Auroral Imaging as a Tracer of Global Magnetospheric Dynamics

Karl Magnus Laundal

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

August, 2010

Preface

This synthesis and a collection of papers are submitted for the degree of philosophiae doctor (PhD) in physics at the Department of Physics and Technology, University of Bergen.

The thesis is divided into an introductory part and a part consisting of four papers submitted to international journals.

Paper I K. M. Laundal and N. Østgaard (2008). Persistent global proton aurora caused by high solar wind dynamic pressure *Journal of Geophysical Research* Vol. 113, doi:10.1029/2008JA013147

Paper II K. M. Laundal and N. Østgaard (2009). Asymmetric auroral intensities in the Earth's Northern and Southern hemispheres *Nature* Vol. 460, doi:10.1038/nature08154

Paper III K. M. Laundal, N. Østgaard, K. Snekvik and H. U. Frey (2010). Inter-hemispheric observations of emerging polar cap asymmetries, *Journal of Geophysical Research* Vol. 115, doi:10.1029/2009JA015160

Paper IV K. M. Laundal, N. Østgaard, H. U. Frey and J. Weygand (2010). Seasonal and IMF dependent polar cap contraction during substorm expansion phase, *Journal of Geophysical Research* in press, doi:10.1029/2010JA015910

This work is funded by the Norwegian Research Council project *IPY-ICESTAR* with project number 176045/S30.

Contents

Acknowledgements

This thesis is the result of five years of working with auroral images, first as a master student, and the last three years as a PhD student. Professor Nikolai Østgaard has been my supervisor since the beginning. I could not have asked for a more encouraging, inspiring, and ambitious supervisor. He has provided me with the basic ideas which got me started with my studies, and he has been enthusiastic about discussing the results, sometimes keeping us occupied for several hours, at his office, and over coffee, lunch or beer.

My five years at the space science group at the University of Bergen have been very enjoyable, much thanks to my fellow students, and the professors and engineers who work there. I would especially like to thank my officemate Thomas Gjesteland, and past and present PhD students and residents of Room 423. I am very happy to have collaborated with Kristian Snekvik, who has always been eager to discuss new ideas and problems. He is also a co-author on one of the papers, and he proof-read this thesis.

Earlier this year, I spent three months at the High Altitude Observatory at the National Center for Atmospheric Research, in Boulder, Colorado. Much of this thesis was written there, between gathering data for, and running the AMIE model. I would like to thank Gang Lu for making my stay in Boulder interesting and productive.

This work would not have been possible if not for the people behind the excellent data sets which we have used. The FUV instrument, and much of the software that was used for data analysis, was made by the FUV team at Space Sciences Laboratory at UC Berkeley. I would particularly like to thank the P.I., Stephen Mende, and Harald Frey, who has also been a co-author on two of the papers in this thesis. Data from the VIS Earth camera was provided by P.I. John Sigwarth.

I am also very grateful to my friends and family for caring about me, and how my studies progress, and for trying (and bravely pretending) to care about the magnetosphere. Finally, I would like to thank Anne Marie for supporting me despite my many late nights at the office, not to mention being left alone (and morning sick) for three months, while I was in Boulder.

Søgne, August 2010 Karl M. Laundal

Chapter 1

Introduction

The aurora is light which is produced when charged particles precipitate from space and collide with the thin gas in the Earth's upper atmosphere, typically at $100-300$ km altitude. Depending on the type of atoms/molecules with which the particles collide, photons with different wavelengths (colors) are produced. A continuous spectrum is also emitted because of the deceleration of the charges. From the ground, we see the aurora as thin discrete arcs, often aligned with circles of magnetic latitude. These arcs are produced by precipitation of electrons, accelerated relatively close to Earth ($\sim$ 1 Earth radius) by quasi-static upward electric fields, and electromagnetic wave activity. On a global scale, the aurora resembles ovals, with centers close to magnetic poles in the northern and southern hemispheres. In addition to the discrete arcs, the auroral ovals are comprised by a background of diffuse aurora. Both the arcs and the diffuse aurora can be seen with appropriate instruments, such as the cameras used in this thesis.

The observational basis of this study is global UV images of the aurora, taken by instruments on the NASA satellites IMAGE and Polar. To some extent, these images can be seen as projections of a much larger region further out in space, to which the Earth is magnetically connected. This region, called the magnetosphere, is constantly changing: Expanding, contracting, twisting, taking in and ejecting plasma, and opening and closing its connection to the interplanetary magnetic field (IMF). Global auroral images can help to identify and quantify these processes.

In Paper I in this thesis, we present observations which demonstrate a very direct relationship between the global aurora and the degree of compression of the magnetosphere. Using images from the SI-12 camera on the IMAGE satellite, which is sensitive to auroral emissions produced solely by proton precipitation, we observe a prompt increase, and persistently elevated intensity when the magnetosphere is compressed. When it expands, the proton aurora immediately diminishes. The compression and expansion of the magnetosphere was deduced from observed changes in the pressure in the solar wind.

In Papers II and III, we study in great detail an event when we had global UV images of the aurora in both hemispheres simultaneously. The unique data from this and a few other events have previously been investigated by *Østgaard et al.* (2004, 2005c), who focused on the location of corresponding auroral features in the two hemispheres (i.e. aurora which

is produced by charged particles moving along the same magnetic field line). They found evidence for a temporally twisted magnetosphere, enforced by the solar wind and IMF. In Paper II, we present an example when the technique employed by Østgaard et al. is not applicable; corresponding large-scale features in the two hemispheres could not be identified. The inter-hemispheric asymmetries in the auroral intensity presented in this paper are suggested to signify currents flowing along magnetic field lines between the two hemispheres. In Paper III we look at the same event, but here we focus on the interior boundary of the auroral ovals, i.e. the poleward boundary of the particle precipitation. The location of this boundary is often assumed to coincide with the boundary between magnetic field lines which are connected to both hemispheres (closed), and magnetic field lines which are connected to the IMF (open). We show that as magnetic flux is exchanged between these regions, the location of the boundary can become different in the two hemispheres, possibly because of differences between the upper atmosphere in the two hemispheres. In Paper IV, we use $\sim 30,000$ auroral images from one hemisphere to study the contraction of the auroral oval statistically. The main result from this study is that, as the oval contracts, average differences appear for different seasonal conditions, as well as for different orientations of the IMF.

This thesis starts with a survey of some fundamental concepts in space plasma physics, followed by a description of the geospace system and the aurora. We then summarize some of the previous work on inter-hemispheric asymmetries. A description of the instruments, and some techniques that we have used in the papers are then presented. The last chapter contains brief summaries of the papers. The main part of this thesis is four scientific papers:

Paper I:

Persistent global proton aurora caused by high solar wind dynamic pressure

Paper II:

Asymmetric auroral intensities in the Earth's Northern and Southern hemispheres

Paper III:

Inter-hemispheric observations of emerging polar cap asymmetries

Paper IV:

Seasonal and IMF dependent nightside polar cap contraction during substorm expansion phase

Chapter 2

Basic concepts in space plasma physics

Space physics is still a relatively young field. Before the launch of the first satellites[1], all knowledge about space was inferred from ground based observations. For example, the existence of a conducting layer in the upper atmosphere, the ionosphere, was postulated around 1900, as the means by which radio waves could be transmitted across the Atlantic ocean. To some extent, the field is still in a descriptive state. This is also reflected by the papers in this thesis, which can be considered primarily as reports of new observational findings. However, we also try to place these findings in a scientific context. The aim of the first part of this thesis is to give a brief review of the current understanding of this scientific context, in a more general way than what is permitted by the format of the scientific papers.

In this chapter we give a brief introduction to space plasma physics. We choose an approach which is inspired by recent works by *Parker* (1996, 1997, 2000, 2007) and *Vasyliunas* (2001, 2005a,b), who present this topic from a somewhat untraditional angle, focusing on causality. This is done by treating the plasma as a fluid, and explaining the dynamics of the fluid in terms of first principles in physics: Newton's laws of motion, and Maxwell's equations. The crude introduction which is presented here is meant to clarify the view which is adopted in the rest of this thesis. For the purpose of brevity, we shall refer to the above citations, and the books by *Baumjohann and Treumann* (1997), *Paschmann et al.* (2003), and *Griffiths* (1981) for mathematical derivations, and elaborations on many of the details which could not be included here.

2.1 Charged particles and electromagnetic fields

The universe is filled with plasma, a gas consisting of electrically charged particles. Only a very few exceptions to this statement exist, and one of them is the cold surface of the Earth, and a surrounding thin layer ($\sim 1/60$ of an Earth radius thick) of gas, called the

[1]The first satellite, Sputnik, was launched in 1957.

atmosphere. The universe is also filled with a magnetic field, and to this there is no known exceptions (*Parker* (2007) p. 14). The physics of space can therefore be said to be that of charged particles moving in magnetic, and sometimes electric fields. The most head-on approach to treat such a system mathematically, is to consider Newton's law of motion for each particle,

$$\frac{d\mathbf{p}_k}{dt} = q_k(\mathbf{E} + \mathbf{u}_k \times \mathbf{B}) \tag{2.1}$$

where $\mathbf{p}_k$ is the momentum of particle number k, q_k is its charge, $\mathbf{u}_k$ is its velocity, and $\mathbf{E}$ and $\mathbf{B}$ are the electric and magnetic fields. As is clear from this equation, the particles move in response to electric and magnetic fields. However, collectively they also change these fields, which behave according to Maxwell's equations,

$$\nabla \cdot \mathbf{E} = \rho/\epsilon_0 \tag{2.2}$$

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E} \tag{2.3}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{2.4}$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \epsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t} \tag{2.5}$$

where ρ is the charge density and $\mathbf{j}$ is the current density, which depend on the location and velocity of every single charged particle.

This kind of approach belongs to kinetic plasma theory. The enormous number of equations, resulting from the enormous number of particles in the magnetosphere, makes it clear that this approach is of little aid in getting a qualitative basic understanding of global phenomena, at least without the aid of massive computer power. Luckily, there is a simpler approach, in which the plasma can be treated as a fluid.

