, $E_i' = VxB$,

because E_c does not vanish in the frame moving with the plasma. Such a relationship between the motions of the plasma and field lines was discussed by several investigators [e.g. *Vasyliunas*,1972; *Cai et al.*, 1995; *Hesse et al.*, 1997]. Because E_i and E_c direct oppositely, total electric field, $E = E_i + E_c$, should be very small, and the observed flow velocity of the plasma $V = E/B$ becomes very small [e.g. *Nagai et al.*, 1997; *Paterson et al.*, 1998].

Under the dawn-to-dusk E_c, the plasma in the off-equatorial region should move equatorward across the field lines with the E_cxB drift, hence, the plasma concentration should occur in the plasma sheet simply due to the E_c. Actually the thin current sheet could be formed in several numerical modeling works where the dawn-to-dusk electric field is imposed at the simulation boundary [e.g. *Pritchett and Coroniti*, 1995; *Xiaogang Wang and Bhattacharjee*, 1995; *Birn et al.*, 1998a; *Rastatter et al.*, 1999]. Since the equatorward moving plasma is influenced by the dusk-to-dawn E_i at the same time, its equatorward motion is very slow [e.g. *Nagai et al.*, 1997]. Due to the E_i, the field lines in the plasma sheet should move away from the equatorial plane as shown in the kinetic simulation of *Pritchett and Coroniti* [1995]. Hence, the field lines tends to be gradually evacuated from the thinning region and accumulated on the tailward and higher latitude side of it as the growth phase progresses, while the plasma moves toward the thinning region and accumulates within it. As a result, the normal component of the magnetic field, Bz in the thinning region decreases significantly, and there appears a local minimum Bz region in the near earth thinning region. Such a very small Bz was observed by the satellites [e.g. *Sergeev et al.* ,1993b] and shown in the simulation works [e.g. *Pritchett and Coroniti*,1995; 1997; *Pritchett et al.*, 1997; *Wu et al.*, 1998].

Pritchett and Coroniti [1995] showed that a thin intense current sheet develops in a thicker plasma sheet, and two kinds of duskward cross-tail current regions appear. One is located in the thin current sheet and its main carrier is electrons, and the main source is the diamagnetic current due to a sharp equatorward pressure gradient and the dawnward drift motions due to an intense equatorward electric field. The other region is situated at higher latitudes around the boundary of the thicker plasma sheet and its main source is the diamagnetic current due to the equatorward pressure gradient of the ions and electrons. They mentioned that electrons with small Larmor radius follow the convective flow and are concentrated in the thin current sheet, whereas the ions with larger Larmor radius tend to spread out across the whole plasma sheet (due to their non-adiabatic motions, though they did not mention so clearly). Hence the structure of the whole plasma sheet is determined by the ion distribution, while that of the current sheet is by the electron distribution. A negative potential develops in the central part of the current sheet due to the electron concentration, which causes the intense equatorward electric field. Though they did not mention clearly, the polarization electric field should direct earthward at the neutral plane and the polarization current (Hall current) should flow tailward. They presumed that the accumulated negative charge or positive divergence of the polarization current could be the source of the upward field-aligned current relating with the equatorwardmost auroral arc associated with the expansion phase onset, i.e. the onset arc. Such a hybrid structure, where a thin current sheet is embedded within a thicker plasma sheet having two kind of cross-tail current source regions, can be seen in a global MHD

simulation results [e.g. in Plate 1 of *Pulkkinen and Wiltberger* [2000]]. In the two-dimensional simulation of *Pritchett and Coroniti* [1995], neither the magnetic flux transport in the azimuthal direction and toward dayside nor the interaction between the magnetosphere and ionosphere were fully considered. Hence their result should be considered as an exaggerated one concerning the nightside magnetospheric process. However, important point is that they elucidated essential and potential characteristics of the night side magnetospheric plasma responding to the imposed convection electric field during the growth phase.

(g.3.4) Local minimum Bz and Steady Magnetospheric Convection

Existence of the minimum Bz in the near earth region (around 10 to 15 Re) was predicted in the consideration about the steady magnetospheric convection (SMC). *Hau et al.* [1989] and *Erickson* [1992] showed in their two-dimensional ideal MHD model calculations that the minimum Bz is necessary to accomplish the steady adiabatic convection in the magnetosphere and to be free from the pressure balance inconsistency problem [*Erickson and Wolf*, 1980]. To maintain the adiabatic relation, pV^{γ} = const., the plasma pressure p does not necessarily increase so largely if the flux tube volume V = ? ds/B increases as approaching the earth. γ is the ratio of specific heats. *Sergeev et al.* [1994] also showed that the minimum Bz is expected for a SMC event using the modified T89 magnetic field model of *Pulkkinen et al.* [1992]. It was shown that during the SMC period, features in the near earth region are very similar to those during the strong growth phase of substorms, while features in the mid-tail region are similar to those during the recovery phase [e.g. *Sergeev et al.*, 1994; *Pulkkinen et al.*, 1994b; *Sergeev et al.*, 1996a]. As for the auroral activity, the double oval configuration developed well during the SMC period and the poleward auroral void area "polar cap" does not change its spatial extent over long period [e.g. *Yahnin et al.*, 1994]. Note that all the SMC events studied begin after a substorm and end with another substorm [*Sergeev et al.*, 1996a]. During the SMC period, the IMF is stable and continuously directs southward, enhanced convection continues without typical signature of the Harang discontinuity in the ionosphere, and no substorm signatures appear both in the ionosphere and magnetosphere. *Galperin et al.* [1992] and *Galperin and Bosqued* [1999] speculated that the upward field-aligned current responsible to the onset arc should be created within the minimum Bz region. Generation mechanism of the current is the divergence of the radial cross-tail current due to the azimuthal pressure gradient with the radial gradient of the magnetic field intensity in that region. Hence their mechanism is a pure MHD process comparing with the Hall effect in *Pritchett and Coroniti* [1995]. *Borovsky et al.* [1998] pointed out that the pressure balance inconsistency could be relaxed by the pressure and mass losses by the reconnection and the auroral particle precipitations which are not included in the discussions by *Hau et al.* [1989] or *Erickson* [1992].

(g.3.5) Cross-tail current increase

Although the total electric field should be very small, the magnetospheric plasma gradually moves earthward under the influence of the dawn-to-dusk E_c. Actually, the slow (about 3 km/sec) inward motion of the earthward edge of the plasma sheet was observed at geosynchronous orbit

during the growth phase by *Shelley et al.* [1971]. Associated electric field was about 0.36 mV/m. Together with the earthward motion of the plasma, the thinning region should moves earthward, and the intense cross-tail current region approaches earthward. *Kaufmann* [1987] showed that a large cross-tail currents (100 mA/m or more) around 7 to 10 Re are required to produce the observed magnetic field perturbations at geosynchronous orbit. Similar amount of the integrated currents are estimated to be necessary around 4 Re to fit the observed magnetic perturbations in a strong growth phase with modified T89 model [e.g. *Pulkkinen et al.*, 1992]. In their all cases of *Pulkkinen et al.* [1992; 1994a], the location of the total integrated current maximum is inside of 10 Re (around 4 to 8 Re), and the current intensity sharply (gradually) decreases at earthward (tailward) side. This is also the case for the three events of *Lu et al.* [1999], and the maximum current density in their cases is 9-23 nA/m^2. *Nakai and Kamide* [1995] also showed that the maximum current of about 130 mA/m is situated around 5.5 Re just prior to onset using their simple numerical model. They depicted that the maximum current of about 80 mA/m is situated around 8 Re during the magnetically quiescent period. Hence the maximum location shifts earthward and the current intensity is amplified. Note that the radial distribution of the integrated current is different from that of the current density [cf. *Pulkkinen et al.*, 1994a]. The enhancement of the ring current during the growth phase [cf. *Iijima and Nagata*, 1972] should largely contribute to the former distribution, while the plasma sheet thinning is essential for the latter one. Considering the radial distance of the estimated onset location of around 10 Re, onset mechanism should be related with the intense current density, rather than the enhanced integrated current intensity. As for the carrier of the intense cross-tail current density, *Mukai et al.* [1998] showed by the direct observation of the GEOTAIL that the dawnward bulk velocity of the electrons is the main contributor, which is consistent with the simulation result of *Pritchett and Coroniti* [1995]. Their data were obtained around 17 Re down the tail, possibly tailward side of the initial NENL, in the final stage of the growth phase (from about 5 min before onset).

(g.3.6) Possible scenario for the plasma sheet thinning

Now it seems that we have enough evidence to check such a conventional view about the thinning of the plasma and current sheets during the growth phase as already mentioned; Increase of the tail flaring angle causes a larger normal dynamic pressure exerted by the solar wind on the tail lobe and a resulting increase of the lobe magnetic field pressure for balancing it; The increase of the lobe magnetic pressure causes a compression and thinning of the plasma sheet which results in the increase of the plasma pressure within it and the magnetic field configuration change toward more taillike. Together with such a process, it is presumed that the transportation of the plasma and magnetic flux from the plasma sheet to the day side is also associated with the thinning [e.g. *Baker et al.*, 1996; *McPherron*, 1994]. There are several counter-evidences to the above view: 1. If the compression by the enhanced lobe magnetic pressure is the cause of the thinning, plasma sheet temperature should increase due to the adiabatic heating, which contradicts the observations which show that the plasma pressure increase is mainly due to the increase of the plasma number density.; 2. Increase of the plasma sheet pressure starts soon after the ionospheric two cell convection pattern

develops after the IMF-Bz southward turning. If the increase of the flaring angle is caused by the transportation of the magnetic fluxes from dayside to the nightside, a substantial time delay should exist until the transported fluxes can enter the tail lobe.; 3. Concentration of the plasma tends to move the magnetic flux and plasma itself tailward not earthward. So the transportation speed of the plasma and magnetic flux to the day side is very slow, and could not play any significant roles for the thinning. It is most probable that the main cause of the thinning is the plasma concentration due to the imposed dawn-to-dusk convection electric field that is propagated from the ionosphere by the Alfvèn wave. The magnetic fluxes are evacuated from the plasma accumulation region tailward and to the higher latitudes due to the diamagnetic characteristics of the plasma, which results in the increase of the lobe magnetic pressure to balance the enhanced plasma sheet plasma pressure. This enhanced lobe magnetic pressure causes the enlargement of the tail lobe radius and the increase of the tail flaring angle. As a result, the pressure balance is maintained also at magnetopause between the solar wind and the tail lobe. Of course, the transported magnetic fluxes from day side should be accumulated in the tail lobe and the dawn-to-dusk electric field induced by the tailward plasma motion around the magnetosheath is directly imposed on the tail lobe region which causes the equatorward motion of the tail magnetic flux and plasma. This process should become more effective as the growth phase progresses and cause further increase of the tail magnetic pressure and the plasma sheet plasma pressure. Hence, both such internal and external mechanisms should be responsible for the thinning especially during the late growth phase. During the early growth phase, the internal mechanism should be dominant. It seems that the conventional view lays so much stress on the magnetic flux transportation, and does not consider enough the importance of the plasma characteristics in the high β plasma sheet.

(g.3.7) Relation between the auroral and magnetospheric processes

The equatorward motion of the auroral oval should be mainly due to the taillike reconfiguration of the nightside magnetic field as estimated by *Sergeev et al.* [1990] because the earthward motion of the source plasma should be very slow. The more rapid equatorward motion of the poleward boundary of the oval implies that the taillike reconfiguration is more significant in the further tailward side of the source region (higher plasma β region) than the nearer earth region (lower β region). The increase of the oval emission intensity during the growth phase indicates the increase of the precipitating energy flux that should be caused by the increase of the energy density (plasma pressure) of the source plasma region. The expansion of the auroral void region in the poleward side, "polar cap", is often considered to be reflecting the increase of the tail lobe magnetic flux, i.e. storage of the magnetic energy in the tail. *Frank and Craven* [1988] tried to estimate the variation of the magnetic flux in the "polar cap" region and the magnetic total energy in the tail using the DE-1 UV imager data. They defined the poleward boundary of the auroral oval for an intensity of about 1 kR. However, as pointed out by *Murphree et al.* [1991] and *Brittnacher et al.* [1999], such a direct association might be a little questionable and even misleading because there are several limitations to the global imager data and the dynamics of the "polar cap" region is not yet fully understood. *Murphree et al.* [1991] pointed out the following limitations of the imager data: 1. The

image data provide information only on regions where significant fluxes of particles precipitate into the ionosphere. Extensive particle populations may exist on field lines with no significant loss cone fluxes.; 2. Even if precipitation could be enough, the resulting emission may not be observed with the wavelength of the imager; 3. or due to the lack of enough sensitivity of the imager; 4. or due to the coarse spatial or temporal resolutions.; 5. The images provide information about features which may or may not be related in terms of their occurrence, hence, they represent the integration of the results of various processes. It is very difficult to distinguish each process.