2.2 Fluid description of a plasma

The key transition in going from a kinetic description to a fluid description of a plasma (or any other substance that can be treated as a continuum), is the introduction of macroscopic quantities. Instead of keeping account of the position of every individual particle, we introduce the density, $n = N/V$, where N is the number of particles in a volume V. Instead of the velocities of each individual particle, $\mathbf{u}_k$, we use the mean velocity, or bulk velocity,

$$\mathbf{v} = \frac{1}{N} \sum_k \mathbf{u}_k \tag{2.6}$$

$\mathbf{u}_k$ can be decomposed as $\mathbf{u}_k = \mathbf{v} + \mathbf{w}_k$, where $\mathbf{w}_k$ is called the thermal velocity of particle number k. By definition of $\mathbf{v}$, the average of $\mathbf{w}_k$ is $\mathbf{0}$. The thermal velocities enter in the pressure tensor:

$$p_{ij} = \frac{1}{V} \sum_k m_k w_{ki} w_{kj} \tag{2.7}$$

where i, j represent the components x, y, z. The macroscopic quantity temperature also depends on the thermal velocities.

We want to be able to use these quantities in differential equations which describe the large-scale dynamics of the system. For this to be possible, two conditions must be fulfilled: 1) The volumes which are considered, $V = \lambda^3$ must be sufficiently large that they contain enough particles that the macroscopic quantities (n, $\mathbf{v}$, pressure, etc.) are not subject to statistical fluctuations. 2) λ must be small compared to the characteristic scale length of variations in the macroscopic quantities (Λ). This is necessary for a differential treatment to make sense.

These two requirements must be balanced according to the need for statistical precision and smallness of λ. In most cases, it is adequate that $\lambda = 10^{-3}\Lambda$ (*Parker* (2007), p. 75). In the case of the solar wind ($n \sim 5$ cm^{-3}) interaction with the magnetosphere ($\Lambda \sim 10^4$ km), we get $\lambda = 10$ km, and $N = 5 \times 10^{18}$. The statistical uncertainty in each cell, $\sim N^{-1/2}$, will be less than 10^{-9}.

The above quantities may be used to describe the plasma as composed by several fluids, e.g. an electron fluid and an ion fluid, in which case two sets of macroscopic variables must be considered, one for each fluid. Another approach is to treat the electrons and ions of a plasma as a single fluid. Multi-fluid theory may be appropriate when the forces acting on the electrons and ions differ, e.g. in the ionosphere, where the friction with neutral particles depends on the particle species. In most of the magnetosphere, a single-fluid approach can be used.

2.2.1 Conservation of mass and momentum

We now look at the differential equations which are used to describe the dynamics of the macroscopic quantities introduced in the previous section. For a derivation of the equations, see e.g. Chapter 7 in *Baumjohann and Treumann* (1997), or Chapter 8 in *Parker* (2007). These equations are conservation laws, and the first law states that the mass of the fluid is conserved:

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial x_j} n v_j = 0 \tag{2.8}$$

The equation is written on component form, and the subscripts j ($= x, y, z$) are to be summed over. This equation has four unknowns: n, and the three components of the velocity. Newton's second law states that the momentum is conserved. For a fluid, this law takes the form

$$nm\left(\frac{\partial v_i}{\partial t} + v_j \frac{\partial v_i}{\partial x_j}\right) = -\frac{\partial p_{ij}}{\partial x_j} + f_i. \tag{2.9}$$

Again, j is a summation index. m is the mass of the particles in the fluid. For an ion/electron fluid, the mass can be written, $m = m_e + m_i = m_i(1 + m_e/m_i) \approx m_i$, since the lightest ion, H^+, has a mass ≈ 1800 times larger than the electron mass. The pressure tensor describes the momentum flux density transported by the thermal motions of the particles. In the case that the pressure is isotropic, $p_{ij} = p\delta_{ij}$, the pressure term can be

understood as a force which accelerates the plasma in the direction anti-parallel to its gradient. f_i is the external force per unit volume (in the i'th direction). For a space plasma, the most important force is the electromagnetic force, which is the subject of the next section. In the following, we neglect gravity, which typically is small compared to other forces.

Assuming the force f_i to be known, Eq. 2.8 and 2.9 constitute four equations, with four unknowns in addition to the unknown elements of the pressure tensor (note that definition 2.7 implies that $p_{ij} = p_{ji}$). To remedy this, one can introduce yet another conservation law, conservation of energy density. This does however introduce another unknown quantity, the heat flux density. In principle, one can continue introducing new equations and new unknowns ad infinitum. Usually, one therefore truncates the set of equations, for example by making an assumption on the pressure using an equation of state.

For the present purpose, the key aspects of the above discussion is that 1) a plasma can be described as a fluid if its density is well-defined on a scale on which the macroscopic variables do not change much, and 2) that the dynamics of the fluid is governed by the pressure, and forces which are applied on it. This force is the topic of the next section.

2.3 The electromagnetic force on a plasma fluid

We have already mentioned that the collective effect of the charged particles in a plasma can change ambient electric and magnetic fields. We present an argument that this leads to the cancellation of any large-scale electric fields, and that this implies that the magnetic field is transported bodily with the plasma.

2.3.1 Maxwell's stress tensor

The electromagnetic force (Lorentz force) per unit volume is

$$\mathbf{f} = \rho\mathbf{E} + \mathbf{j} \times \mathbf{B} \tag{2.10}$$

where ρ is charge density, and $\mathbf{j}$ is current density. Using Maxwell's equations (2.2 and 2.5) to replace ρ and $\mathbf{j}$, and some vector calculus, the force per unit volume can be written (see *Griffiths* (1981), p. 351 for full derivation):

$$\mathbf{f} = \nabla \cdot \mathbf{T} - \epsilon_0\mu_0 \frac{\partial \mathbf{S}}{\partial t} \tag{2.11}$$

where $\mathbf{S}$ is the Poynting vector, $\mathbf{S} = (\mathbf{E} \times \mathbf{B})/\mu_0$. The Poynting vector term in Equation 2.11 contains the rate of change of the electromagnetic energy entering the volume per unit area per time. $\mathbf{T}$ is the Maxwell stress tensor. Element ij can be written

$$T_{ij} = \epsilon_0 \left(E_i E_j - \frac{1}{2}\delta_{ij} E^2 \right) + \frac{1}{\mu_0} \left(B_i B_j - \frac{1}{2}\delta_{ij} B^2 \right) \tag{2.12}$$

where $\delta_{ij} = 0$ when $i \neq j$ and 1 when $i = j$. The physical meaning of the force associated with element ij of the stress tensor is the force per unit area (or stress) in the i'th direction, exerted on a surface element normal to the j'th direction. The elements on the diagonal represent electric and magnetic pressure, and the off-diagonal elements are shears. The total electromagnetic force exerted on a volume element can be found by integrating Equation 2.11 over that volume.

2.3.2 The relative strengths of magnetic and electric stress

Equation 2.11 follows from Maxwell's equations, and is therefore exact. It is this force which is to be inserted in Equation 2.9. However, it can be simplified by considering the relative strengths of electric and magnetic stresses in space plasmas. The Poynting vector term will be treated later. A plasma is populated by highly mobile electrically charged particles (a potential difference of one volt will accelerate an electron to 600 km/s). Because of these charges, any electric field which arises, will be rapidly canceled. Therefore, there can be no significant large-scale electric fields in the local frame of reference moving with the plasma. The magnetic field will remain, since there are no equivalent magnetic charges or currents to cancel the magnetic field.

Assume for the moment, that the electric field, $\mathbf{E}'$, in the plasma frame of reference is exactly canceled. It is clear that, in this frame of reference, only the magnetic stress plays a role. Since the plasma velocity in general is far from uniform, this frame of reference is highly irregular, and so we want to calculate the forces in a different coordinate system. To do this, we have to use the Lorentz transformations for the electric and magnetic fields ($v^2/c^2 \ll 1$ is assumed). In a coordinate system in which the plasma is seen to move at velocity $\mathbf{v}$ (all quantities are assumed functions of $\mathbf{r}$ and t), the fields are given by:

$$\mathbf{E}' = \mathbf{E} + \mathbf{v} \times \mathbf{B} \tag{2.13}$$

$$\mathbf{B}' = \mathbf{B} - \frac{\mathbf{v} \times \mathbf{E}}{c^2} \tag{2.14}$$

Assuming $\mathbf{E}' = \mathbf{0}$, the electric field observed from our fixed coordinate system is $\mathbf{E} = -\mathbf{v} \times \mathbf{B}$. In general, the v/c^2 factor in Equation 2.14 does not mean that this term is small compared to $\mathbf{B}$, due to the relative magnitude of the two fields in SI units[2], which is used here. However, in this case, it is true, since by inserting $\mathbf{E} = -\mathbf{v} \times \mathbf{B}$ in the last term of Eq. 2.14 we get $\frac{v^2}{c^2}\mathbf{B}$. This means that $\mathbf{B}$ in conducting plasmas is invariant under non-relativistic coordinate transformations. We also note that the component of the electric field which is parallel to the magnetic field, $E'_{\parallel}$, is also invariant (but here we have assumed this component to be 0).

To compare the magnitude of the forces from $\mathbf{E}$ and $\mathbf{B}$, we see from Equation 2.12 that the electric and magnetic stress scale as $\frac{F_E}{F_B} \sim \frac{E^2}{B^2}\epsilon_0\mu_0 = \frac{E^2}{B^2c^2}$, where $c = 1/\sqrt{\epsilon_0\mu_0} \approx$

[2]This is what *Parker* (2007) refers to as a "grotesque" asymmetry of the SI unit system, whose description as a "practical" he compares to Eric the Red's naming of his icy discovery, Greenland.

3×10^8 m/s is the speed of light. In the case that $\mathbf{E}' = \mathbf{0}$, the ratio of electric to magnetic stress is $\frac{F_E}{F_B} \sim \frac{E^2}{c^2 B^2} = \frac{v^2}{c^2}$, when we use that $E = vB$. We can conclude that, when $\mathbf{E}' = \mathbf{0}$, the electromagnetic forces felt by the plasma are those of magnetic fields, and that electric fields play an insignificant role in the dynamics of a plasma. This is true in any frame of reference moving at non-relativistic speed compared to the plasma.

We also note that, since $\frac{F_E}{F_B} \sim \frac{E^2}{c^2 B^2}$, any $E' \neq 0$ will have to be of the order $c^2 B^2$ for the electric stress to directly affect the dynamics of the plasma fluid. For a magnetic field strength of 5 nT (e.g. the tail neutral sheet), the electric field would have to be 2.25 V/m, which can be regarded as a very strong field in the tail (electric fields in tail reconnection are in the order of $\sim 10^{-3}$ V/m (*Østgaard et al.*, 2005b)). For a magnetic field strength of 10,000 nT (e.g. the distant part of the auroral acceleration region), the electric field would have to be 9 MV/m. Electric fields of this magnitude have never been observed in this region of space (for comparison, the dielectric strength of air is $\sim$ 3MV/m).

2.3.3 Frozen-in magnetic field

The above considerations show that electric fields are negligible as a driving force of convection in space plasmas. However, we have not yet looked at how the plasma and the magnetic field are coupled. Including the magnetic stress in the momentum equation (Eq. 2.9) introduces the magnetic field as a new unknown, and we have to add three new equations to complete the system (assuming an equation of state can be used for the pressure). To do this, we rewrite Faraday's induction law (Eq. 2.3), using Equation 2.13:

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) - \nabla \times \mathbf{E}' \tag{2.15}$$

With $\mathbf{E}' = \mathbf{0}$, it can be shown (see e.g. *Parker* (2007), p. 93) that Equation 2.15 implies that the magnetic field moves bodily with the plasma. The bulk motion of the plasma does not cross magnetic field lines, and we say that the magnetic field is frozen-in. This is an important conceptual notion, and it is a good approximation almost everywhere in the universe. In the magnetosphere, we assume the magnetic field to be frozen-in everywhere except at shock fronts (e.g. sunward of the magnetopause), in magnetic reconnection (we will describe reconnection later), and in the acceleration region of auroral particles. In these regions, $E' \neq 0$, and $\nabla \times \mathbf{E}'$ can become comparable to $\nabla \times (\mathbf{v} \times \mathbf{B})$. This means that, while electric fields play no direct part in the forces acting on a plasma fluid in space, they can affect how the magnetic field evolves, and thus indirectly affect the plasma dynamics. We will discuss some of the regions in which this happens in Section 2.6.

With $\mathbf{E} = -\mathbf{v} \times \mathbf{B}$, the Poynting vector can be written

$$\mathbf{S} = \frac{1}{\mu_0} (-\mathbf{v} \times \mathbf{B}) \times \mathbf{B} = \frac{1}{\mu_0} \left(\mathbf{v} B^2 - \mathbf{B}(\mathbf{v} \cdot \mathbf{B}) \right) = \frac{B^2}{\mu_0} \mathbf{v}_\perp \tag{2.16}$$

which implies that, in the frame of reference of the plasma, the Poynting vector vanishes. This means that there is no transport of electromagnetic energy relative to the plasma.

The rate of change of $\mathbf{S}$ in a given volume therefore depends on $\mathbf{v}$. Since the Poynting vector term in Equation 2.11 is scaled by a factor of $\epsilon_0\mu_0 = c^{-2}$, this term is small compared to the Maxwell stress tensor as long as $v^2/c^2 \ll 1$.

2.4 Ideal MHD

The result of the above discussion is that the electromagnetic force entering in Eq. 2.9 is dominated by the divergence of the magnetic stress, and we get the following equation for conservation of momentum:

$$nm\left(\frac{\partial v_i}{\partial t} + v_j\frac{\partial v_i}{\partial x_j}\right) = -\frac{\partial p_{ij}}{\partial x_j} + \frac{1}{\mu_0}\frac{\partial}{\partial x_j}\left(B_iB_j - \frac{1}{2}\delta_{ij}B^2\right) \tag{2.17}$$

With the assumption that $\mathbf{E}' = \mathbf{0}$, Equation 2.15 reduces to (in component form)

$$\frac{\partial B_i}{\partial t} = \frac{\partial}{\partial x_j}(B_jv_i - B_iv_j) \tag{2.18}$$

Along with Equation 2.8 and an equation of state to handle the pressure, these equations constitute a full set, describing the large-scale dynamics of space plasmas. This description, which follows from $\mathbf{E} = -\mathbf{v}\times\mathbf{B}$, is called ideal magnetohydrodynamics (MHD). When $\mathbf{E}' \neq \mathbf{0}$, the main modification of these equations appears in Faraday's law, describing the evolution of the magnetic field.

2.4.1 Force balance between thermal and magnetic pressure – plasma β

We now look at an important special case of Equation 2.17, which leads to the definition of a useful plasma parameter, β. Assume that the left hand side of Equation 2.17 is zero ("hydrostatic" equilibrium), that we can neglect the off-diagonal terms in the particle and magnetic stress tensors, and that the diagonal terms are all equal, $p_{ii} = p$, and $T_{ii} = B^2/2\mu_0$ (isotropic thermal and magnetic pressure). Then Equation 2.17 reduces to

$$\nabla\left(p + \frac{B^2}{2\mu_0}\right) = 0 \tag{2.19}$$

which means that under these conditions, the total pressure is uniform. The relative importance of thermal and magnetic pressure defines the plasma β:

$$\beta = \frac{p2\mu_0}{B^2} \tag{2.20}$$

The solar wind is an example of a high β plasma, which is why the magnetic field is being pulled almost radially outward, instead of holding the plasma back. The magnetospheric lobes are examples of low β plasma, in which the magnetic field is in charge of the dynamics.

2.5 Contrasting paradigms

Currents and electric fields do not appear in Equations 2.8, 2.17 and 2.18. Having solved the above set of equations, the electric field and current density can be calculated from Ampere's law (2.5), and from Equation 2.13, respectively. It can be shown that **E** and **j** automatically will satisfy Maxwell's equations when the above procedure is followed (*Parker*, 1996).

For our purpose, the importance of the above discussion is that it sets the premises for how to qualitatively describe the large-scale dynamics of the system. Magnetic stress and plasma pressure should always be the primary suspects when large variations in the magnetosphere are observed. Currents and electric fields are derived quantities. In this paradigm, drawing electrical circuits in order to explain variations in the magnetic field by means of currents, would be to go at the problem in the wrong causal direction.

Despite the apparent simplicity of this approach, and the transparent relations to first principles of physics, it remains controversial. A common view, at least until very recently, is that the electric field drives convection. This view arises, in part, from Equation 2.13, which relates **v** and **E**. While this equation holds in ideal MHD, it does not say what is cause and what is effect. *Vasyliunas* (2001) showed more formally than what has been presented above, that **v** is the cause of **E**, and that an initial **E** is unable to maintain any significant flow of plasma. Later, he showed that in space plasma, **B** generally causes **j**, and not vice versa (*Vasyliunas*, 2005a,b). This view, which is called the **B**, **v** paradigm (in contrast to the **E**, **j** paradigm), has also been elaborated in papers (*Parker*, 1996, 1997, 2000) and more recently in a book by Eugene Parker (2007).

The success of the **E**, **j** paradigm in describing the average behavior of the magnetosphere can be attributed to the fact that the average behavior is approximately a steady state. **v** does not change, and the forces balance. Then, electric fields and currents can be described using circuit theory, and the observed steady velocity will follow. However, if the system changes, so does the current paths, and this approach becomes very difficult. Arguably, the above described approach, in which plasma reacts to forces according to Newton's law, is more intuitive in this case.

The importance of the distinction between **E**, **j** and **B**, **v** can also be said to be more fundamental than choosing the parameters which are most practicable. If **E** can not cause **v** in a space plasma, it becomes absurd to explain convection using electric fields, since it should be the other way around. This should also be considered when qualitatively discussing the cause for observations in near-Earth space.

2.5.1 Excitation of ionospheric convection

One example which illustrates the contrast between treating **E** and **j** as primary variables, and using **B** and **v**, is the excitation of ionospheric convection and currents (*Song et al.*, 2009). The high latitude ionospheric convection is often described as (or modeled as) an effect of an electric field, which maps down to the ionosphere along magnetic field lines, from the magnetosphere. Horizontal currents can then be calculated, using Ohm's

law, $\mathbf{j} = \mathbf{\Sigma} \cdot \mathbf{E}$, where $\mathbf{\Sigma}$ is the conductance tensor. Another approach is to start with field aligned currents, which map down from the magnetosphere, and use Ohm's law to calculate the electric field (and hence convection).

This view has been very successful in explaining and describing various observed phenomena. For instance, Ohm's law gives an intuitive explanation for why regions with high conductance, typically where auroral particles precipitate and in the sunlit part of the ionosphere, are associated with suppressed convection. The high conductance in these regions cancels some of the electric field, and thus reduces the convection. It also explains why these regions are associated with stronger horizontal currents.

However, according to *Vasyliunas* (2005b,a), this approach presupposes stable equilibrium. An alternative approach was adopted by *Song et al.* (2009), who used the equations of motion, and an imposed flow at the top boundary of the ionosphere as the driver of ionospheric convection. The imposed flow creates a force acting on the charged particles, which also experience a resisting force (friction), due to collisions with the neutral constituents in the ionosphere (only $\sim 1\%$ of the particles in the ionosphere are charged). The collision frequency is higher for ions than it is for electrons. Therefore, the friction term is also different, and a single fluid description is no longer appropriate. A coupled system of ≥ 3 equations of motion must be considered, one for each particle species: neutrals, electrons, and each type of ion which is present.

The energy dissipation associated with ionospheric flow is usually expressed as Joule heating, $\mathbf{E} \cdot \mathbf{j}$, in the $\mathbf{E}$, $\mathbf{j}$ paradigm. According to *Vasyliunas and Song* (2005), the energy dissipation is not really Joule heating, but frictional heating, arising from the collisions between the charged particles and the neutrals. According to these authors, it is "largely a coincidence" that the energy dissipation takes the form of $\mathbf{j} \cdot \mathbf{E}$ (with $\mathbf{E}$ in the frame of reference of the neutrals).

We return to a more detailed description of ionospheric convection patterns in Section 5.2.4, and in Section 5.3.2 we discuss how the ionosphere may act back on the magnetospheric driving, and even change the geometry of the magnetosphere.

2.6 Breakdown of ideal MHD

The electric field is not always zero in the frame of reference of the plasma. Localized regions exist where the electric field becomes important, and the frozen-in approximation breaks down. In this section, we look at when this happens. We focus in particular on one important example: Magnetic reconnection. Another example is the auroral acceleration regions, which is discussed in more detail in Chapter 4.

The degree to which ideal MHD holds can be quantified using the generalized Ohm's law (see *Baumjohann and Treumann* (1997) for a derivation of this equation):

$$\mathbf{E} + \mathbf{v} \times \mathbf{B} = \eta \mathbf{j} + \frac{1}{ne}(\mathbf{j} \times \mathbf{B} - \nabla \cdot \mathbf{P}_e) + \frac{m_e}{ne^2}\frac{\partial \mathbf{j}}{\partial t} \tag{2.21}$$

This equation is essentially the momentum equation of the electrons in the frame of reference of the single-fluid plasma. We recognize that it has the same form as Eq. 2.13,

with $\mathbf{E}'$ replaced by the terms on the right hand side. This expression can therefore be used to quantify the electric field, $\mathbf{E}'$, in the rest frame of the plasma. In ideal MHD, $\mathbf{E}' = \mathbf{0}$, and all the terms on the right hand side vanish.