(g.3.8) Flux change of energetic particles

Another remarkable signature of the growth phase is the flux depletion of the energetic electrons and ions in the near earth region. More extreme depletion or dropout is observed at off-equatorial region, and relatively small depletion is observed near the equatorial plane [e.g. *Sauvaud and Winckler*, 1980; *Lopez et al.*, 1989; *Baker and McPherron*, 1990]. Such a flux decrease progresses as the configuration of the magnetic field lines become more taillike. The field line passing by the satellite at off-equatorial region is stretched further tailward and is connected with the equatorial region of the smaller population. Such a tailward change of the probing point is considered to be a main cause of the extreme depletion at off-equatorial region. On the other hand, near the equatorial plane, the probing point is almost fixed, and a main cause of the small depletion is considered to be the earthward shift of the drift trajectory of the energetic particles. As the growth phase progresses, magnetic field intensity B decreases near the equatorial plane, and the equi-B contours shifts earthward. Since the magnetic drifts of the energetic particles follow the equi-B contours, a larger flux region shifts more earthward and the satellite should observe a smaller flux region that exists at tailward side before. *Sauvaud et al.* [1996] proposed a different mechanism where the deceleration by the inductive electric field associated with the taillike reconfiguration could be the main cause of the flux depletion. However, they did not show the above difference associated with the latitudinal position of observation.

As the flux decreases, the pitch angle distribution gradually changes to the cigar-shaped [*Baker et al.*, 1978] or butterfly type [*Sauvaud and Winckler*, 1980], where there appears a flux depression at 90 deg, both for the energetic electrons [*Baker et al.*, 1978] and ions [*Sauvaud and Winckler*, 1980]. Such a development of the field-aligned anisotropy during the growth phase was analyzed by *Tsyganenko* [1989b] considering the conservation of the first and second adiabatic invariants. As B decreases at the equatorial plane, perpendicular energy should decreases due to the first invariant conservation, whereas, as the field line length between the fixed location at the equatorial plane and the foot on the earth is gradually getting shorter, the parallel energy tends to increase due to the second invariant conservation. As a result, the pitch angle distribution tends to become cigar-shaped.

After the onset, a sudden recovery or further increase of the flux occurs, and the pitch angle distribution changes to the pancake-shaped where there is a maximum at 90 deg. Such an abrupt transition is associated with the sudden change of the magnetic field configuration to more dipolar one. For the off-equatorial observation, the tailward probing point moves back earthward to the

higher flux region with the earthward motion of the dipolarized field line, while, for the near equator observation, higher flux region moves back tailward with the tailward shift of the equi-B contours. In addition, especially in the lower energy range, particles are newly injected from the tailward side, which causes the further increase of the flux beyond the pre-substorm level. In such a dipolarization period, adiabatic heating of the perpendicular component due to the first invariant conservation should be more effective compared with the Fermi acceleration due to the second invariant conservation. As a result, the pitch angle distribution tends to become pancake-shaped. *Birn et al.* [1997a] showed in detail the characteristics of the dispersionless injections (simultaneous flux enhancement at all energy channels) and associated plasma properties both for the electrons and ions observed at geosynchronous orbit. *Birn et al.* [1997b; 1998b] showed by the test particle simulations combined with the three-dimensional MHD simulations that such energizations and injections could be caused by the cross-tail trajectory of the energetic particles along the electric field direction and the fast earthward transport by ExB drift under the intense electric field induced in the dipolarization process in the near earth region. Such an intense inductive electric field is produced in the MHD simulations as a result of the dipolarization at earthward side of the near earth neutral line (NENL) [*Birn and Hesse*, 1996]. Such a dipolarization is caused by the magnetic flux transport and its pileup by the fast earthward flow that is created in the reconnection process at the NENL.

(g.3.9) Field-aligned current distribution

Growing stressed configuration of the magnetic field and change of the plasma pressure distribution and anisotropy during the growth phase should cause a new production and re-distribution of the cross-tail currents in the magnetosphere and their divergence which results in the generation of the field-aligned currents (FACs) flowing into and out from the ionosphere. In the ionosphere, additional electric potential should appear to close the FACs by additional Pedersen currents. Hence, in the ionosphere, both effects from the dayside and the nightside processes are co-existed, and the global potential distribution gradually changes such that the two-cell convection pattern expands equatorward and into the nightside, and the cross-polar cap potential increases. At the same time, the auroral precipitations gradually increase as mentioned above, which causes the increase of the ionospheric Pedersen and Hall conductivities. These increases in the electric field and the conductivities cause the gradual intensification of the equivalent currents observed by the ground magnetometers. Note that the observed equivalent currents are mainly due to the Hall currents because the Pedersen current effect is almost masked with the field-aligned current effect in the nearly uniform conductivity ionosphere [e.g. *Fukushima*, 1969; 1971; 1976]. Such an evolution of the equivalent current pattern can be seen, e.g., in the Fig. 2a of *Kamide and Baumjohann* [1985], and the spatial relationship between the auroral precipitations and the intensified equivalent currents can be seen, e.g., in *Shue et al.* [2000]. The spatial relationship between the auroral particle precipitations and the FACs during the growth phase were analyzed in detail by *Watanabe and Iijima* [1993] by using the DMSP satellite data. They categorized the precipitation region into two domains. One is called the "inner plasma sheet (IPS)" which is located

at equatorward-most part and characterized by an enhanced energy flux of ions and less structured energy spectra of ions and electrons, and the other is called the "outer plasma sheet (OPS)" which is located at higher latitude of the IPS and extended to the plasma sheet-polar cap boundary, and is characterized by a weaker energy flux of ions and a highly structured intensity enhancement for both ions and electrons. They found that the region 2 field-aligned current system roughly collocated with the IPS region and the region 1 with the OPS, and the current intensity is almost balanced latitudinally between the two systems. In the region 1 (region 2) system, the currents flow away from (into) the ionosphere in the evening side and flow into (away from) in the morning side [e.g. *Iijima and Potemra*, 1978]. Although the distribution of the IPS and OPS precipitations is continuous through the so-called Harang discontinuity MLT sector around 23 hr MLT [cf. *Maynard*, 1974], there exists a complicated structure of the multiple field-aligned current sheets around that MLT sector which is the transition region between the evening-type and morning-type region 1 and 2 systems. They presumed magnetospheric processes from their low altitude observations, but they did not consider seriously the processes around the transition region.

(g.3.10) Directly driven and loading/unloading processes

Gradual evolution both of the Pedersen current and the auroral particle precipitations indicates that the energy dissipation process progresses during the growth phase, concurrently with the energy storage process in the magnetosphere. After the expansion phase onset, the stored energy is released and dissipated explosively in various manner: 1. Joule heating with the ionospheric Pedersen currents and atmospheric heating by direct precipitation of the auroral and higher energy particles; 2. Heating and energization of the magnetospheric plasma and resultant enhanced ring current and internal plasma convection; 3. Release and severance of a part of the plasma sheet (plasmoid and post-plasmoid plasma sheet) into the interplanetary space; 4. Electro-magnetic wave generation. For a typical substorm of the total energy about 3×10^{15} J, estimated dissipated energy in each process is about 10^{15} J for the process 1 to 3 [e.g. *Baumjohann and Kamide*, 1984; *Baker et al.*, 1985; *Brittnacher et al.*, 1999; *Harel et al.*, 1981; *Scholer et al.*, 1984; *Ieda et al.*, 1998], and about 4×10^{10} J for the process 4 assuming 1 hr duration with 10^7 W for the AKR (Auroral Kilometric Radiation) power generation [*Gallagher and Gurnett*, 1979]. Hence the total released energy is roughly equipartitioned among all the processes except the last one [*Baumjohann*, 1986; *Kamide and Baumjohann*, 1993]. Such a energy storing/release process is called the loading/unloading process, while the dissipation process during the growth phase is called the directly driven process [e.g. *Baker et al.*, 1985; *Baumjohann*, 1986]. Note that the directly driven process also occurs even after the onset, concurrently with the loading/unloading process. Ionospheric equivalent current variation associated with the directly driven process has been called the DP2 current system, while that with the unloading process the DP1 [e.g. *Nishida*, 1971]. As mentioned above, the DP2 current system should consist both of the dayside and nightside effects, hence, it is driven both externally and internally.

(g.3.11) Riometer absorption

The equatorward motion of the auroral oval during the growth phase is also manifested by the equatorward shift of the maximum cosmic noise absorption (CNA) region observed by the ground-based riometer networks [e.g. *Hargreaves et al.*, 1975; *Pytte et al.*, 1976c; *Ranta et al.*, 1981]. This weak pre-onset absorption was called the "absorption bay" due to the shape of its time evolution. The equatorward speed is 200-500 m/sec [*Hargreaves et al.*, 1975], 83-167 m/sec [*Pytte et al.*, 1976c], and 60-600 m/sec (median is 100-200 m/sec) [*Ranta et al.*, 1981], and the latitudinal width is about 50 km, whereas the longitudinally it may extend for at least 2000 km [*Ranta et al.*, 1981]. The absorption intensity gradually increases during the equatorward movement by about 1.5 times [*Hargreaves et al.*, 1975]. Riometer absorption is considered to be caused mainly by the increase of the electron density in the D-region by the precipitation of energetic (> 30 keV) electrons [e.g. *Friedrich and Torkar*, 1983; *Ranta et al.*, 1985]. *Kirkwood and Eliasson* [1990] showed, using the EISCAT, VIKING, and riometer data, that the precipitation region was situated around the isotropic boundary (IB) for the energetic electrons and those electrons were different from the lower energy plasma sheet origin (they used the term "trapping boundary", but their definition is the same with the IB as already mentioned). They presumed that the precipitation mechanism is the pitch angle scattering by the chaotization of the energetic electrons in the taillike configuration in the near earth region. Such an association was also considered by *Sergeev et al.* [1990]. *Kirkwood and Eliasson* [1990] also showed that the onset arcs were located immediately poleward of the IB, which is consistent with the result of *Yahnin et al.* [1997]. *Collis et al.* [1986] showed, using the EISCAT and riometer data, that the increase of the absorption during the growth phase was mainly due to the increase of the precipitating electron fluxes without significant change of their energy spectra, whereas abrupt spike-like increase of the absorption at the onset should be associated with the sudden hardening of the precipitating spectrum. This suggests that a different precipitation mechanism should be working at the onset. The intense field-aligned acceleration of the energetic electrons and ions observed by the geosynchronous GEOS-2 satellite just after the local onset [*Chen et al.*, 2000, especially for their event 4] could be a possible correspondence to the absorption spike. *Ranta et al.* [1981] showed that the maximum absorption at the onset could be positively correlated with the maximum absorption during the preceding bay, and the absorption is intensified after the onset by about 2 to 4 times larger than the preceding values.

(g.3.12) HF radar characteristics

During the growth phase, the HF radar backscatter region also moves equatorward [e.g. *Lewis et al.*, 1997; 1998]. The equatorward border of the backscatter region can be identified clearly and moves slowly, while the poleward part is more disturbed and moves more rapidly equatorward. Hence, as the growth phase progresses, the latitudinal width of the backscatter region becomes narrower. These are very similar to the optical auroral signatures. Since the HF radar backscatter returns from the field-aligned electron density irregularity in the F-region which is considered to be produced by some instabilities, e.g. gradient drift instability due to the appropriate directional relationship between the density gradient and horizontal electric field [e.g. *Tsunoda*, 1988], the distribution of the backscatter strongly depends on the condition in the F-region, and the spatial

relationship between the HF backscatter and optical auroral distributions during the growth phase has not yet been clarified.

(h) Precursor phenomena prior to the onset

While the growth phase evolution is very gradual taking about 1 to 1.5 hours, the expansion phase starts abruptly within possibly less than 1 sec. It can be expected that there should be some transition phase just prior to the onset, and such precursor phenomena should be very important to understand the onset mechanism.