The first term is a resistive term, which can be neglected in most of the magnetosphere (e.g. *Paschmann et al.* (2003), p. 50), but it can become important in the ionosphere. The next term is called the Hall term. It can be neglected when the characteristic scale size, $\Lambda \gg c/\omega_{pi}$, where c/ω_{pi} is the ion inertial length. With $n \sim 0.5$ cm^{-3}, the ion inertial length is ~ 300 km. Scale lengths of this size can be found in reconnection regions, and in the auroral magnetosphere (*Paschmann et al.* (2003), p. 50). In these regions, ions are demagnetized, but electrons are still frozen to the magnetic field. The last terms are the electron pressure tensor, and the electron inertial term. These are the most relevant terms contributing to parallel electric fields (*Paschmann et al.* (2003) p. 52).

We emphasize that the regions in which the terms on the right hand side of Equation 2.21 become significant occupy a very small fraction of the volume of the magnetosphere. In between these regions, where the fields vary smoothly, the MHD description applies, just as hydrodynamics applies to the regions between shock fronts (*Parker*, 1997). This means that MHD can not be used to address e.g., the acceleration of the charges in these regions, but it can predict the locations at which the acceleration occurs.

2.6.1 Reconnection

The concept of magnetic reconnection is illustrated in Figure 2.1 (insertion). The oppositely oriented magnetic field lines moving inwards (thick arrows) form a sharp gradient in the magnetic field, where the frozen-in approximation breaks down. These field lines then merge at certain points, forming what is known as the X-line, because of the X-shaped boundary between incident magnetic flux and plasma and the highly curved field lines which are ejected from the reconnection region (thin arrows). The interior white area denotes a region in which ions are de-magnetized (ion diffusion region), due to the Hall term in Eq. 2.21. In this region, electrons are still frozen to the magnetic field, and will be carried away from the X-line, producing a net transport of negative charge.

The figure also shows two regions in which reconnection can change the topology of the magnetosphere. On the dayside, the interplanetary magnetic field (dashed lines) can merge with the closed field lines of the Earth (solid) to form open field lines (dotted). The momentum of the solar wind then transports the open field lines anti-sunward, and stretch them into a tail, which can extend out to several hundred R_E (Earth radius) on the nightside of the Earth. When open field lines with footpoints in opposite hemispheres meet in the tail, they may reconnect, and form new closed field lines. The X-line in the tail can form both at very high altitudes, $\gtrsim 100 R_E$ (distant X-line), and closer to the Earth, at $\sim 20 R_E$, in what is called the near Earth neutral line (NENL). In the NENL, reconnection can occur between open field lines, or closed field lines which are stretched. Section 3.1 describes in more detail the consequences of dayside and nightside reconnection for the flow of mass and energy throughout the magnetosphere.

The magnetic field and plasma which is ejected from the X-line has a high velocity

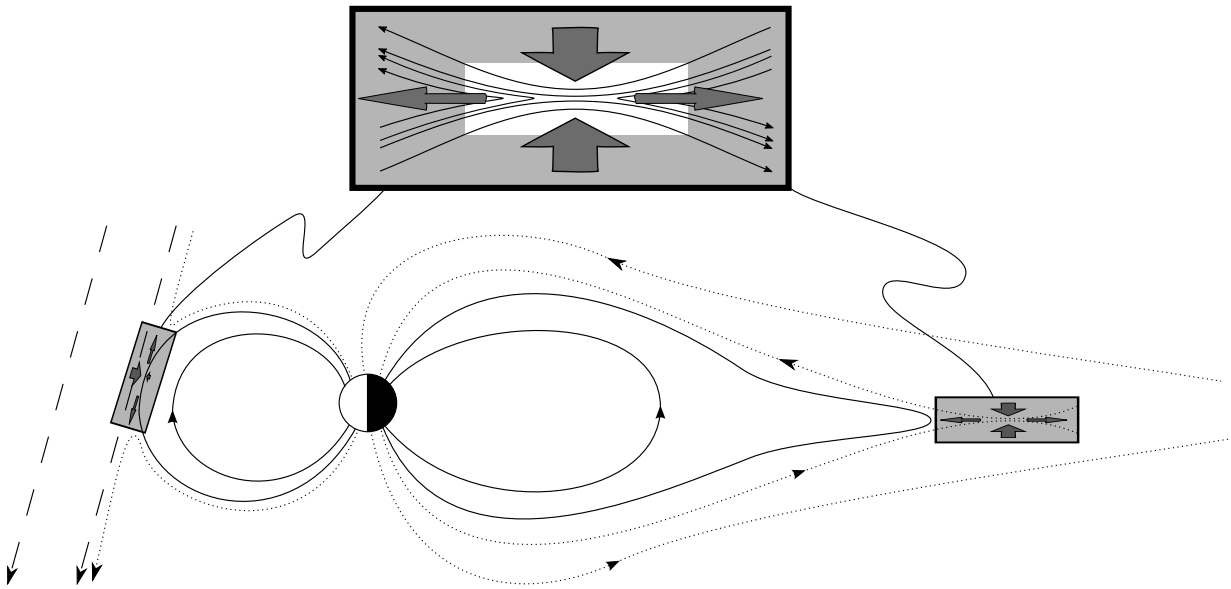

Figure 2.1: Magnetic reconnection shown at two regions in the magnetosphere. On the dayside (to the left in the figure), the interplanetary magnetic field (dashed) merges with closed field lines (solid) of the Earth, producing open field lines (dotted). The open field lines are transported anti-sunward, to the magnetotail. The opposite orientations of the open field lines in the tail enables open field lines to reconnect, forming new closed field lines. The reconnection geometry shown in the insertion illustrates slow (thick arrows) convection of plasma and magnetic field into the X-line (note the asymmetric inflow to the dayside X-line), in which the magnetic field is zero, and fast jets (long arrows) out from the X-line. The interior white region shows where the ions are de-magnetized, due to the sharp gradients in the magnetic field.

($\sim 0.9 v_A$, where v_A is the Alfven speed, $B/\sqrt{\mu_0 nm}$). In-situ measurements of localized regions of high speed plasma are therefore often interpreted as a crossing of these reconnection jets. Another signature of magnetic reconnection observed by spacecraft in the magnetosphere is magnetic perturbations associated with the currents which are implied by the ejection of electrons from the ion diffusion region (e.g., *Snekvik et al.* (2008)).

2.7 Summary

For the present purpose, the key points in the above discussion is that the solar wind-magnetosphere-ionosphere system can be treated largely as a fluid of plasma and magnetic field, which is subject to mechanical forces: Plasma pressure gradients and magnetic stress. $\mathbf{B}$ and $\mathbf{v}$ can be seen as the primary variables, while $\mathbf{E}$ and $\mathbf{j}$ can be derived once $\mathbf{B}$ and $\mathbf{v}$ are known. Of course, if $\mathbf{j}$ and $\mathbf{E}$ are known from measurements (which to some extent can

be true with auroral images), they can be used to make inferences on $\mathbf{v}$ and $\mathbf{B}$, but the physical process leading to these measurements should be described in terms of convection and magnetic field.

The localized regions in which fluid description, and ideal MHD in particular, is not applicable, are of great importance to acceleration of particles and exchange of magnetic flux with the interplanetary magnetic field. These processes facilitate changes in magnetic field geometry, and changes in plasma populations, which then indirectly affects global dynamics.

Chapter 3

Energy and mass flow in the geospace system

In this section we describe the basic constituents of geospace, and look at some important processes governing the flow of mass and energy throughout the system.

Figure 3.1 shows a map of the magnetosphere. Being in a state of perpetual change, this map is at best an average picture. However, the various regions are relatively well-defined, and we will here give a description of the principal processes by which the various regions interact. We also try to keep track of what are the driving processes, and what processes are modulating the driver (feedback). To do this, we take the approach argued in the previous section, that the principal parameters in space plasmas are $\mathbf{B}$ and $\mathbf{v}$.

3.1 The Dungey cycle

When the IMF has a component which is anti-parallel to the geomagnetic field, reconnection can occur between the closed magnetic field lines on the dayside magnetopause, and the IMF. This creates new open field lines which, due to the momentum of the high β solar wind, will be pulled anti-sunward. Since the open magnetic field lines are connected to the Earth, they will be deformed by the anti-sunward motion on higher altitudes, thus forming the magnetotail. In this process, some of the kinetic energy of the solar wind is being converted to magnetic energy, i.e. there is Poynting flux through the magnetopause. The resulting magnetic energy is stored in the lobes, which are regions of highly rarefied plasma ($n \sim 10^{-2}$ cm^{-3}) (*Baumjohann and Treumann* (1997), p. 7), and low β. The lobes are comprised by open magnetic field lines. Being oppositely oriented in the two hemispheres, the implied shear between the two lobes necessitates a cross-tail current, directed from dawn to dusk.

When the lobes are pushed together, reconnection may form new closed magnetic field lines. In this process, magnetic field is annihilated, and magnetic energy is transferred to the plasma, thus producing the high β plasma of the region called the plasma sheet. The newly closed field lines of the plasma sheet are stretched, and the plasma is therefore

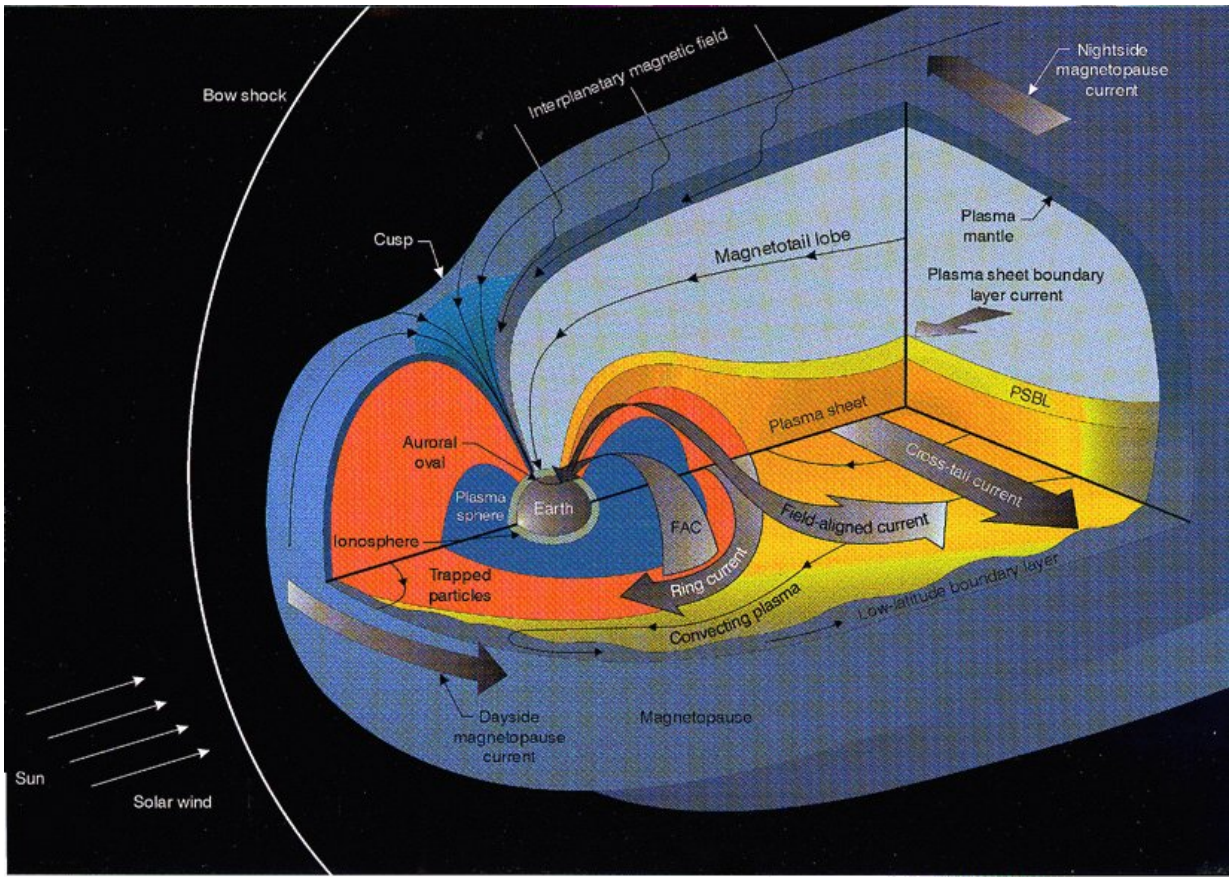

Figure 3.1: Figure showing currents and regions with different plasma populations in the magnetosphere. The *ionosphere*, which consists of a partially ionized, relatively cold and collisional plasma is shown as a thin layer surrounding the Earth. The *plasmasphere* (blue region) consists of co-rotating cold plasma which resides on closed field lines mapping to low latitudes. The *ring current*, whose bulk of energy is carried by energetic ($\sim 10 - 200$ keV) protons, circles the Earth outside (and sometimes interact with) the plasmasphere. The *plasma sheet* is a high-β plasma which maps approximately to the auroral zone. In the tail, it is separated from the low-β plasma of the *lobes* by the plasma sheet boundary layer (*PSBL*). The demarcation between the solar dominated plasma and the region which is dominated by magnetic field lines connected to the Earth, is called the *magnetopause*. The low latitude boundary layer (*LLBL*) and the plasma *mantle* can be seen as the spatial extension of this boundary. Outside the magnetopause is the *solar wind*. A somewhat simpler version of this figure appears in *Kivelson and Russell* (1995), p. 22; the present version is of unknown origin.

subject to magnetic stress (curvature forces), which pushes the plasma towards the Earth. Observations show that the earthward plasma flow takes the form of localized jets, called bursty bulk flows, rather than a large scale, slow convection (*Angelopoulos et al.*, 1994). As the plasma approaches the Earth, it encounters an increased total pressure, which will divert the flow towards the flanks of the magnetosphere. When reaching the dayside, the magnetic flux can once again merge with the IMF, and the above described process can be repeated.

This cycle is called the Dungey-cycle, after the scientist who first suggested its basic principles (*Dungey*, 1961). It is believed to encompass the principal processes responsible for convection of plasma and magnetic flux through the high latitude magnetosphere. The force responsible for the excitation of the convection is magnetic stress and plasma pressure. Reconnection acts simply as a gate opener, changing the topology of the system to allow the magnetic field to relax. The geometry of the reconnection is however of great importance for the geometry of the convection, and hence also for the geometry of the magnetosphere. This will be discussed further in Chapter 5. Several other processes also modulate the cycle, and we will discuss some of them shortly.

Opening of flux on the dayside and closing of flux on the nightside change the total open magnetic flux of the magnetosphere. The ionospheric footpoint of the open flux (the lobes) is called the polar cap. While reconnection only indirectly excites convection, it directly affects the location of the open/closed field line boundary (or polar cap boundary), and the size of the polar cap. Expansion (contraction) of the polar cap therefore signifies flux opening (closure) in excess of closure (opening). Flux closure and opening are quasi-independent processes, since they occur $\sim 30R_E$ (near Earth neutral line) to $> 100R_E$ (distant neutral line) apart. The idea of a quasi-independent expanding/contracting polar cap was suggested by *Siscoe and Huang* (1985), and expanded upon by e.g., *Lockwood et al.* (1990); *Cowley and Lockwood* (1992). This is discussed further in Paper III, where we use auroral images to identify the polar cap boundary. This method is also discussed in Section 4.1.1.

3.2 Different reconnection geometries

The Dungey cycle described above is observed when the IMF has a southward component. When it points northward, the IMF can merge with lobe magnetic field lines. The resulting convection is sunward in the ionospheric polar cap (*Crooker*, 1992). One can also imagine several other geometries in which different kinds of reconnection (and subsequent convection) are important (*Watanabe et al.*, 2007b). However, our main focus is on the Dungey type reconnection, and associated convection, since this cycle is generally believed to dominate the energy transfer from the solar wind to the magnetosphere, and the magnetospheric dynamics during the most active times. It is also worth noticing that, except for the relatively rare case of dual lobe reconnection (small IMF $|B_y|/B_z$ and $B_z > 0$) (*Imber et al.*, 2006), the Dungey-type reconnection is the only one capable of changing the open flux content in the magnetosphere.

3.3 Circulation in the inner magnetosphere

Not directly included in the above description of the Dungey cycle is the circulation in the inner magnetosphere. The plasmasphere is known to co-rotate, albeit not perfectly (*Burch et al.*, 2004), and co-rotation is also a characteristic motion of cold plasma at ring current altitudes. The driving force of the co-rotation comes from below: At low latitudes (to which the forces associated with the Dungey cycle do not usually penetrate) the upper atmosphere rotates with the Earth. Via frictional forces, the ionospheric plasma is brought into the same motion. Since this plasma is frozen to the magnetospheric plasma higher up on the same flux tubes, magnetic stress acts on the magnetospheric plasma until its motion matches the co-rotation.

During periods of particularly strong driving from the solar wind, hot plasma can be injected to the ring current from the plasma sheet, where it can reside for days. These periods are called geomagnetic storms. The energetic particles which carry the ring current undergo gradient and curvature drifts, crossing magnetic field lines. The bulk motion of the single-fluid plasma in the ring current region, on the other hand, does not cross magnetic field lines, since the frozen-in condition still holds there.

In this view, there is a region of the magnetosphere, approximately at ring current altitudes, where the domination of solar wind driving (Dungey cycle) and the more sluggish ionospheric driving (co-rotation) overlap.

3.4 A component view of geospace

In Figure 3.2 we have divided geospace into eight component regions. The arrows indicate how mass and/or energy flows between the various regions, and their labels indicate the process by which this happens. The driving processes in the Dungey cycle and thermospheric winds are both present in the chart. However, from the large number of other interactions, it is clear that other driving processes exist, and that the drivers can be heavily modified by numerous feedbacks. Another process which is thought to be important when the IMF is northward, and can not effectively reconnect with closed field lines on the dayside of the Earth, is viscous interaction (*Axford and Hines*, 1961). Viscous interaction arises from the shear between the solar wind and the magnetosphere, which can trigger instabilities (e.g., Kelvin-Helmholtz), and drive plasma on high latitude closed field lines anti-sunward.

Feedback mechanisms which affect the flow imposed by the driving Dungey cycle processes include i) ionospheric drag, ii) precipitation of particles increasing the ionospheric conductivity, iii) formation of plasmaspheric plumes which in turn affects wave particle interactions, which can increase the precipitation to the ionosphere, and iv) outflow of ionospheric ions (O^+) which can modulate the reconnection efficiency (*Shay and Swisdak*, 2004). These few examples are also interrelated, adding to the complexity. It is clear that the system is highly non-linear, and that a basic qualitative understanding of the global system can not include all eventualities. A common approach is that each observed phe-