(h.1) Auroral activity

Weakening of the emission intensity of the discrete auroral arc just prior to the onset has been known as the "auroral fading" [e.g. *Pellinen and Heikkila*, 1978; *Morse and Romick*, 1982; *Kauristie et al.*, 1995; *Kauristie et al.*, 1997]. *Kauristie et al.* [1997] examined ten such events and summarized that the duration of the fading is typically about 2 min, and a local breakup occurred about 2-3 minutes after the fading ended. Hence, the emission intensity recovered just prior to the onset. Usually, the arc that broke up, i.e. the onset arc, had faded earlier. In many cases, the fading occurred just prior to the local onset, not only to the global initial onset. In the 6 cases, the fading showed longitudinal dependence (the fading region propagated in an east-west direction or only part of the arc faded). They presented three possible mechanisms for the fading: 1. Downward field-aligned current generation due to the azimuthal plasma pressure gradient in the nightside magnetosphere could suppress the precipitation of the electrons; 2. Decreasing of the magnetic intensity at the equatorial plane in the taillike configuration should cause the narrowing of the loss cone for the auroral electrons; 3. Increase of the eastward inductive electric field due to the increased cross-tail current should decelerate the electrons that move eastward by the magnetic drift, and should cause the decrease of the precipitating electron energy flux. The role of the strong inductive electric field was also discussed by *Pellinen and Heikkila* [1984]. They called the last few minutes of the growth phase including the auroral fading period as the "trigger phase". They presumed that the auroral arc could be produced by the direct precipitation of the electrons which are energized by the dawn-to-dusk convection electric field during their eastward magnetic drifts. However, it seems to be very difficult to imagine that the very narrow latitudinal width of the arc could be produced by such a precipitation mechanism. In addition, the auroral arcs usually have the inverted-V type energy spectra, which suggests that the electrons are accelerated by the field-aligned electric field which should be closely related with the upward field-aligned currents [e.g. *Knight*, 1973]. Hence, the generation mechanism of the field-aligned current in the magnetosphere should be considered more seriously. The auroral fading was also reported from the satellite observation. Using the VIKING data, *Shepherd and Murphree* [1991] mentioned that a general fading of the aurora seems to occur some 2 min prior to the onset, and then the onset seems to occur perhaps in the middle of the 4-hour MLT sector in the vicinity of a pair of bright spots in the poleward component of the aurora. They showed that the longitudinal extension of the fading region was much larger than the field-of-view coverage of the ground-based all-sky observation.

Elphinstone et al. [1995a] showed that the azimuthally spaced auroral forms (AAFs) appeared within the main UV oval within a few minutes of substorm onset prior to the poleward expansion of the bulge. The AAFs can distribute in 8 hours of local time, while the onset region itself is much localized in about 1 hour. Azimuthal separation (wavelength) of the AAFs is about 130 to 600 km with an average of 307±115 km at 120 km altitude, and associated wave mode number is about 30 to 135. The average location of the onset accompanied by the AAF (AAF onset) was 22.9±1.1 hour MLT and 63.8±3.3 deg CGMLAT, hence at about 2 deg lower in latitude compared with the average location for all the onsets studied by them. The AAF onsets occur when the solar wind pressure is relatively high. *Elphinstone et al.* [1995a] discussed that a possible onset mechanism producing such AAFs could be a modified flute/ballooning instability by *Ivanov et al.* [1992]. However, it is still unclear whether the source mechanism of the AAFs is same with or closely related with the mechanism of the onset, or the two phenomena are completely independent of each other.

(h.2) Equivalent current

A rapid increase of the ionospheric equivalent current in the vicinity of the onset region just prior to the onset was reported by *Kawasaki and Rostoker* [1979]. They showed that the equivalent currents, flowing along the onset-associated discrete arc (onset arc), were intensified about 2 or 3 minutes before the sudden appearance of the surge. Accompanying with the enhancement of the equivalent current, the emission intensity of the onset arc also increased toward the onset. The onset of the Pi2 pulsation coincided with the appearance of the surge. The equivalent current pattern suddenly changed to the surge-associated one after the onset. Their observations implied that both the equatorward electric field across the east-west aligned onset arc and the electron precipitation responsible to the onset arc are rapidly intensified from 2 or 3 minutes before the onset at just eastern side of the onset region. *Kepko and Kivelson* [1999] also showed that the ionospheric current began to increase clearly about 90 sec before the onset of the Pi2 pulsations, and the start time of the increase roughly coincided with a large earthward flow burst observed by the ISEE 2 near X=~17 Re. Although their event was a very low level disturbance, not a well-developed substorm, it seems a good case for studying the temporal relationship between the evolutions of the earthward bursty flows, Pi2 onset, and the substorm current system. After the Pi2 onset, the DC component of the magnetic variation increased. The Pi2 pulsation was clear during about the first 4 minutes after the onset, and then it was damped and its wave form became disturbed. The ISEE 2 did not observe the large bursty flows during the precursor 90 sec and subsequent about 2 minutes after the onset when the clearer Pi2 appeared. After that time, more disturbed flows were observed for about 2 minutes, and then more continuous earthward flows appeared for about 12 minutes which can be categorized in the BBF defined by *Angelopoulos et al.* [1994].

(h.3) Magnetospheric phenomena

(h.3.1) Cross-tail current increase

A rapid characteristic magnetic field variation observed in the near earth region by the

AMPTE/CCE satellite was reported by *Ohtani et al.* [1992b]. The vertical and radial components showed remarkable and rather smooth decrease in their amplitudes from about 1 minute before the local onset of the turbulent variation followed by the dipolarization. After the onset, both amplitudes increased initially and showed a large amplitude longer period (about 1 min) fluctuation overlapped with smaller amplitude shorter period (about 10 sec) turbulence during the first about 5 min, and then the vertical component became steadily enhanced and the radial amplitude depressed, hence, the dipolar configuration was established. They called the about 1 minute period prior to the onset the "explosive growth phase" and interpreted that the rapid variation should be caused by the rapid increase of the cross-tail current intensity on the tailward side of the satellite. They discussed that the magnetic variations could not be explained by the azimuthal expansion of the substorm current wedge system. However their schematic drawing for the observed magnetic variations in their Fig. 12(c) is not correct at least for the radial component final variation, and they almost neglected the fact that the azimuthal component also showed significant, sometimes largest, variations during the "explosive growth phase" in all their cases, indicating a significant influence of the field-aligned currents. The satellite should observe both effects from the cross-tail currents as depicted in their Fig. 11 and from the field-aligned currents in their Fig. 12(a). It seems very difficult, almost impossible, from one point magnetic observation to distinguish whether the observed onset is the initial real onset or the consequence of the expansion of the dipolarization region. Important points are that such an explosive variation should be situated at just outside or appear just prior to the dipolarization, and there is a turbulent region just inside the taillike and dipolelike boundary, or a turbulent period of about 5 min before the full dipolarization. It should be noted that the period of the large amplitude variation during the turbulent period is within the range of the Pi2 pulsation. *Slavin et al.* [1997] reported an event where the rapid decrease of the vertical component was observed simultaneously at wide area in the near earth region by the GOES 9, GEOTAIL and the WIND satellites with similar amplitudes of 5-7 nT a few minutes before the Pi2 onset. Their observation implies that the current source of the "explosive growth phase" could be distant and/or very large scale relative to the spacing of those satellites about 10 Re.

As already mentioned, the event studied by *Mitchell et al.* [1990] and *Sergeev et al.* [1993] also showed that a large duskward and earthward ion bulk flow appeared with a very small magnetic field about 1 min before the local onset of the dipolarization. *Mitchell et al.* [1990] interpreted that the intense ion flow should be caused by the Speiser motion of the non-adiabatic ions and could carry the increased cross-tail currents. Their event was apparently not the initial onset, but the local onset due to the azimuthal expansion of the substorm current wedge, as shown by *Sergeev et al.* [1993]. *Mukai et al.* [1998] showed that around X=-17 Re, the cross-tail current density significantly increased about 5 min before the global onset, and was further enhanced about 1 min before the onset. This increase of the current density coincided with the increase of the tailward flow, the decrease of the Bz, and the increase of the electron and ion temperatures.

(h.3.2) "TCR" variation

A rapid configuration change of the magnetic field lines, starting from a few minutes before the

onset, was also observed by the IMP 8 satellite in the midtail lobe region (X=-26 to -38 Re) [e.g. *Taguchi et al.*, 1997; 1998a; 1998b]. The horizontal (Bx) and total (Bt) intensities sharply increased, reached a peak value, and then sharply decreased to the lower level than the values before the increase. The start time of the increase was about 1-2 min before the ground Pi2 onset [*Taguchi et al.*, 1998b], and the duration of the above variation as a whole was about 4-5 min. The peak time nearly coincided with the Pi2 onset. The vertical component (Bz) showed the bipolar or unipolar transition from a higher value to a lower negative value, and then a recovery to higher level than the minimum value. These variations are similar to the case of the TCR (Traveling Compression Region) observed in the distant tail [e.g., *Slavin et al.*, 1984; 1993], which was interpreted as a consequence of the tailward passage of the plasmoid underneath the satellite. *Taguchi et al.* also interpreted the midtail variation as a TCR due to the formation of the plasmoid. Standing at their point of view, the near-Earth neutral line (NENL) should start to be formed about 1-2 min before the ground onset, and the plasmoid should start to move tailward about 1-2 min after the starting of the NENL formation, hence, around the time of the ground onset. *Taguchi et al.* [1998c] found from the GEOTAIL observation that the plasma flows toward the neutral sheet was enhanced during the later period of the above "TCR" variation in about 3-5 min. They interpreted this velocity enhancement as a consequence of the start of the lobe reconnection at NENL. In their examples, the start time of the velocity enhancement almost coincided with the time of the current wedge formation and the Pi2 onset at ground, and the duration of the enhancement is about 1-2 wave period of the Pi2. They discussed that the plasma sheet reconnection before the lobe reconnection might be important for the ground onset signatures.

As already mentioned, the magnetospheric pressure increase is an essential feature of the growth phase, and should correspond to the increase of the cross-tail current. The rapid increase of the lobe field could be produced by the rapid increase of the cross-tail current that is not necessarily related with the plasmoid formation. After the onset, the cross-tail current intensity should be reduced, and the reduction is more significant in the earthward region of the "TCR" observation. It seems to be possible to explain the "TCR" variation by such a change of the cross-tail current distribution in the tail before and after the onset. The enhancement of the plasma flows toward the neutral sheet implies an enhancement of the dawn-to-dusk electric field. This could correspond to the enhancement of the reconnection rate at the NENL, but it seems very difficult to say definitely whether this enhancement implies the start of the lobe reconnection or not. If the flow enhancement is a consequence of the lobe reconnection, they should answer the question why the enhancement stop soon after the Bz recovers, i.e., why the lobe reconnection stop soon after the plasmoid moves tailward. Basically any kind of magnetic field variations should be related with the variation of the various kind of current system. It seems unfair not to argue about the current system in the discussion of the magnetic field variation.

(h.3.3) NENL formation

Now there seems to be enough evidences supporting the view of the NENL formation before the onset. The fast earthward (tailward) flows with the northward (southward) Bz are expected

earthward (tailward) of the neutral line. Event studies showed that such fast flows appeared, e.g. about 2-3 min before [*Sergeev et al.*, 1995], 5 min before [*Mukai et al.*, 1998], 4 min before [*Ohtani et al.*, 1999], and 90 sec before [*Kepko and Kivelson*, 1999] the ground onset. Statistical studies also showed that both the earthward and tailward flows were observed 1 to 5 min before the ground onset [e.g. *Nagai et al.*, 1998; *Nagai and Machida*, 1998; *Machida et al.*, 1999; *Miyashita et al.*, 2000]. These statistical studies also showed that the NENL is formed initially in the premidnight plasma sheet between XGSM = -20 Re and -30 Re [*Nagai et al.*, 1998], -22 Re and -30 Re [*Nagai and Machida*, 1998], centered around (X, Y) ~ (-20, 5) Re [*Machida et al.*, 1999], and around (-19, 6) Re [*Miyashita, et al.*, 1999; 2000]. All these results showed that the NENL initial location should be tailward from the estimated initial onset location around X ~ -10 Re. The formation of the NENL could be associated with the energization of both ions and electrons and creation of the field-aligned currents, hence some kind of auroral activity associated with the NENL formation can be expected in the ionosphere especially several minutes before the onset. To our knowledge, there are no works ever done before which clearly show the relationship between some specific auroral activities and the NENL formation before the onset. As already mentioned, the discrete auroral activity at higher latitude side of the double oval configuration should be related with the reconnection process at the activated and retreated NENL in the late expansion and recovery phases, and one type of the onset can occur while such configuration still exists. In such cases, the higher latitude auroral activity should not be referred as the consequence of the initial NENL formation.

(i) Onset signatures

(i.1) General description

As mentioned above, timing relation between the phenomena during the several minutes before and after the onset is very important to understand the onset mechanism. Hence, it is important to choose what kind of phenomena to define the onset time. Well-known onset signatures in the ionosphere are the auroral initial brightening and subsequent breakup, sudden increase of the westward electrojet at the auroral latitudes represented by the AE index, start of the magnetic positive and negative bays in the H and D components at the mid- and low-latitudes, start of the Pi2 magnetic pulsations observed at mid- and low-latitudes, and start of the Pi1 magnetic pulsation burst (Pi1B) observed at auroral to low-latitudes. In addition, characteristic magnetic variations around the auroral bulge or surge can be used to identify the onset [cf. *Kisabeth and Rostoker*, 1973; *Tighe and Rostoker*, 1981; *Opgenoorth et al.*, 1980]. Onset signatures in the magnetosphere are the start of the magnetic variation of the dipolarization or the current disruption observed in the near earth region, the start of the dispersionless flux recover or increase (injection) of the energetic particles in the near earth region, the sudden total pressure decrease observed in a broad region, and the enhancement of the auroral kilometric radiation (AKR). Because the onset occurs in a very limited localized region both in the ionosphere and magnetosphere, and the onset disturbances propagate and expand both radially and azimuthally afterward, signatures which can be observed in a wider region almost simultaneously beyond the limited onset region are more favorable for the onset identifier. Otherwise, global and dense network of the limited FOV (field of view)

observations or a global FOV observation are needed. In the first sense, the mid and low latitude signatures (positive and negative bay, Pi2, and Pi1B) in the ionosphere and the total pressure decrease in the magnetosphere are more favorable, and in the second sense, the long period observation of the satellite global auroral imager or magnetospheric imaging are desirable. To determine the onset location, the global auroral imager data are very essential.

(i.2) Timing relation between the auroral brightening and the Pi 2 onset

Superiority of the auroral breakup as the onset indicator was discussed by *Liou et al.* [1999; 2000a]. *Liou et al.* [2000a] checked the timing relation between the Pi2 observed at Kakioka and the auroral breakup observed by the POLAR UV imager, and found that the Pi2 onset tends to lag behind the auroral breakup by 1 to 3 min with the peak around 1 min. They determined the auroral breakup time by the time of the apparent start of the luminosity increase and the bright area expansion, while they determined the Pi2 onset time by subtracting a quarter of the wave period from the time at the wave amplitude peak. Their conclusions seem to strongly depend on these definitions. Especially the definition of the Pi2 onset time seems too artificial and automatic. There are no reasons to believe that the maximum amplitude should appear first at the onset. Actually in their one example shown in their Fig. 2, much smaller amplitude wave preceded and nearly coincided with the auroral breakup, and it seems better to choose this smaller one as the Pi2 signature at the onset. Note that the maximum amplitude of the Pi2 coincided with the acceleration of the positive bay in that case. They certainly mentioned that a sharp increase in the field intensity of the Pi2 pulsations coincides with the auroral breakups. If we consider this sharp increase as the onset signature of the Pi2, then their result will become different such that both onset signatures occur almost simultaneously. Moreover, as already mentioned, the auroral luminosity often starts to increase several minutes before the onset. Hence the apparent increase of the luminosity does not always indicate the start of the expansion phase. As a summary, when we discuss the timing relation between the various onset phenomena, we must consider and know better about the features and physical meanings of the phenomena before discussing only the timing.

(i.3) Substorm Current Wedge (SCW) system evolution

At the onset time, the cross-tail current density increased during the growth phase should be reduced suddenly, hence the cross-tail current disruption should occur in a very localized region in the near earth magnetosphere [e.g. *Ohtani et al.*, 1991]. The disrupted currents are diverted along field lines, flowing into the ionosphere on the dawnside and flowing out from the ionosphere on the duskside of the localized region, which forms so-called the substorm current wedge (SCW) system [e.g. *McPherron et al*, 1973].

Ionospheric closure of the SCW system is considered as the increased westward electrojet flowing initially in a limited longitudinal range at the auroral latitudes, and it is manifested by the appearance of the DP1 equivalent current system. In a highly inhomogeneous ionospheric conductivity distribution after the expansion phase, ground magnetometer can observe both the effects of the Pedersen currents and the field-aligned currents in addition to the Hall currents

especially at remote sites apart from the conductivity enhanced region [e.g. *Opgenoorth et al.*, 1983]. The positive and negative bays in the H and D component at the mid- and low-latitudes are usually considered to be caused by such field-aligned current effects. However the ionospheric electric potential distribution after the onset shows that the positive and negative potentials are situated at north-east and south-west sides of the conductivity enhanced region, respectively [e.g. *Lu et al.*, 1998], rather than simply at east and west sides as assumed by the current wedge model. Hence the intense westward electrojet flowing in the main body of the bright auroral region is mainly due to the Hall currents, and the currents most effective to the mid- and low-latitudes magnetic variation should be also the Hall currents returning from duskside to dawnside in the lower latitude side of the main body. Such a spatial relationship between the auroral bulge and the SCW field-aligned current location deduced from the mid-latitude magnetogram as shown by *Sergeev et al.* [1996b] could be reasonably understood by the flow pattern of the Hall currents. Similar kind of return currents appear in the higher latitude side polar cap region, as usually observed [e.g. *Kokubun and Iijima*, 1975]. Such a potential distribution should be a combination of the primary one directly related with the current wedge system and the secondary one related with the charge accumulation in the ionosphere caused by the primary poleward Hall currents in the enhanced conductivity region as predicted by *Fukushima* [1969]. Hence most of the intense westward electrojet should not be the main part of the ionospheric closure of the SCW system. *Aikio and Kaila* [1996] showed, using the EISCAT data and the all-sky camera data, that a very high conductance region is situated at the poleward boundary of the poleward expanding bulge and intense both Pedersen and Hall westward currents flow in that region. Within the diffuse auroral region at lower latitudes, the electric field mainly directs southward and the intense westward current is mainly Hall current. Hence it is supposed that the ionospheric closure current of the SCW system might flow in the latitudinally narrow region around the poleward boundary of the auroral bulge which may be too narrow to identify in the numerical calculation deduced from a rather sparse distribution of the magnetometer stations as *Lu et al.* [1998]. Corresponding to the north-south separation of the electric potentials and the southward directed electric field, a latitudinal closure current system should exist in addition to the longitudinal closure of the SCW system, which indicates that an additional field-aligned current - tailward radial current system should be created in the magnetosphere during the expansion phase.

Apparent start of the magnetic bay at mid and low-latitudes often coincides with the ground Pi2 onset [e.g. Fig. 3 of *Saka et al.*, 1996], or shows an appreciable delay from the Pi2 onset [e.g. Fig. 13 of *Shiokawa et al.*, 1998a; Plates 2 and 4 of *Liou et al.*, 1999]. Since the characteristics of the magnetic bay in the H and D component is closely related with the location of the observatory relative to the onset region, we must check both the H and D components at various stations distributed in a wide area around the onset region, when we try to discuss the onset timing. For example, in the case of *Shiokawa et al.* [1998a], the Pi2 onset recorded at Hermanus and Ottawa seems to nearly coincide with the positive bay onset in the D component at Sanae, hence it could be considered that both onset signatures appeared almost simultaneously though the magnetic bay activity should be rather localized initially. In their event, more global evolution of the magnetic

bay, i.e. further evolution of the SCW system started being coincident with the start of the more disturbed fast earthward flows in the magnetosphere. Similar kind of correspondence between the fast flows and SCW evolution can be seen in the event of *Kepko and Kivelson* [1999].

In the current disruption (CD) model [e.g. *Lui*, 1996], the diversion of the cross-tail current should be associated with the current reduction in the CD region due to the magnetic field turbulence caused by some kinetic instability (cross-field current instability (CCI)) in that region. It should be noted that such a turbulence usually appears during the transition phase to the dipolelike configuration and its duration is rather short (about 5-10 min), while the duration of the SCW evolution is generally much longer. As discussed by *Lui* [1996], the field-aligned current generation mechanism should be related with the less disturbed more dipolar region rather than the disturbed CD region. This point seems not to be explained consistently in the CD model.

Birn and Hesse [1996] and *Birn et al.* [1999] showed in their MHD simulations that the dipolarization and the SCW field-aligned current formation could be produced in the near earth region around X=-10 Re by the magnetic and plasma flux transport by the fast earthward flows and their braking and azimuthal deflection, and such fast flows are produced by the reconnection process at the NENL on the more tailward side around X=-23 Re. *Birn et al.* [1999] analyzed their results in detail and found that the most effective source mechanism of the SCW field-aligned current system might be the divergence of the diamagnetic currents due to the enhanced pressure gradient in the Z-direction combined with the changes in By and Bx. As already mentioned, *Birn et al.* [1997b; 1998b] could also reproduce the dispersionless injections of the energetic particles by the creation of a large dawn-to-dusk inductive electric field in the near earth region due to the dipolarization. Hence, in their simulations, almost all the essential onset signatures in the magnetosphere could be sub-products of the reconnection process on the tailward side.

(i.4) Dipolarization, injection, particle precipitation and Pi2 generation

Due to the reduction of the cross-tail current, the field line configuration tends to change suddenly from the very stretched taillike one to the more dipolelike one, hence the dipolarization occurs in the localized region. These configuration change and the cross-tail current reduction should cause the recover and further enhancement (injection) of the energetic particle fluxes, as already mentioned, and the acceleration and energization of the auroral particles both in the perpendicular and parallel directions to the field lines. The parallel accelerated auroral and more energetic particles precipitate in the ionosphere, and cause the auroral brightening and the sudden increase of the riometer absorption. The AKR is intensified possibly relating with the enhanced discrete auroral activities at the western side or poleward-most of the auroral bulge. It is very natural to suppose that such a large configuration change causes a large scale field line oscillation with a characteristic period of the Pi2 pulsation range, 40-150 sec.

(i.5) Magnetospheric pressure decrease

In the NENL model, the sudden pressure decrease of the magnetosphere can be interpreted as the sudden decrease of the tail lobe magnetic flux by the start or increase of the reconnection at the

NENL. As already mentioned, it is more probable that the pressure increase in the growth phase is mainly due to the plasma pressure increase in the plasma sheet. If we stand at such a point of view, the sudden pressure decrease should be caused by the sudden plasma pressure decrease in the onset region.

(i.6) Fast flows

Statistical studies showed that the velocities of the fast earthward and tailward flows clearly increase just after the Pi2 onset [e.g. *Nagai and Machida*, 1998; *Machida et al.*, 1999; *Miyashita et al.*, 2000], which implies an enhancement of the reconnection rate at the NENL just after the onset. Such an enhancement of the reconnection rate can also be seen in the results of *Taguchi et al.* [1998c] as already mentioned. In the event studies, it seems a common feature that the nature of the fast flow changes from the rather smooth variation before the onset or in the preceding phase to the more disturbed or bursty one after the onset or in the subsequent phase [e.g. *Kepko and Kivelson*, 1999; *Shiokawa et al.*, 1998a; *Mukai et al.*, 1998; *Ohtani et al.*, 1999]. In the near earth region, the fast earthward flows are observed almost always concurrently with the current disruption features [e.g. *Angelopoulos et al.*, 1992; *Fairfield et al.* 1999].

(i.7) Relationship between the current disruption, Pi2, and Pi1

As already mentioned, the current disruption features seem to be often composed of two distinct stages: a large amplitude and lower frequency (Pi2 range) variation overlapped with a less disturbed smaller amplitude and higher frequency (Pi1 range) one, followed by more disturbed and bursty nature of the higher frequency and lower amplitude variation overlapped on the rather steady transition to the dipolar configuration [e.g. *Sakurai and McPherron*, 1983; *Takahashi et al.*, 1987; *Ohtani et al.*, 1992b; *Saka et al.*, 1996]. *Yumoto et al.* [1989] showed that the current disruption onset observed in the near earth region by the AMPTE/CCE tends to be preceded by the Pi 2 pulsation onset observed at low-latitude ground stations by about 0-3 min. Since it might be a frequent case that the satellite observes after the disruption region expands to the location as shown by *Lopez and Lui* [1990], this observation suggests that the current disruption onset does not so often precede the ground Pi 2 onset or both should occur almost simultaneously. They also showed that in the event of *Takahashi et al.* [1987], Pi 1 pulsations of similar period were observed concurrently both at the AMPTE/CCE and at the ground with similar duration. Their ground conjugate observation showed that the Pi 1 pulsation should be an odd-mode oscillation of the geomagnetic field lines, and the satellite observation showed that the pulsation is a compressional mode. It should be noted that the concurrent Pi 1 oscillation appeared about 90 sec after the dipolarization onset and about 30 sec after the ground Pi 2 onset. Hence, there seems to be two distinct stages also in the magnetic pulsations observed at ground similar to the current disruption features: rather pure Pi 2 pulsation period is followed by the more disturbed period with the Pi 1 pulsations. Such a temporal relationship between the Pi 2 and Pi 1 pulsations can be seen in the other examples [e.g. *Bösinger and Wedeken*, 1987; *Yumoto*, 1990].