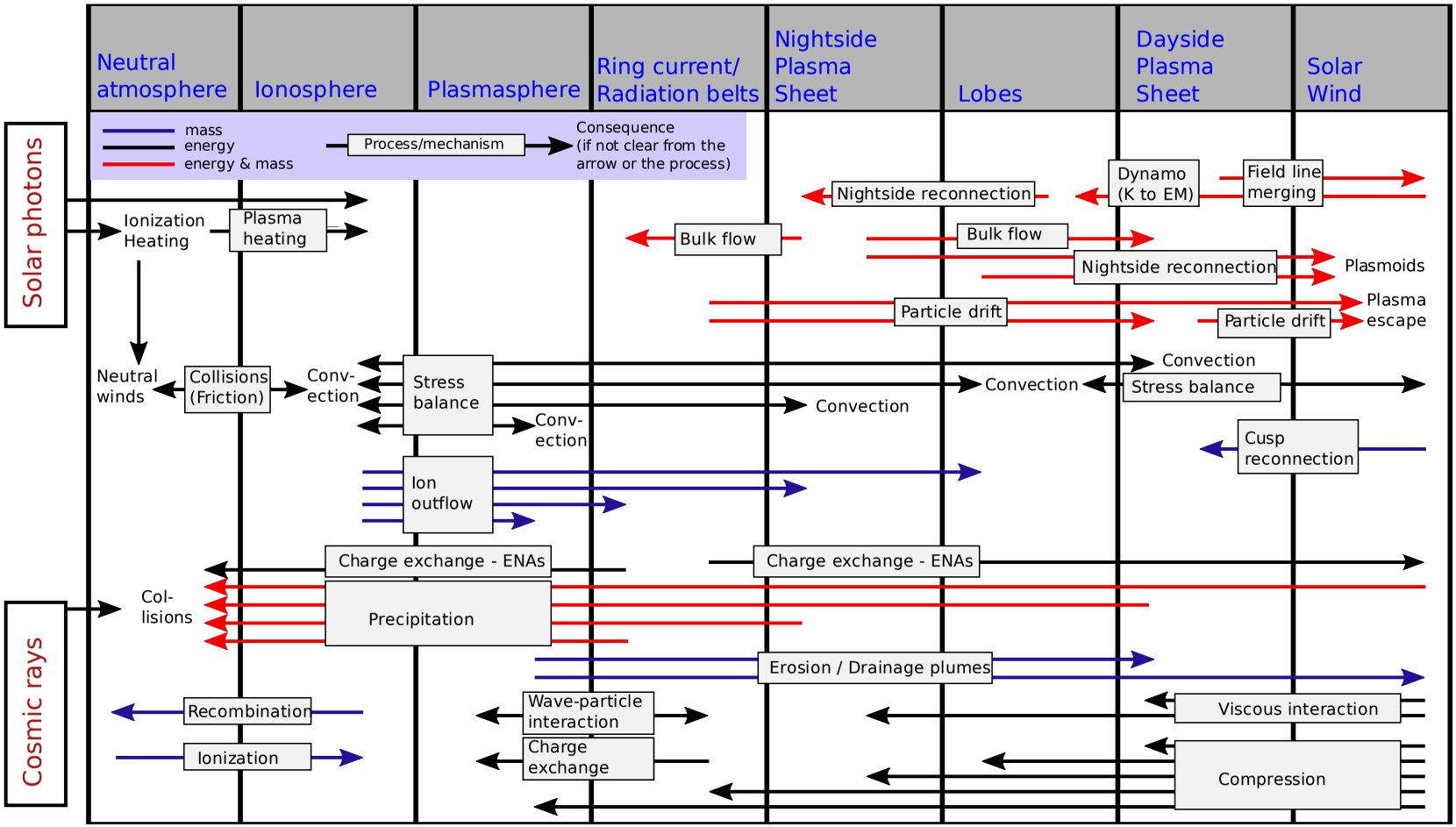

Figure 3.2: Flow chart, showing how mass and energy flow between various regions in geospace. The regions are shown as vertical columns. The flow of mass and energy is from the start to the end of the arrows; arrows crossing a region is not meant to represent flow through that region. The color of the arrow denotes flow of mass and/or energy (see key in upper left corner). The separation into mass and energy is not strictly consistent. One example is ENAs escaping to the solar wind, which represents a significant energy flux, but only a negligible loss of mass. Acronyms: K: Kinetic [energy], EM: Electromagnetic [energy], ENA: Energetic Neutral Atom. This figure was made in collaboration with the System Level Science group at the International Space Science Institute.

nomenon is interpreted in terms of maybe one of the processes in Figure 3.2, believed to be of particular importance.

3.5 The role of substorms

Being an important element in Papers II, III and IV, the role of substorms should also be discussed in the context of global geospace dynamics. Defining and explaining the formation of substorms have been controversial topics the last few decades. We will not attempt a survey of the controversy here, but rather focus on some descriptive aspects of substorms.

In global auroral images, substorms can be identified as a local brightening, followed by an expansion in latitude and longitude, lasting for a few 10s of minutes[1] (*Akasofu*, 1964; *Frey et al.*, 2004; *Gjerloev et al.*, 2008). One example, from 23 June 2000, is shown in Figure 3.3. In this event, the aurora was very faint prior to the onset. Then, at 13:05 UT, a local brightening is observed, and in the subsequent images a longitudinal and latitudinal expansion.

The intensification seen in Figure 3.3 was associated with a contraction of the oval (or equivalently, poleward expansion of the aurora). The contraction signifies a decrease in the open flux in the magnetosphere (see Section 4.1.1). Substorms can thus be seen as a process by which the magnetosphere releases energy, and rids itself with excess open flux through enhanced tail reconnection (*Milan et al.*, 2007). Substorms are relatively frequent, occurring at a cadence of ~ 3 hours during persistent strong solar wind driving (for reasons unknown). The probability of a substorm onset occurring has been shown to increase with the level of open flux (*Boakes et al.*, 2009), and the magnetic flux closure is more significant when the initial level of open flux is higher (*Milan et al.*, 2009a). It should be noted that flux closure is also believed to occur outside of substorm expansion phases, then most importantly by reconnection at the distant neutral line.

The expansion of the aurora from a localized region to a large fraction of the auroral zone reflects processes in the magnetotail. The onset maps approximately to the transition between dipolar field lines and the stretched field lines of the tail. In the subsequent minutes the tail becomes increasingly dipolar, the dipolarization spreading from the onset region to become a global phenomenon. The dipolarization region has been suggested to map to the region of most intense aurora, called the bulge (*Liou et al.*, 2002).

From the description of the Dungey cycle given above, in which reconnection acts as a gate opener for convection, substorm expansion phases are also expected be associated with an increase in convection. This is discussed in Chapter 5.2.4.

[1]The term substorm stems from another substorm signature: A negative perturbation in the horizontal magnetic field seen at ground magnetometers in the auroral zone in the northern hemisphere. These perturbations were long (erroneously) believed to be subgroups of the more long-lasting and global magnetic perturbations which signify geomagnetic storms. Another substorm signature is observations of dispersionless abrupt increase of energetic particles at geosynchronous orbit. However, we will stick to the definition from auroral images.

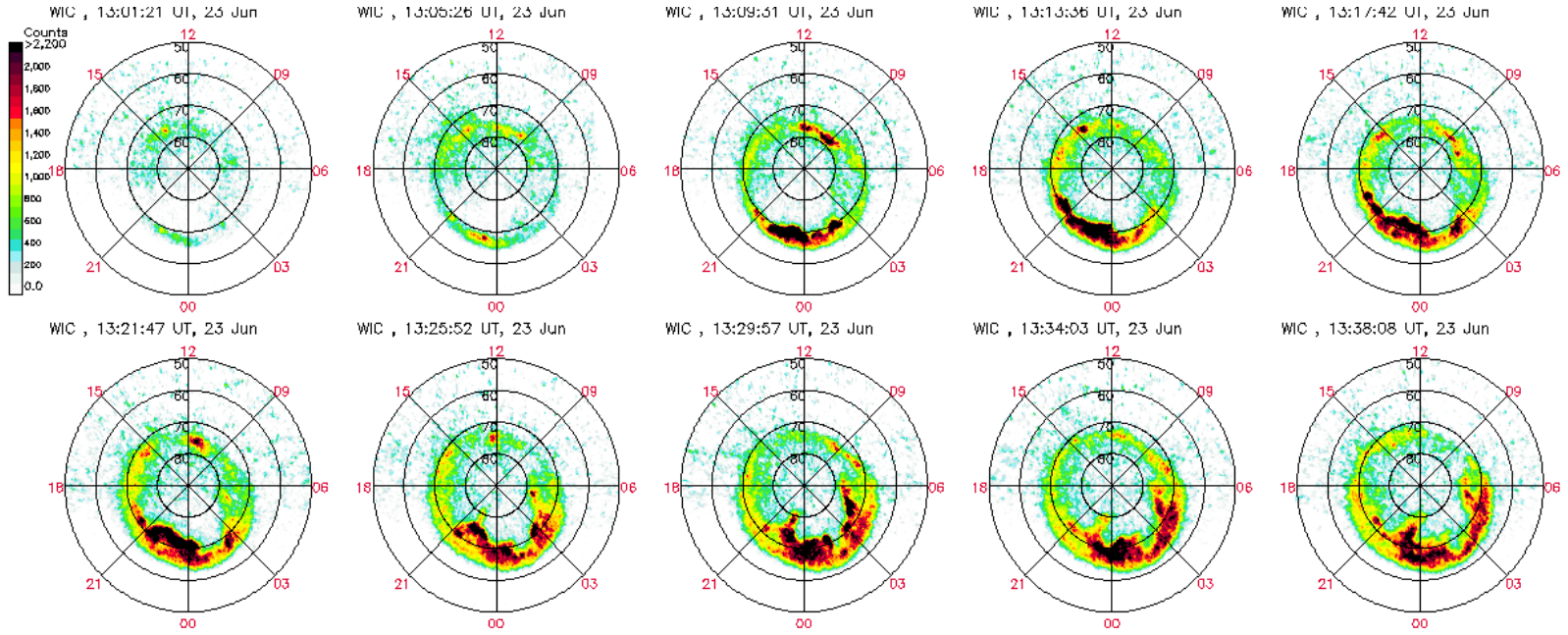

Figure 3.3: Sequence of images, covering ~ 40 minutes, and showing the onset and expansion of an auroral substorm, on 23 June 2000. The onset is seen at 13:05 UT, at 23 MLT. In the next image, four minutes later, the region of bright aurora has expanded to encompass almost six hours MLT. The brightening on the dayside was most likely due to a concurrent increase in solar wind pressure. The images were taken by the WIC camera on the IMAGE satellite.

Chapter 4

The aurora

In this chapter, we look at how auroral imaging can be used as a tracer of magnetospheric dynamics, and in particular how it can be used to estimate the open flux content of the magnetosphere. Then we describe different types of aurora.

It is important to keep in mind that the aurora is not merely an illumination of magnetospheric processes, but also a signature of processes taking place, which modulate the global magnetospheric dynamics. Some examples: 1) Particles producing the aurora also increase the ionospheric conductivity, changing the interaction between the ionosphere and the magnetosphere, and therefore changing the convection in both the magnetosphere and the ionosphere (Chapter 5.3.2). 2) Associated with auroral precipitation is often an upflow of ionospheric ions, which increases the mass content in the magnetosphere, thereby changing the global dynamics. For example, outflow of O^+ ions is believed to affect magnetospheric dynamics in the main phase of geomagnetic storms. 3) The aurora often signifies the existence of parallel electric fields, which are important for particle acceleration. Parallel electric fields also imply that the frozen-in condition no longer holds.

4.1 Mapping to the magnetosphere

For auroral imaging to be used at its full potential, we must know how the precipitating particles relate to source regions and processes in the magnetosphere. Many studies have used auroral features as signs of localized magnetospheric activity. Examples include north-south aligned arcs (streamers) which are interpreted as an ionospheric signature of bursty bulk flows (*Sergeev et al.*, 1999), the cusp spot as a tracer for high-latitude magnetopause reconnection (*Frey et al.*, 2003b), and sudden brightenings signifying the onset of magnetospheric substorms. Several examples of various auroral forms (outside the main oval), and their proposed mapping to magnetospheric processes are found in *Frey* (2007).

4.1.1 Open and closed magnetic field lines

Identifying the magnetospheric regions from which auroral particles precipitate can be done by studying their energy spectra (e.g., *Newell et al.* (2004)). When using UV images, however, this becomes more difficult, since the spectra are generally not known.