(i.8) Characteristics of the Pi2 pulsation

The Pi 2 pulsation is considered to be an odd-mode fundamental oscillation of the geomagnetic field lines anchored at the onset location in the ionosphere [*Kuwashima and Saito*, 1981]. The Pi 2 amplitude has a maximum around the onset latitude and decreases toward lower and higher latitudes. It has a smaller secondary maximum around the plasmapause latitude [e.g. *Saito et al.*, 1976; *Kuwashima and Saito*, 1981; *Lester and Orr*, 1983; *Yumoto et al.*, 1994]. A phase reversal of the H component can be seen around that location [e.g. *Saito et al.*, 1976; *Lester and Orr*, 1983; *Yumoto et al.*, 1994]. Pi 2 is observed almost simultaneously over the wide latitudinal range from the auroral region through low latitudes with a common predominant period [e.g. *Kuwashima and Saito*, 1981] even in the dayside especially at low latitudes [*Yumoto*, 1990]. At polar cap stations, Pi 2 tends to be observed at earlier time than in the lower latitudes possibly because of the higher Alfvèn velocity in the lobe region [*Uozumi, et al.*, 2000]. The Pi 2 period is closely related to the latitude of the associated onset location. It becomes shorter (longer) when the onset occurs at lower (higher) latitudes [e.g. *Saito et al.*, 1976; *Kuwashima and Saito*, 1981], which can be understood by the oscillation of the closed field lines anchored at the onset location. The occurrence of the Pi 2 is concentrated in the nightside with a maximum around 2200-2300 LT [*Fukunishi*, 1975a]. Similar local time dependence of the occurrence is observed at geosynchronous orbit [e.g. *Sakurai and McPherron*, 1983]. In the local time sector 1900-0300 LT, satellite observes the Pi 2 events almost always simultaneously without significant time lag from the ground Pi 2 events within 1 min. In the magnetosphere, Pi 2 pulsation generally have a compressional component with a transverse one. The initial perturbation in the azimuthal component is in the same sense as the DC shift caused by the SCW field-aligned current system, i.e. eastward in premidnight, westward in postmidnight, when the satellite is located in the northern region [e.g. *Sakurai and McPherron*, 1983; *Saka, et al.*, 1996], which suggests that there may be a very close association between their causative mechanisms. *Holter et al.* [1995] showed that the Pi 2 oscillation observed at off-equatorial region by the geosynchronous satellite GEOS 2 had signatures of standing waves in a coupled mode of the shear Alfvèn and slow magnetosonic waves. They suggested that the standing waves should be trapped in the thin current layer that developed prior to the onset and could be observed during the transient phase only from the taillike to dipolelike configurations. *Takahashi et al.* [1992] showed that the Pi 2 oscillations observed in the inner magnetosphere (L=2-5) by the AMPTE/CCE could be identified as standing waves in the odd-mode and poloidal mode in the field-aligned direction. They also showed that there should be a standing wave structure also in the radial direction possibly because of the reflection of the MHD fast-mode waves between the inner and outer boundaries, and the node location of the standing wave structure should be around 3 Re. There often appear several Pi2 activities associated with several intensifications in one substorm event. The first one should be considered as the onset indicator [e.g. *Rostoker et al.*, 1980].

(i.9) Characteristics of the Pi1 pulsation

The Pi 1 magnetic pulsations (oscillation period: 1-40 sec) have been known to be categorized into two types, Pi1C and Pi1B [e.g. *Heacock*, 1967]. The Pi1C is characterized by its gradual start

and long (1-3 hours) continuous narrower frequency band appearance in the frequency-time (f-t) spectra with a broad maximum frequency around 0.1 Hz. Its occurrence has a maximum around 0400 MLT, and is well correlated with the evolution of the magnetic bay and auroral pulsations in the night and morning hours. *Oguti et al.* [1984] showed that the Pi1C pulsations could be caused by the spatial and temporal variations in the ionospheric electric conductivity produced by the pulsating precipitation of auroral electrons into the pulsating auroral patches. Hence the occurrence of the Pi1C should be localized in the pulsating auroral locations. Pi1B is characterized by its impulsive start and short (1-20 min) bursty broadband appearance in the f-t spectra with amplitude maximum around 0.2-1.0 Hz. Its occurrence well coincides with the substorm onset, hence Pi1B has been considered as an indicator of the onset. Its observed amplitude is maximum on the ground just underneath active auroral forms, and decreases steeply both at higher and lower latitudes. The high and mid latitude observations exhibit a high degree of correlation in the start time, duration and temporal fine structure, but not in the amplitude behavior. Pi1Bs are often observed simultaneously within 5 sec over a large range of latitudes and longitudes in the nightside hemisphere [*Bösinger and Wedeken*, 1987].

(i.10) Summary of the timing relation between the onset signatures

Now we summarize the timing relation considering the following points:

(s1) Pi2 onset on the ground (Pi2(G)) occurs with the Pi2 onset in the magnetosphere (Pi2(M)) within 1 min. [e.g. *Sakurai and McPherron*, 1983]

(s2) Pi2(G) onset almost coincides with the auroral brightening (AB) within 30 sec. [e.g. *Liou et al.*, 2000a]

(s3) Pi2(G) onset precedes the Pi1B onset on the ground (Pi1B(G)). [e.g. *Bösinger and Wedeken*, 1987; *Yumoto*, 1990]

(s4) Pi2(G) onset precedes the start of initial phase of current disruption (CD) by 0-3 min. [*Yumoto et al.*, 1989]

(s5) CD is composed of two distinct stages: CD1 and CD2. CD1 contains large amplitude Pi2 range variations, and CD2 does not. [e.g. *Sakurai and McPherron*, 1983; *Ohtani et al.*, 1992b; *Saka et al.*, 1996]

(s6) Fast flow enhancement (BBF2) starts about 0-2 min after Pi2(G) onset. [e.g. *Nagai and Machida*, 1998; *Machida et al.*, 1999; *Miyashita et al.*, 2000]

(s7) CD occurs simultaneously with the fast flow. [e.g. *Angelopoulos et al.*, 1992; *Fairfield et al.* 1999]

(s8) Dipolarization (DP) is completed after the CD features and BBF2 end. [e.g. *Fairfield et al.* 1999]

(s9) Substorm current wedge (SCW) system often shows two step evolution during the early stage of the expansion phase: SCW1 and SCW2. Start of SCW1 nearly coincides with the Pi2(G) onset, and start of SCW2 with the start of BBF2. [e.g. *Shiokawa et al.*, 1998a; *Kepko and Kivelson*, 1999]

(s10) Auroral bulge expansion (ABE) also shows two step evolution
corresponding to SCW1 and SCW2: ABE1 and ABE2, as discussed below.
[e.g. *Liou et al.*, 1999; *Liou et al.*, 2000a]

(s11) Precursor fast flow (BBF1) appears before the Pi2(G) onset
[e.g. *Nagai et al.*, 1998; *Nagai and Machida*, 1998; *Machida et al.*, 1999; *Miyashita et al.*, 2000]

Possible timing relation is as follows:

BBF1 -> Pi2(M), CD1? -> Pi2(G), AB, ABE1, SCW1, CD1, BBF2?
-> CD2, BBF2, Pi1B(G), SCW2, ABE2 -> DP, SCW2, ABE2

We suppose that the Pi2 waves are first launched in the equatorial plane and propagate down to the ionosphere with the Alfvèn velocity. Point (s1) suggests that the Alfvèn wave travelling time from the equatorial plane to the ionosphere should be smaller than 1 min. Detailed timing between CD, BBF2, and Pi1B(G) is unclear. The fast flows, BBF1 and BBF2, are considered to be produced by the reconnection process at the NENL. Some schools tend to consider that the Pi2 pulsations should be caused by the fast earthward flows [e.g. *Baker et al.*, 1996; *Shiokawa et al.*, 1998a]. However such a cause-effect relationship between BBF1 and Pi2(M) has not yet been clarified. Evolution of the BBF1 seems rather gradual compared with the sudden and pure appearance of the Pi2(G). Hence, it seems a little difficult to suppose that the BBF1 could be a direct cause of the Pi2(G) onset. The direct cause should be related with the other precursor phenomena as already mentioned, and the period just prior to the onset when such precursor phenomena appear could be called the "trigger phase" as speculated by *Ohtani et al.* [1999]. From the point (s4), CD itself might not be the direct cause of the Pi2(M) onset. From the point (s5), CD1 could be the same with Pi2(M). As already mentioned, the main characteristics of CD2 and Pi1B(G) are very similar to each other, especially concerning their bursty nature, oscillation frequency range, and duration of appearance. Pi1B(G) could be an ionospheric feature of CD2, which propagates from the equatorial plane to the ionosphere with the Alfvèn velocity along the field-lines. From the point (s7), when CD occurs, NENL should already exist, and such a simultaneity suggests a close relationship between CD and BBF2. From the point (s9), further development of the SCW system (SCW2) might be closely related with BBF2 and DP.

As for the auroral activity, the initial bulge rapidly expands poleward during the period of SCW1, almost reaches a ceiling latitude at the start time of SCW2, and the bulge expands mainly in the azimuthal direction after that time [e.g. two examples in *Liou et al.*, 1999; one example in *Liou et al.*, 2000a]. Hence the auroral bulge expansion (ABE) also seems to have two distinct stages corresponding to the SCW1 and SCW2: ABE1 and ABE2.

In summary, it can be said that the ground Pi2 onset is a good indicator of the substorm expansion phase onset, because the Pi2 onset might be one of the signatures which appear first after the onset. Above estimation of the timing relation is based on various observations concerning

various combinations of the various onset signatures. It should be necessary to do more detailed event and statistical analyses concentrating on the timing relation using more complete data set including the important onset signatures as many as possible observed both in the ionosphere and magnetosphere. In such data set, global auroral imager data are essential.

(j) Evolution after the onset

(j.1) Auroral bulge expansion

(j.1.1) Two step evolution in the early stage

As mentioned above, there seems to be two distinct stages in the auroral bulge expansion during the early stage of the expansion phase. The first stage is characterized by a rapid poleward expansion and a rapid asymmetric azimuthal expansion after the onset, and the subsequent stage is by a much slower poleward expansion and a slower symmetric azimuthal expansion [e.g. Plates 1 and 3 of *Liou et al.* 1999; Fig. 1 of *Liou et al.*, 2000a]. *Kaneda and Yamamoto* [1991] noticed that there should exist such kind of two-stage evolution, and showed two examples of the AKEBONO UV imager data. They named the two stages as the initial brightening and the flaring-up, respectively. They discussed mainly about the evolution of the azimuthal extent of the bulge. *Shepherd and Murphree* [1991] showed an example where the poleward expansion of the bulge stopped around the latitudes of the pre-existing arc about 1 min after the onset, and after that time, the bulge expanded azimuthally. Duration of the first stage is about 4 min for the two events of *Liou et al.* [1999], about 3 min for the one event of *Liou et al.* [2000a], about 5 min and 3 min for the two events of *Kaneda and Yamamoto* [1991], and about 1 min for the event of *Shepherd and Murphree* [1991], hence the duration or the interval between the first and second stages is rather short within 1-5 min. Although this kind of two-stage evolution seems to be a common feature of a typical auroral substorm, there are very few previous works showing such evolution clearly except above mentioned examples, and only *Kaneda and Yamamoto* [1991] pointed out its importance in the auroral dynamics.

(j.1.2) Further poleward expansion in the later stage

A well developed bulge often shows a rather sudden further expansion of its size both in the poleward and azimuthal directions. Such a further expansion can be seen in the two examples of *Brittnacher et al.* [1999], one example of *Cummer et al.* [2000], and three examples of *Henderson et al.* [1998]. Start time of the further expansion after the onset is about 12 min and 15 min for the events of *Brittnacher et al.* [1999], and between 31 min and 39 min for the event of *Cummer et al.* [2000]. Onset times were not described for the events of *Henderson et al.* [1998]. As shown in *Henderson et al.* [1998], the further expansion often occurs with an activation of the most poleward arc system, and the "double oval" configuration clearly appears after the further expansion. This kind of stepwise expansion is also identified in the three examples of *Craven and Frank* [1987], and the start time of the further poleward expansion after the onset is about 18 min, 16 min, and 20 min for each event, although the temporal resolution of the DE 1 imager is about 6 min. The third event of *Craven and Frank* [1987] was discussed by *Hones* [1985] as a possible auroral manifestation of

the "poleward leap" of the auroral electrojet which coincides with the thickening of the plasma sheet in the outer magnetotail (beyond ~15 Re) late in a substorm [e.g. *Hones et al.*, 1970; *Pytte et al.*, 1978; *Hones et al.*, 1984].

Hones et al. [1987] analyzed an event using the VIKING UV auroral data, ground magnetometer data, geosynchronous satellite and near tail (X~-17 Re) ISEE 1 data. They found that an apparent "poleward leap" of the aurora occurred about 20 min after the onset. This auroral "poleward leap" did not immediately initiate the subsidence of the auroral zone electrojet and "poleward leap" of the electrojet to higher latitudes, contrary to the expectation from their previous analyses [e.g. *Hones et al.*, 1984]. Actually, the auroral "poleward leap" occurred around the time when the auroral zone electrojet was re-intensified and a high latitude bay started to develop. At the ISEE 1 location, the plasma sheet thinning occurred at the expansion phase onset and its thickening or recovery started when the auroral zone electrojet subsided enough and the high latitude bay reached its peak intensity, hence the central location of the electrojet shifted to higher latitudes, which is consistent with their previous works. At that time, the auroral "poleward leap" almost stopped and a "double oval" configuration clearly appeared in the UV data with a very dim diffuse auroral region at lower latitudes and a relatively intense poleward-most discrete auroral activity. This event showed that the "poleward leap" of the auroral bulge does not coincide with that of the electrojet, and the "poleward leap" of the electrojet should be closely related with the variation of the latitudinal distribution of the ionospheric conductivity due to the auroral bulge evolution, hence, the concept of "poleward leap" seems to be too simplified one and very misleading as pointed out by *Rostoker* [1986]. Although the "poleward leap" of the electrojet is a very complex phenomena as above mentioned, the "poleward leap" of the auroral bulge was a real phenomena clearly observed in their data, and was very similar to other examples as already mentioned.

In summary, this further expansion of the auroral bulge in the later stage seems to be also a very common feature of a typical well-developed substorm, although there are also very few previous works which clearly show such kind of the further expansion except above mentioned examples, and only *Hones et al.* [1987] pointed out the importance of the further expansion and tried to understand it in context of the magnetospheric substorm. Depending on the above examples, the further expansion starts about 12-39 min after the onset. Considering the two step evolution in the early stage mentioned in the previous section, there should be three distinct stages in the auroral bulge expansion, and the duration of the first stage is about 1-5 min and that of the second stage is about 7-38 min. To our knowledge, there are no previous works showing a full process of such a three stage evolution of auroral bulge and pointing out its importance.

(j.2) Magnetospheric phenomena

(j.2.1) Changes in the magnetospheric configuration and the NENL location

As already mentioned in the section (g.3.2), timing of the plasma sheet thinning and subsequent thickening depend on the distance from the earth. This can be interpreted by the near earth neutral line (NENL) formation initially around the distance of 15 Re and its subsequent tailward retreat [e.g. *Nishida and Nagayama*, 1973b; *Nishida and Fujii*, 1976; *Pytte et al.*, 1976a;

Hones et al., 1987].

Asano et al. [1998] studied the relation between the fast flows, which should be originated from the NENL, and the dipolarization, which should be closely related with the plasma sheet thickening, observed by the GEOTAIL in the near tail region. They found that in the near earth region, dipolarization occurs just after the earthward flows have ceased, and it propagates down the tail with a speed of about 30 km/s. In the region of X_{GSM} < -15 Re, flow reversal events from tailward to earthward are observed. In such events, the first earthward flows are usually observed in the plasma sheet boundary layer (PSBL) and rather field-aligned, and the dipolarization occurs later after the earthward bulk flows appeared in the central plasma sheet (CPS). Hence, on the tailward side of the initial NENL, the plasma sheet recovery often occurs before the dipolarization, and sometimes the dipolarization itself is unclear. It should be noted that at the VELA satellite location, radial distance of about 18 Re, the time of maximum Bz is about 60-70 min after the onset judging from the Fig. 5 of *Asano et al.* [1998], which is roughly consistent with the time delay of the plasma sheet recovery at the VELA location from the onset, which is shown in *Hones et al.* [1984].

Baumjohann et al. [1999] also analyzed the relation between the fast flows and the dipolarization observed by the GEOTAIL. They found that the dipolarization moves tailward from the innermost region (11-16 Re) with an average velocity of about 35 km/s, very similar value to *Asano et al.* [1998], and the NENL location clearly starts to move tailward after the dipolarization front reaches that location around 21 Re about 45 min after the onset. *Baumjohann et al.* [1999] concluded that the recovery phase should start at that moment. It can be seen in their Figures 2 and 4 that on the tailward side of the initial NENL, further thinning occurs after the onset before the arrival of the dipolarization front. It should be also noted that in the innermost region, magnetic configuration gradually changes from an extra-dipolar one to a normal one when the dipolarization or plasma sheet thickening proceeds in the more distant region. This plasma sheet recovery to more relaxed configuration or plasma sheet re-thinning occurs soon after the extra-dipolarization about 15 min after the onset.

(j.2.2) Plasmoid evolution

In the NENL model, the plasmoid is formed after the reconnection of the plasma sheet field lines begins, and it starts to move tailward after the reconnection of the lobe field lines begins [e.g. *Hones*, 1977]. Statistical analysis by *Ieda et al.* [1998] showed the evolution of various characteristics of the plasmoid observed by the GEOTAIL in the distance of X=-16 to -210 Re. *Nagai et al.* [1994] estimated the tailward velocity of the plasmoid using the B_Z zero crossing time of the bipolar B_Z variations, and *Nagai and Machida* [1998] did using the arrival time of the tailward flows. They derived similar average velocities of 775 km/s and 760 km/s in the distance of X=-50 to -210 Re, respectively. Using the time delay between the Pi2 onset and the arrival time of the traveling compression region (TCR) observed by the ISEE 3, *Slavin et al.* [1993] derived a similar value of 794 km/s in the distance of X=-50 to -250 Re. More complete analysis by *Ieda et al.* [1998] showed that the tailward velocity depends on the distance: 350, 470, 680, 590 km/s for the distance range of (-16,-30), (-30,-50), (-50,-100), (-100,-210) Re, hence a little lower than the

values of other works. Considering the possible initial location of the NENL around X=-20 to -30 Re [e.g. *Nagai et al.*, 1998] and the plasmoid analyses by *Nagai et al.* [1994] and *Ieda et al.* [1998], plasmoid should start to move tailward about 10 min after onset in average, hence the lobe field line reconnection should start around that time. As a reference, in a global MHD simulation by *Ogino and Walker* [1998], the tailward ejection of a plasmoid starts 15 min after the start of the NENL formation.

(j.2.3) Radial propagation of the current disruption

Radial propagation of the current disruption region in the near tail was analyzed by *Jacquey et al.* [1991; 1993] and *Ohtani et al.* [1992a] using the ISEE 1 and 2 magnetic field data. Their results suggested that the propagation generally proceeds tailward from earthward location of the satellites situated at X=-11~-22 Re with a velocity of about 150-300 km/s, about one order of magnitude larger than the tailward velocity of the dipolarization mentioned in the section (j.2.1). Their data also showed that the current disruption front reached the satellite distance about 6, 5, 6, 2, and 3 min after the onset for the distances of 20.7, 14.4, 15.5, 11.3, and 12.8 Re, respectively, about one order of magnitude faster than the dipolarization. Hence it could be concluded that the initial information of the dipolarization (current disruption) should propagate tailward in the near earth region with a much faster velocity than the velocity of the tailward propagation of the full dipolarization.

(j.2.4) Azimuthal expansion of the substorm current wedge (SCW)

Nagai [1991] showed in his statistical analysis that the dipolarization at the geosynchronous orbit occurs in a narrow local time sector centered at 2330 MLT, and expands both eastward and westward with velocities of about 3.5 hr MLT/11 min and 3.5 hr MLT/15 min, i.e. about 4.8 deg/min and 3.5 deg/min, respectively. Outside of the dipolarization region, the field becomes more taillike, hence the growth phase feature continues until the start of the local dipolarization. The plasma injection observed at geosynchronous orbit also shows similar features of the localized initial onset around the premidnight region and afterward longitudinal expansion both in the eastward and westward directions with about 10-100 km/s, hence about 0.8-8.2 deg/min [e.g. *Arnoldy and Moore*, 1983]. *Arnoldy and Moore* [1983] showed that the azimuthal expansion speed expressed by deg/min is about 1.8 times larger in the ionosphere than at the geosynchronous orbit possibly because of the magnetic field configuration stretched toward midnight. They also suggested that the strong acceleration and precipitation of the particles associated with the injection should take place at only just behind of the azimuthally expanding fronts lasting for about 10 min. *Pulkkinen et al.* [1995] compared the longitudinal expansion of the auroral bulge observed by the VIKING UV imager with that of the SCW deduced from the satellite observations in the near earth tail, and showed that both are very similar especially for the eastward expansion. This could be an evidence indicating that the auroral bulge should correspond to whole of the dipolarization and injection region in the magnetosphere, not just to the source region of the upward field-aligned currents, from very beginning of the bulge formation. The expansion speed in their case is about 5

deg/min in the magnetosphere.

(j.2.5) Pressure change in the near earth tail

Kistler et al. [1992] showed that both the plasma sheet plasma pressure and total pressure tend to increase after the injection in the near earth tail within about -14 Re, contrary to the total pressure decrease in the more tailward region. *Baumjohann et al.* [1991] showed that as an average feature, ion temperature increases and density decreases after onset in the central plasma sheet in X=-10 to -19 Re, while at geosynchronous orbit, *Birn et al.* [1997a] showed that temperature, density, and pressure of ions increase all together after the injection onset. These results suggest that the nonadiabatic heating process shown in the result of *Baumjohann et al.* [1991] should be effective in the region of about X=-14 to -19 Re, and in the nearer earth region, adiabatic heating should become effective. The extra-dipolar configuration and the enhanced plasma pressure in the nearer earth region should cause a re-distribution of the cross-tail current density in the plasma sheet, and could produce an enhancement of the integrated cross tail current, though the maximum current density itself should be reduced. This could be an explanation for the result of *Nakai and Kamide* [1995] where the integrated current intensity should increase in the nearer earth region within about -13 Re, and in the more tailward region, it should decrease below the values at the time just prior to the onset. Hence, it might be concluded that the "current disruption" in the nearer earth region should mean a disruption or decrease of the current density, not of the integrated current intensity. As for the azimuthal pressure gradient in the dipolarized plasma sheet, *Shiokawa et al.* [1998b] showed that the gradient directs toward midnight both for the plasma thermal pressure and the total pressure as an average in the region of X=-10 to -15 Re. It should be noted that even in the quiet intervals, their data showed a similar magnitude of the thermal pressure gradient toward midnight of about 0.15-0.2 nPa/30 deg, hence, about 0.2-0.3 nPa/10Re at X~-13 Re. Similar order of the midnight-directed azimuthal pressure gradient in the near earth region was reported by *Hori et al.* [2000], and estimated by *Wing and Newell* [1998].

(j.3) Summary of the evolution after onset

In summary, the most dipolar configuration in the innermost region of *Baumjohann et al.* [1999] should occur around the time of the plasmoid ejection, and subsequent tailward propagation of the dipolarization and tailward retreat of the NENL occurs well after that. Considering the auroral bulge evolution mentioned in the section (j.1), the plasmoid ejection should start well after the fist stage of the evolution, and around the time between the second and third stages. In addition, considering the timing relation mentioned in the section (i.10), a clear enhancement of the reconnection rate at the NENL, which is manifested by the enhancement of the fast flows about 0-2 min after the onset, might start during the first stage of the bulge evolution. Hence it can be supposed that there might be some close relationship between the auroral bulge evolution and the reconnection process at the NENL, and a possible time sequence is as follows:

AB -> ABE1 -> ABE2 -> ABE3
BBF1 --------> BBF2 ------> PME
CFRN ----------> OFRN ------>

, where ABE3 and PME denote the third stage auroral bulge expansion and the plasmoid ejection, respectively, and CFRN and OFRN do the closed and open field line reconnections, respectively. Other abbreviations are the same as before. As discussed in the section (j.2.3), the tailward propagation of the initial information of the dipolarization (current disruption) should occurs mostly during the first stage of the bulge evolution, ABE1, whereas the tailward propagation of the full dipolarization proceeds in the later stage of the ABE3.

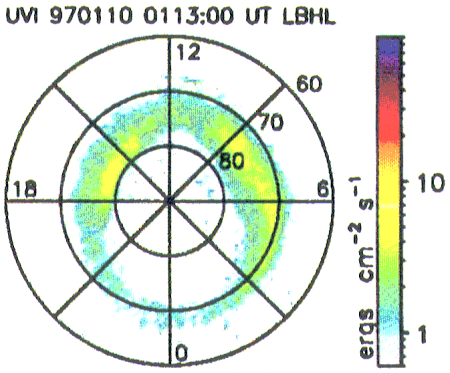
Auroral bulge should be an ionospheric projection of the whole dipolarization and current disruption region in the magnetosphere. The three stage evolution of the bulge suggests that there should be also three stages in the evolution of the substorm current wedge, possibly relating with the reconnection process at the NENL.

(k) A typical example of the global auroral substorm evolution

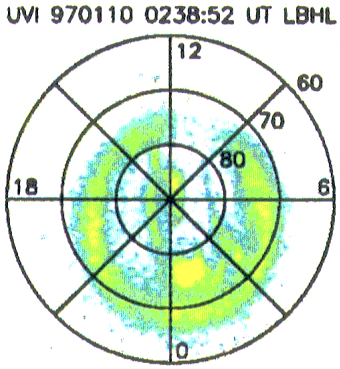
To summarize the morphology of the global auroral substorm that is currently understood, typical examples are shown in Figures 1.2 and 1.3. Both are observations by the POLAR UV auroral imager. Figure 1.2 is a reproduction of the Plate 2 in *Brittnacher et al.* [1999], and shows the evolution during the growth phase. After the southward turning of the IMF, the transpolar arcs develop in the polar cap region (Fig. 1.2(b)) and then they are swept out. The auroral oval gradually moves equatorward, its latitudinal width becomes narrower and its emission intensity gradually increases. Figure 1.3 shows the evolution after the expansion phase onset. The auroral breakup occurs in the pre-midnight localized region within the auroral oval (Fig. 1.3(c)). The auroral bulge rapidly expands in the both poleward and azimuthal directions (Fig. 1.3(d)). The azimuthal expansion is often asymmetric. In the subsequent stage, the poleward expansion is suddenly decelerated, and the azimuthal expansion proceeds in symmetric way (Fig. 1.3(e)). In the following stage, both the poleward and azimuthal expansions are accelerated again (Fig. 1.3(f)). Emission intensity in the middle of the bulge gradually decreases, and eventually the "double oval" configuration appears. Intense discrete auroral activities appear at the poleward-most region and the diffuse auroral region exists in the equatorward-most region (Fig. 1.3(h)). During the recovery phase, emission intensity of the both poleward-most and equatorward-most auroral activities gradually decreases, and the poleward-most activities gradually move equatorward (Fig. 1.3(i)-(j)).