However, one region which can often be recognized in UV images of the aurora is the polar cap. This region, which is defined as the ionospheric footpoint of the open magnetic field lines, is largely void of energetic particle precipitation. The low energy precipitation ($\sim$ eV) which can be found there, called polar rain, causes too faint emissions to be observed from global imagers. Just equatorward of this boundary, the precipitation is harder and more intense, causing detectable emissions, which can be used to identify the boundary.

It is however important to be aware that the method for identifying the open/closed boundary also has its problems. First of all, some visible precipitation occurs on open field lines, most notably in the cusp. The cusp is comprised by newly opened field lines, and it borders to precipitation on closed field lines, so that the two regions often are indistinguishable in global UV images. With in-situ particle observations, the cusp can be identified by its lower energies, compared to particles on closed field lines (e.g., *Newell et al.* (2004)).

The perhaps most problematic aspect of using global images to determine the OCB is that the aurora is often faint (e.g., the image prior to substorm onset in Figure 3.3). This might either make it impossible to assess a boundary, or it may produce a wrong boundary. The latter may happen if the precipitation from the center plasma sheet (close to Earth) is intense, while the precipitation from the distant plasma sheet is weak. One example is substorm onset, which maps to a region quite close to Earth, and may be embedded in an otherwise sub-visual auroral oval. Low-altitude in situ particle measurements are much more sensitive than global imagers, and therefore less susceptible to this problem. Comparisons between the open/closed boundaries inferred from these measurements (particle precipitation measured by DMSP in this case), with the boundaries from global UV images, have given credence to the imaging method; *Carbary et al.* (2003) and *Boakes et al.* (2008) found good agreement between the boundaries determined from imaging and in-situ particle observations, but they also identified a systematic bias: On the dawn side, the UV determined boundary was on lower latitudes than the boundaries determined by DMSP, and on the dusk side, the UV boundary was slightly poleward of the DMSP boundaries. This offset could be due to the fact that the field-aligned currents close to the OCB (Region 1) go up at dusk, and down at dawn, and therefore are associated with electron precipitation at dusk and proton precipitation (or at least not high energy electron precipitation) at dawn. This makes the boundary more prominent at dusk, compared to dawn.

4.2 Different types of aurora

In this section we look at different types of aurora. The description given here is focused on the characteristics of the precipitation, without answering why the mechanisms appeared

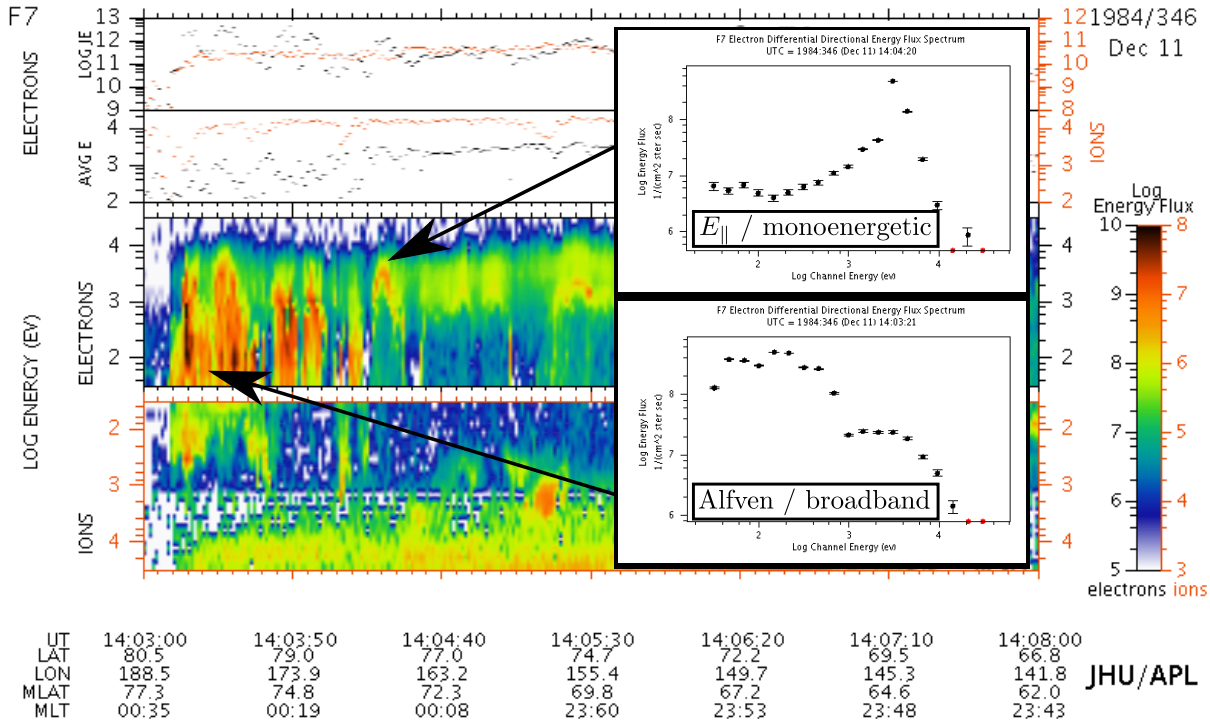

Figure 4.1: Example of a DMSP pass, with in-situ measurements of precipitating electrons and ions. The colors show differential flux in 19 different energy bins, spaced logarithmically. Two kinds of discrete aurora are seen: Alfven wave accelerated, broadband precipitation and monoenergetic precipitation, accelerated by a potential drop parallel to the magnetic field ("inverted V"). Two spectra are shown to illustrate the two types of aurora. Plots from APL website, example from *Newell et al.* (2009).

in the first place: We know that the electrons in an auroral arc can be accelerated by parallel electric fields, but the origin of the parallel electric field is still an active field of research, which we will not go in to here.

4.2.1 Monoenergetic aurora

Low-altitude satellites passing through auroral arcs often observe a monoenergetic spectrum, meaning that a small range of energies dominate completely over other energies. This characteristic is consistent with the electrons having been accelerated by a potential drop somewhere along the field line (*Evans*, 1973). The acceleration region is believed to be at 1000 km up to a few R_E. In some cases, the potential drop tends to gradually decrease away from its maximum, making the spectrogram from a satellite pass display a shape looking like an inverted V (Λ). An example of an inverted V, and a monoenergetic

spectrum is shown in Figure 4.1.

While the dynamics of parallel electric fields is a topic of extensive research, their existence can be shown to be reasonable by a quite simple argument (e.g. *Parker* (1996)). If the flow of plasma creates a curl in the magnetic field, Ampere's law (Eq. 2.5) implies a current and/or a change in electric field. The charges which are to carry the current are subject to Newton's inertia law, and one can presume that this might prevent $\mu_0 \mathbf{j}$ to immediately match $\nabla \times \mathbf{B}$. This implies an increase in electric field, in a direction such that the charges are accelerated to increase the current. This explanation is also supported by the observation that parallel electric fields are stronger when the ionospheric conductivity is low (*Newell et al.*, 1996; *Liou et al.*, 2001), which means that the number of available charges is less. Therefore, the charges must be accelerated more to carry the imposed current (see also Section 5.3.4).

4.2.2 Broadband aurora

Another energy spectrum which is often observed in auroral arcs is characterized by a high flux in a broad range of energies. Figure 4.1 also has an example of "broadband" precipitation. *Chaston et al.* (2003) showed that precipitation with this kind of spectrum is associated with electromagnetic Poynting flux, carried by Alfven waves[1] downward towards the ionosphere. Their study showed that broadband precipitation is likely energized by these waves at altitudes between 1 and 2 R_E. Alfven wave aurora often appears in the dayside cusp, and on the nightside, close to the polar cap. These locations indicate a connection between reconnection and the excitation of dispersive Alfven waves.

Newell et al. (2009) showed that the total energy carried by broadband precipitation is very sensitive to solar wind driving, increasing more than other types of aurora. They found that broadband aurora is responsible for a large fraction (28%) of the particle number flux to the ionosphere during active times. Solar wind driving was quantified using a solar wind coupling function: $v^{4/3} B_T^{2/3} \sin^{8/3}(\theta/2)$, where v is the solar wind velocity [km/s], B_T is the IMF magnitude [nT] in the plane perpendicular to the Sun-Earth line, and θ is the IMF clock angle.

4.2.3 Diffuse aurora

Diffuse precipitation is always present, and consists of particles which have not been accelerated in localized regions. It is generated when trapped magnetospheric particles are scattered into the loss cone, or new particles are injected from the tail. Diffuse electrons are scattered primarily by waves, especially by broadband electrostatic waves (*Newell et al.*, 2009). Protons can also be scattered by the stochastic motion that occurs when their gyroradius is comparable to changes in magnetic field. This is typically seen at magnetotail field lines whose apex are highly curved (*Sergeev et al.*, 1983). The latitude separating the

[1] Alfven waves are magnetohydrodynamic waves. They can be pictured as string like oscillations of magnetic field lines.

protons on curved field lines, which precipitate, and protons on dipolar field lines, whose loss cone is empty, is called the isotropic boundary. According to *Mende et al.* (2003b) the equatorward auroral boundary seen by the proton aurora imager on IMAGE, SI-12, possibly coincides with the isotropic boundary.

Newell et al. (2009) showed that the largest fraction of the total precipitating energy (61%) is carried by diffuse electron precipitation, and that the amount of energy which precipitates increases with solar wind driving.

4.2.4 Proton aurora

Electrons carry most of the energy in the night side precipitation, but the proton aurora has been shown to dominate at certain instances (*Frey et al.*, 2001; *Galand and Lummerzheim*, 2004). Even if protons are accelerated in localized regions, the proton aurora will be diffuse. This is because of their large gyro-radius, and because the precipitating protons charge exchange with the neutral atmosphere and become neutral hydrogen atoms, which can move large distances before they become ionized again (and this process can be repeated thousands of times), thus spreading the deposited energy over a large area in the ionosphere. Precipitating protons also produce secondary electrons, which then can excite other emissions. The aurora seen by global UV imagers can therefore not be attributed to electron precipitation alone. In Paper I, we use the IMAGE SI-12 camera, which so far is the only camera that have provided global images of the aurora produced only by protons. We will describe the principle behind this instrument in Chapter 6.1.4, and how it can be used to estimate and subtract the proton contribution to the other cameras in the FUV package.