Growth phase evolution
POLAR UVI data

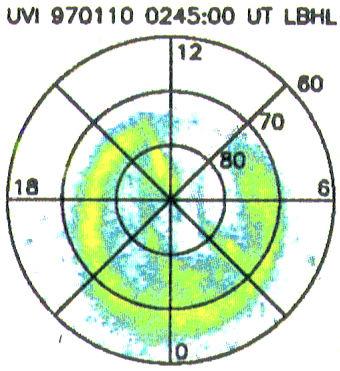
(a)



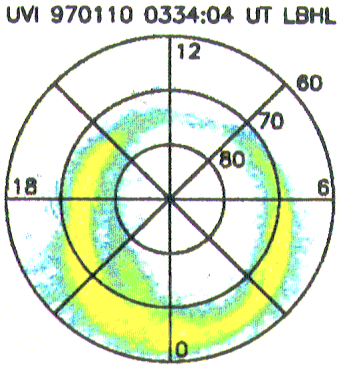
(b)



(c)



(d)



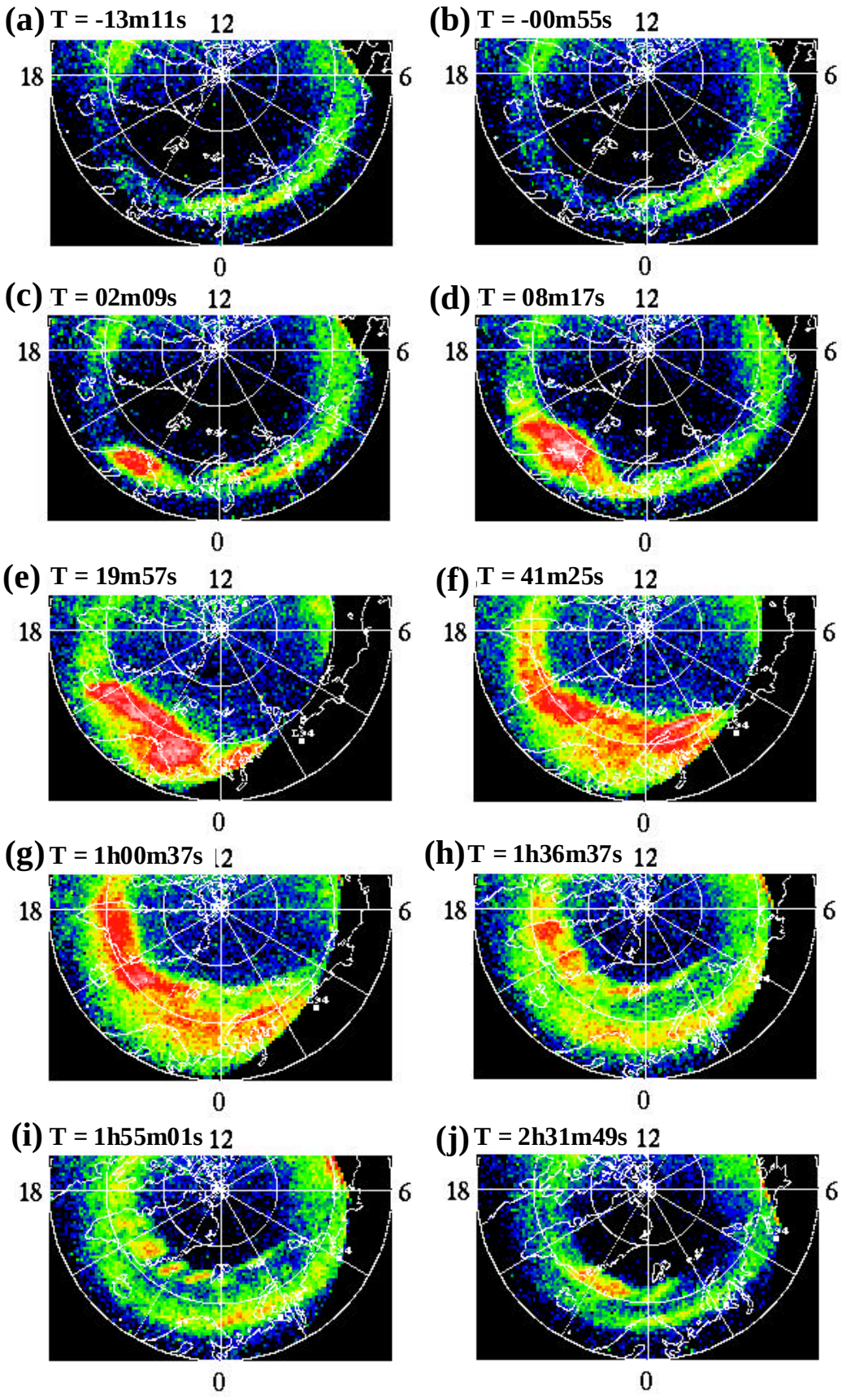


Figure 1.3. POLAR UVI data showing a typical global auroral evolution after the expansion phase onset of a substorm.

1.1.2 A summary of the still un-resolved problems

In the followings, we will summarize some of the still un-resolved important problems in the substorm study, based on the discussions in the section 1.1.1.

(1) Source mechanism of the onset arc, and its relationship with the onset mechanism
(2) Onset mechanism
(3) Premidnight preference of the onset region
(4) Localization of the onset region
(5) Explosiveness of the onset phenomena
(6) Causal relationship between the various onset signatures: NENL formation, CD, dipolarization, injection, Pi2, Pi1B, SCW, DP1 current system, and the auroral brightening
(7) Causal relationship between the phenomena during the expansion phase and the recovery phase: NENL activity, CD, dipolarization, injection, Pi2, Pi1, SCW, DP1 current system, and the auroral bulge evolution
(8) Mechanism to create the "double oval" configuration during the recovery phase.

Problems (1), (3), (4), (5) and (6) should be closely related with (2), and the understanding of the growth phase evolution should be essential for solving these problems. (8) is one prominent issue involved in (7). It should be important for understanding both the recovery phase of the substorm and the condition for the steady magnetospheric convection (SMC).

1.1.3 Existing substorm models

It seems a very tough work to survey all of the existing substorm models and describe each of them in detail because there have been so many models so far. In the followings, we will list out some representative ones.

(1) Classical NENL model
[e.g. *Hones*, 1977; 1979; 1984]
(2) Modified NENL model
[e.g. *Baker et al.*, 1996; 1999]
(3) Flow braking model based on the NENL model
[e.g. *Shiokawa et al.*, 1998a; *Birn et al.*, 1996; 1999]
(4) Current Disruption (CD) model
[e.g. *Lui*, 1996]
(5) Synthesis model based on the CD model
[e.g. *Lui*, 1991a; 1991b; 1998; *Lui and Murphree*, 1998]
(6) Field line resonance and Shear flow ballooning instability model
[e.g. *Samson et al.*, 1996; 1998]

(7) Ballooning instability model
[e.g. *Roux et al.*, 1991a; 1991b; *Korth et al.*, 1991; *Pu et al.*, 1992; 1997]
(8) Interchange, kink instability model
[e.g. *Pritchett and Coroniti*, 1997; *Pritchett et al.*, 1997]
(9) External trigger or the reduction of the magnetospheric electric field model
[e.g. *Lyons*, 1995]
(10) Magnetosphere-Ionosphere (M-I) coupling model
[e.g. *Kan and Sun*, 1985; 1996; *Kan et al.*, 1988; *Kan*, 1993; 1998]
(11) Boundary layer dynamics model
[e.g. *Rostoker and Eastman*, 1987; *Rostoker*, 1991; 1994; 1996]
(12) Thermal catastrophe model
[e.g. *Smith et al.*, 1986; *Goertz and Smith*, 1989]

Many models focus on the onset mechanism only that is actually the most challenging problem in the substorm study. However, it seems necessary now to construct a more comprehensive model to understand the whole process from the growth phase to the recovery phase, based on the rich observations in the both magnetosphere and ionosphere. The essential observational results to be explained in the model are as follows:

F1. Growth phase starts after the southward turning of the IMF.
F2. Plasma sheet thinning and taillike reconfiguration proceed during the growth phase.
F3. The NENL is formed frequently before the expansion phase onset.
F4. Plasmoid is released during the early stage of the expansion phase.
F5. Onset location is around 10 Re, while the location of the NENL is around 25 Re.
F6. Onset occurs around the transition region of the tail configuration from dipolelike to taillike.
F7. Onset occurs in the premidnight localized region explosively.
F8. Onset signatures (Pi2 wave, auroral brightening, SCW formation, dipolarization (CD), and injection) appear at onset.
F9. Westward traveling surge (WTS) develops on the western side of the bulge.
F10. Auroral bulge and SCW show eastward, westward and poleward (radial) expansions, often in a stepwise way.
F11. "Double oval" configuration appears during the late expansion and recovery phases.
F12. Plasma sheet recovers during the late expansion and recovery phases in the near and midtail regions.

In the followings, we will briefly discuss each of the models listed above.

(1) Classical NENL model

In this model, onset time of the expansion phase is the time when the reconnection of the plasma sheet field lines begins at the NENL, and the disruption of the cross-tail current occurs at

the location of the reconnection. Hence the onset region is the location of the NENL. These are inconsistent with the observations F3 and F5. This model includes the plasmoid release and subsequent recover of the plasma sheet. This is basically a model for the magnetospheric substorm, and considers very little about the ionospheric process. Especially relationship between the magnetospheric process and the auroral bulge evolution is not mentioned clearly. In addition, this model can picture the evolution of the magnetospheric process in the X-Z plane, whereas the azimuthal evolution is not depicted.

(2) Modified NENL model

This modified model considers the observations F3 and F5, and assumes that the expansion phase onset occurs when the reconnection of the lobe field lines begins. During the period when the plasma sheet field line reconnection proceeds, the cross-tail current is diverted in the plasma sheet from the NENL location to the inner and outer regions, and the inward shifted enhanced cross-tail currents cause the extreme thinning and taillike reconfiguration of the plasma sheet around the near earth onset region. A possible onset mechanism is the ion tearing mode instability due to the chaotic motion of the thermal electrons in the extreme thinned plasma sheet. This model assumes that the near earth onset occurs roughly concurrently with the start of the lobe field line reconnection. This kind of very speculative descriptions are very difficult to criticize. It is not clear whether the cross-tail current can be diverted from the NENL location around 25 Re to the near earth onset region around 10 Re, hence in the distance of about 15 Re, and whether such an azimuthal localization of the onset region about 1 Re can be possible by such a diversion. This model does not mention about the relationship between the near earth onset process and the reconnection process at the NENL after the onset and between the magnetospheric processes and the auroral bulge evolution, at least in their schematic drawings. This model seems to put less stress on the near earth onset process comparing with the NENL activity.

(3) Flow braking model based on the NENL model

This model elucidates the importance of the earthward fast flows which can transport a significant magnetic flux and plasma flux into the near earth region, and cause the dipolarization, injection, and generation of the SCW system in the near earth region apart from the generation region of the fast flows, possibly NENL location, on the tailward side. Owing to this model, the source region of the auroral brightening should be located around the flow braking point, and the poleward expansion of the bulge should correspond to the tailward shift of the braking point with the tailward velocity of the dipolarization, which seems too slow to explain the rapid expansion of the bulge, especially in the early stage, as discussed in the section 1.1.1(j). In addition, this model does not mention about the relationship between the magnetospheric process and the auroral bulge evolution clearly, especially in the azimuthal direction. Though the flux transport by the fast flows is undoubtedly an important ingredient of the magnetospheric substorm evolution, it might play its important role mainly in the late expansion phase and the recovery phase. In this model, the premidnight localization of the near earth onset region should depend on the localization of the

NENL activity on the tailward side. It is not explained consistently why such a premidnight localization of the NENL activity can be realized. This model seems to assume that the particles associated with the auroral brightening should be transported from the NENL location to the near earth onset region with the reconnected field lines. In other words, initially those particles should be situated around the NENL location. However, the model magnetic field line mapping indicates that those particles should be initially situated around the near earth onset region in the extremely stretched field line configuration just prior to the onset, not in the dipolar configuration after the onset. This is also an unclear point in this model.

(4) Current Disruption (CD) model

This model has been often referred as an antithesis to the NENL model, because the NENL formation is not considered to be an essential phenomenon to the expansion phase onset in this model. This model is basically a model for the onset mechanism. Now the NENL formation and the plasmoid release are established observational facts (F3 and F4), and it can be said that any models neglecting these facts are incomplete one. In this model, the main onset mechanism is considered to be the cross-field current instability (CCI) [e.g. *Lui et al.*, 1991] where the unmagnetized motion of ions significantly increases the cross-tail current density just prior to onset and causes some kind of kinetic instability to create the magnetic field turbulence observed in the near earth region at the onset. As discussed in the section 1.1.1(i.6), the current disruption features are observed almost always concurrently with the fast earthward flows, and such fast flows should be created by the reconnection process at the NENL. Hence the current disruption feature should be closely related with the NENL formation on the tailward side. Considering these points, it might be said that the current disruption feature itself cannot provide any antitheses to the NENL model, rather it is an important ingredient of the substorm evolution associated with the NENL activity.

(5) Synthesis model based on the CD model

This model speculates that a rarefaction wave is launched tailward as a result of the current disruption and causes the rapid earthward motion of the plasma and magnetic field (convection surge). This rarefaction wave causes further thinning in the near tail and midtail regions and eventually causes the NENL formation. After the NENL formation, substorm evolution is mainly governed by the NENL activity. Hence, after that time, this model is almost the same as the NENL model. One main difference between this model and the modified NENL model is in that this model explicitly associates the near earth onset mechanism with the NENL activity by assuming the rarefaction wave. With a typical speed of the fast mode wave about 1000 km/s, it will take just about 1.6 min for the distance of 15 Re from the estimated near earth onset location about 10 Re to the NENL location about 25 Re, which is enough shorter than the duration of the current disruption feature typically 5-10 min. Hence, it could be said that soon after the first current disruption onset, the substorm evolution is governed by the NENL activity, not by the subsequent spread of the current disruption region. This model assumes that the radial and azimuthal expansion of the SCW is caused by the spatial expansion of the current disruption region in a rather irregular and chaotic

fashion. For the CCI, sufficient increase of the cross-tail current density should be required. It is not clear whether such sufficient increase can be realized especially in the further tailward region where the magnetic field intensity becomes small and the current density should become small comparing at the initial onset location. The speculated irregular expansion seems very different from the rather systematic stepwise expansion of the auroral bulge as mentioned in the section 1.1.1(j.1). This model tries to include various mechanisms proposed by various models with almost no critical arguments for their feasibilities. It is a little doubtful whether such a kind of synthetic approach is a correct way to a final goal. As for the auroral activity, this model depicts its evolution corresponding to the magnetospheric processes, but in a very rough and ambiguous way.

(6) Field line resonance and Shear flow ballooning instability model

This is also basically a model for the onset mechanism. Observational evidence for this model seems too weak. There are no clear data showing that the auroral arc associated with the field line resonance is really the onset arc. Discussion about the auroral activity mainly depends on the data obtained with the meridian scanning photometers. For discussing the onset phenomena and onset timing, it is essentially important to know the location of the observer relative to very localized onset region. For that purpose, global auroral imager data are indispensable. In the schematic drawing by *Samson* [1998], the ballooning instability onset occurs during the explosive growth phase just prior to the onset, and the onset mechanism itself is a tearing mode or localized reconnection. After the onset, features due to the ballooning instability cannot be seen. In the drawing, it is not clear how the closed field line reconnection in the near earth region can proceed to the open field line reconnection on the tailward side. Azimuthal evolutions of the SCW and the auroral bulge are not mentioned at all.

(7) Ballooning instability model

This model is mainly based on the observations by the geosynchronous satellite GEOS 2, and basically a model for the onset mechanism. Only one event was described in detail relating with the auroral activity by *Roux et al.* [1991a; 1991b]. In the other papers, this group mainly concentrates to check whether the plasma parameters at the onset can satisfy the conditions for the growth of the ballooning instability. It seems very dangerous to derive any definite conclusion from the only one point observation at geosynchronous orbit. As for the event study by *Roux et al.* [1991a; 1991b], their interpretation (in their Figure 10) seems inconsistent with the observations (in their Figure 9). The observation showed that the energetic electron flux increases (decreases) in the dipolelike (taillike) configuration, whereas in the ballooning instability situation, the flux should decrease (increase) in a more dipolelike (taillike) configuration. Moreover, they discussed only about the source mechanism for the auroral surge form, and mentioned nothing about the discrete auroral activity extending from the surge in the east-west direction. It seems impossible to explain the prominent auroral activity by their ballooning instability model. In addition, they discussed only about the variation in the early stage of the dipolarization, and mentioned nothing about the afterward more essential trend to the full dipolarization.

(8) Interchange, kink instability model

This model is based on the numerical simulation works, and is a model for the onset mechanism. In the simulation by *Pritchett and Coroniti* [1997], the interchange mode produces an earthward plasma flow concentrating into the localized onset region. The north-south component of the magnetic field becomes negative in that region, and eventually that region moves tailward due to both the effects of a reconnection and a nonlinear growth of the interchange mode. Hence, in this model, the onset region is collocated with the reconnection region, which is inconsistent with the observation F5. They considered little about the observational facts, nothing about the auroral activity corresponding to the evolution of their instability.

(9) External trigger or the reduction of the magnetospheric electric field model

This model puts a strong stress on the fact that substorms are often triggered by the northward turnings of the IMF. This model assumes that the plasma pressure gradient should direct outward from the substorm current wedge region, which is inconsistent with observations [e.g. *Shiokawa et al.*, 1998b]. This model also assumes that the duskward electric field should decrease within the current wedge, which is also inconsistent with the observations showing an appearance of the intense duskward electric field responsible to the fast earthward flows. This model considers only the static electric field projected from outside, and neglects the inductive electric field generated by the internal process. This model cannot explain the eastward expansion of the current wedge. In this model, the dipolarization region propagates tailward with the speed of the discontinuity about 200 km/s, which is about one order of magnitude larger than the observed values as mentioned in the section 1.1.1(j.2.1). It is also unclear whether the azimuthal localization of the onset region can be realized by the mechanism mentioned in this model. In conclusion, this model is too speculative and is neglecting too many observational facts.

(10) Magnetosphere-Ionosphere (M-I) coupling model

In the original M-I coupling (MIC) model by *Kan and Sun* [1985] and *Kan et al.* [1988], the role of the magnetosphere is just to reflect the Alfvèn wave propagating from the ionosphere, and the magnetospheric source mechanism for the SCW system is not considered at all. Hence, the calculated results are very different from the observations, for example, as for the ionospheric potential pattern or current pattern. The calculated potential pattern is just a modification of the DP2 pattern, and does not indicate the characteristics of the DP1 pattern, e.g. as shown in the results by *Lu et al.* [1998]. In an extensive consideration about the substorm evolution including the magnetospheric process and the MIC process, *Kan* [1993] described the difference between the MIC model and the NENL model. In the MIC model, the expansion phase onset occurs when the upward field-aligned current (FAC) in the premidnight sector exceeds some threshold value about 1 $\mu A/m^2$, which is driven by the enhanced convection in the ionosphere. Closure of this upward FAC leads to disruption of the cross-tail current in the magnetosphere. This disruption induces the duskward electric field, which propagates down to the ionosphere, enhances the convection, and

drives further intense upward FAC. This process is called the unloading instability and proceeds as a positive feedback. In this model, the unloading energy is stored on the closed field lines in the near earth plasma sheet, and the NENL formation is not required. This model assumes that the auroral brightening region corresponds to the intense upward FAC region of the SCW system, and does not mention about the source mechanism of the entire bulge and its azimuthal expanding motion. In more advanced calculation by *Kan and Sun* [1996], they showed that an artificial intense localized convection is necessary to reproduce the signatures at the expansion phase onset. Hence, they admitted that the expansion phase onset does not occur due to the ionospheric process only as considering in the original MIC model. This intense localized convection pattern is very similar to the DP1 pattern. The purpose of the model should be to reproduce such a pattern, not to give the pattern a priori. This seems to show a limitation of the calculation of the MIC model. In more recent work, *Kan* [1998] speculated that the source mechanism of the DP1 pattern should be the braking of the fast earthward flows and the dipolarization-induced electric field. Hence, again they admitted that the MIC process is not important for the onset of the magnetospheric substorm. In summary, now the MIC model seems not the model for the substorm, just shows a coupling process between the ionosphere and magnetosphere. More realistic MIC process should be studied in a global simulation including both the ionospheric and magnetospheric processes.

(11) Boundary layer dynamics model

In the most recent paper by *Rostoker* [1996], this model seems to become just a model for the auroral surge formation after the onset. In this model, auroral surge is formed as a result of the Kelvin-Helmholtz instability which is generated by the enhanced velocity shear at the Low Latitude Boundary Layer (LLBL)/Central Plasma Sheet (CPS) interface due to the enhanced reconnection rate at the neutral line somewhere from about 30 to 100 Re. This enhanced reconnection is triggered by the rarefaction wave launched from the near earth onset region as a result of the sudden dipolarization. This story is very similar to the synthesis CD model. This model stresses the spatially periodic appearance of the auroral vortex that is clearly seen in a well-developed auroral bulge. The auroral surge itself does appear at just beginning of the bulge formation, and the source region for the bulge formation should be located in the near earth plasma sheet, not at the LLBL/CPS interface. Hence this model cannot be applied to the surge formation in the early stage of the expansion phase. This model does not mention about the auroral bulge formation and its evolution. Since evolution of the surge is closely related with that of the bulge, this model is very incomplete one even for the surge formation.

(12) Thermal catastrophe model

This is basically a model for the onset mechanism. A resonant absorption of the Alfvèn wave energy in the plasma sheet boundary layer (PSBL) causes a catastrophic increase of the temperature in the PSBL and adjacent central plasma sheet at the onset. The absorbed Alfvèn waves are considered to be initially excited by the K-H instability (KHI) at the magnetopause. For the normal Parker spiral form of the IMF, an effective excitation by the KHI should occur at duskside, and an

effective resonant absorption can be expected in a limited local time range in the pre-midnight sector. Hence, this model can give an explanation for the pre-midnight preference of the onset region. In this model, both the plasma sheet temperature and density gradually increase as the lobe magnetic field strength increases during the growth phase. *Baumjohann et al.* [1991] showed that the temperature of the central plasma sheet increases soon after the onset, while the temperature of the PSBL starts to increase well after the onset during the recovery phase, which does not seem to be explained by this model. This model cannot address the generation of the SCW system and its two- or three-dimensional evolution, since this is an one dimensional steady state model.

As discussed above, it can be said that even now there are no comprehensive models which can give consistent answers to all the problems listed in the section 1.1.2, especially concerning with the auroral activity. One reason for that may be in that the people based on the magnetospheric data tend to concentrate on and try to explain mainly the magnetospheric phenomena, and less consider the ionospheric phenomena including the auroral activity and consistency between both phenomena. On the other hand, the people based on the ionospheric data tend to make a, often unrealistic, speculation about the magnetospheric process without considering or knowing the magnetospheric phenomena strictly or exactly. In addition, the people based on the numerical calculations or their own theories often concentrate on one or several aspects in the substorm phenomena and almost neglect other aspects. To construct a more comprehensive model, such an approach as putting an equal-weighting on the both ionospheric and magnetospheric phenomena and trying to find a better consistency between them should be necessary. Now, the global auroral evolution can be monitored by the satellites (e.g. POLAR or IMAGE), various global observation networks have been constructed on the ground, and the magnetospheric phenomena can be observed with the sophisticated instruments aboard the spacecrafts (e.g. GEOTAIL or CLUSTER-2). In addition, the further realistic schemes for the global numerical simulation has been developed. Hence, now we seem to be standing at the point to proceed such an approach to get a complete understanding of the long standing fascinating mystery, substorm.

1.2. Purpose and contents of this study

Main purpose of this study is to analyze the auroral substorm evolution in detail, which was observed by the global auroral UV imager aboard the AKEBONO satellite and by the ground-based optical instruments at Syowa and Asuka Stations in Antarctica in 1989. The satellite UV imager can provide the global feature of the auroral substorm evolution, and the ground-based observation can reveal the detailed processes in the evolution. These global and local observations are complementary to each other, and both are considered to be equally essential for the auroral substorm study. Based on the detailed analysis and a great amount of the previous works by other researchers, as introduced in the above sections, we try to propose a comprehensive model for the substorm evolution in which we can find some consistency between the evolutions of the auroral

substorm and the "magnetospheric substorm".

The observations by the AKEBONO UV imager and the observations at Syowa and Asuka stations in 1989 are overviewed in Chapter 2. Detailed study of the one special event on June 6-7, 1989 is presented in Chapter 3. In Chapter 3, we present a comprehensive model for the substorm evolution from the growth phase to the expansion phase. A statistical study of the observations by the meridian scanning photometers at Syowa and Asuka stations is presented in Chapter 4. The summary and conclusions of this study are described in Chapter 5.