One of the principal advantages of the SI-12 camera was summed up by *Mende et al.* (2003a): "[...] protons, especially on the nightside, tend to be fairly energetic, with mean energies above 10 keV, and they are only minimally modulated by the field-aligned electric fields which have a fundamental influence on the electron aurora. Therefore, auroral protons are expected to be much better tracers of magnetospheric plasma populations than electrons." Some of the existing studies utilizing the SI-12 camera have focused on the statistical morphology of the precipitation (*Coumans et al.*, 2006), the hemispheric power associated with proton precipitation (*Hubert et al.*, 2002), the statistical behavior of proton aurora during substorms (*Mende et al.*, 2003a,b), and estimates of the oval radius/open flux, due to the low contamination of dayglow in the camera (*Hubert et al.*, 2006; *Milan et al.*, 2009b).

In Paper I, we found that the proton aurora increases when the solar wind dynamic pressure is high. We also presented data from two storm main phases, which showed fundamentally different intensities, possibly because of the large difference in the solar wind dynamic pressure between the two events. An interesting follow-up study would be to look at a larger data set of SI-12 images in relation to solar wind dynamic pressure, and ring current dynamics, manifested by a drop in the SYM-H index (Section 6.4).

Chapter 5

Hemispheric differences

Papers II, III and IV are all concerned with differences between hemispheres inferred from auroral imaging. In Paper II we observe an unusual event in which the intensity of the aurora was very different in the two hemispheres. This was interpreted as an effect of magnetic field aligned currents flowing between the hemispheres. In Paper III, we studied the same event, but with a different approach, looking at spatial asymmetries in the polar cap boundary. These observations led to conjectures about the magnetosphere-ionosphere coupling, which would have implications for all events with rapid flux closure. The statistical study of images in one hemisphere, presented in Paper IV supports some aspects of these findings, but many questions remain.

In this chapter we look at what we mean by asymmetric magnetic field line footpoints. Then we survey previous studies, particularly on magnetic field asymmetries, but we also touch upon studies of convection. We then try to synthesize some mechanisms which are thought to be responsible for the inter-hemispheric asymmetries.

5.1 The meaning of asymmetric field line footpoints

In the context of the Earth's magnetosphere, the footpoints of a magnetic field line are asymmetrical if the two points are not part of the same International Geomagnetic Reference Field (IGRF) line. The IGRF is a model of the magnetic field which is generated in the Earth's interior, and it represents what the Earth's magnetosphere would look like in the absence of external influence, primarily from the solar wind and IMF. Not all types of external forcing produce asymmetric field line footpoints. An obvious example is the magnetotail, which is always present, but generally does not affect the symmetry of field line footpoints. A measured inter-hemispheric asymmetry signifies a perturbation of the IGRF which is not balanced in the opposite hemisphere.

To quantify asymmetries, it is clear that we need a coordinate system which is based on the IGRF. Such coordinate systems are Apex coordinates (*VanZandt et al.*, 1972; *Richmond*, 1995) and the Altitude Adjusted Corrected Geomagnetic (AACGM) coordinates (*Baker and Wing*, 1989).

5.1.1 The Apex coordinate system

In the papers in this thesis, we primarily use the Apex coordinate system. Consider a point at some given altitude above the surface of the Earth, given in geographic coordinates. To find the Apex coordinates of this point, we start tracing along the IGRF magnetic field line, until we reach the point that lies at the highest altitude above the earth (the slightly spheroidal shape of the Earth is taken into account). This point is the apex of the field line. Its radius (in R_E), A, is

$$A = 1 + \frac{h_A}{R_{eq}}, \tag{5.1}$$

where h_A is the apex altitude, and R_{eq} is the equatorial radius of the Earth. The apex latitude is then defined by

$$\lambda_A = \pm cos^{-1} A^{-\frac{1}{2}}. \tag{5.2}$$

This equation is the relation between the equatorial distance to a dipole magnetic field line (in this case the distance is A), and the latitude of its intersection with a unit sphere. The plus (minus) sign corresponds to the northern (southern) hemisphere. The geomagnetic dipole longitude[1] of the apex defines the apex longitude, ϕ_A. This definition ensures that points which trace out to the same apex, and thus belong to the same field lines, get symmetrical coordinates in the two hemispheres. This definition depends on altitude: Points which are on the same geographic longitude and latitude, but at different altitudes will get different magnetic coordinates. For the auroral images in this study, an altitude of 130 km was used.

The other much used coordinate system, AACGM, is defined in a similar way: Instead of tracing to the field line apex, the tracing stops at the geomagnetic dipole equatorial plane. Both coordinate systems are shown on geographic coordinates in Figure 5.1, AACGM in black and Apex in red. The deviation between the two systems is less than the uncertainty in pixel localization in global auroral imagers.

5.1.2 Effects of the non-uniform magnetic coordinate system

Several interesting phenomena can be understood by inspecting the mapping between magnetic and geographic coordinates revealed by Figure 5.1. In the absence of asymmetries, corresponding grid cells in the two hemispheres are ionospheric mappings of the same magnetic field lines, or flux tubes. Since the sizes of conjugate grid cells may be different, precipitating particles can be spread over a larger geographic area in one hemisphere, leading to a difference in particles per unit area, and hence auroral intensity. However, this effect is countered by the stronger magnetic field, and thus mirror force, in the hemisphere with the smallest area. In the case of inter-hemispheric asymmetries in magnetic field

[1]The geomagnetic dipole longitude is measured eastward of the western hemisphere half of the meridian connecting the *centered* dipole poles and the geocentric poles. A nice overview of magnetic coordinate systems, but especially the AACGM, is found in the pdf presentation file by K. Baker at http://superdarn.jhuapl.edu/tutorial/Baker_AACGM.pdf

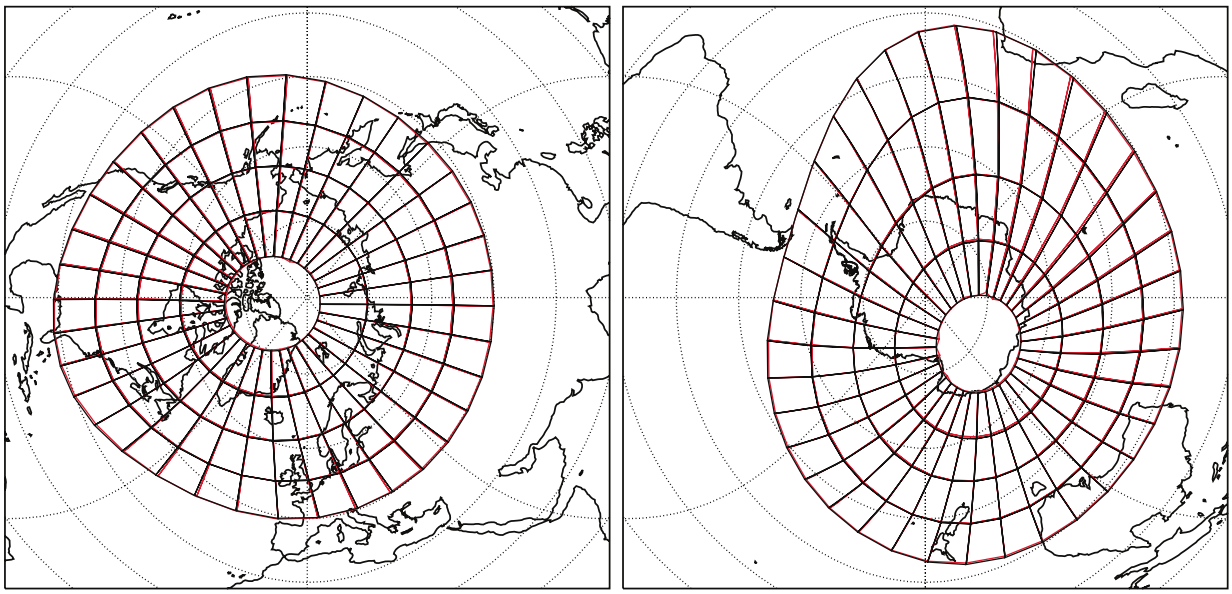

Figure 5.1: The figure shows the apex (red) and AACGM (black) coordinate systems mapped to geographic coordinates. Shown are circles of magnetic latitude at 40, 50, 60 and 80 degrees. The magnetic meridians are separated by 10 degrees magnetic longitude. The difference in total area enclosed by the apex and AACGM coordinate boxes is $\approx 0.5\%$ of the total area.

line footpoints, the corresponding grid cells in the two hemispheres will shift, and the effect of the non-uniform coordinate system can increase or decrease. Note that, due to the magnitude of the magnetic field on low altitudes, a shift in the field line footpoint is almost exclusively due to perturbations on high altitude.

The non-uniformity of the magnetic coordinate system, seen from a geographic coordinate system, may become important when comparing measurements which are ordered by the two different coordinate systems. Three examples are: 1) Studies which relate the solar illumination in the ionosphere, which is ordered by geographic coordinates, to phenomena which are ordered by magnetic coordinates, such as the onset location of substorms. 2) Dynamics of high latitude ionospheric convection: The friction and the effective inertia of the ionosphere should be higher in large grid cells (i.e., the size of the magnetic grid in geographic coordinates), since they contain more mass, while the forces driving the convection might be better ordered by a magnetic coordinate system. 3) Studies of the transition regions between plasma which is controlled by the co-rotation of the ionosphere, and plasma which is controlled by the interaction with the IMF.

One can also imagine, by looking at Figure 5.1, that the frequency of phenomena that are ordered by magnetic coordinates, e.g. substorm onsets, are longitude dependent when sorted by geographic coordinates. This is because these phenomena are spread out in a larger geographic area in large grid cells, compared to the smaller grid cells on the

same latitude, while each cell has the same number of events. A more curious effect of the irregular magnetic coordinate system is that the resolution of auroral images, when plotted in magnetic coordinates, depends on the low altitude magnetic field strength: Since the pixels are evenly spread out in geographic coordinates (on average), the large grid cells (weak magnetic field) will contain more pixels.

5.2 Previous observations of inter-hemispheric asymmetries

Here we give a review of some of the previous works reporting observations of inter-hemispheric asymmetries. A summary of some suggested causes for inter-hemispheric asymmetries is given in Chapter 5.3.

5.2.1 Simultaneous auroral images

Table 5.1 contains a list of conjugate studies of the aurora, using imaging. In the following, we summarize the findings of the works which are most relevant to the papers in this thesis. We focus on studies using imagers, although conjugate studies of the auroral zone have also been carried out using magnetometers (review by *Wescott* (1966)), riometers (e.g., *Hargreaves* (1969)), X-ray detectors carried by balloons (e.g., *Anderson et al.* (1962)), scanning photometers (*Sato et al.*, 1986), and in-situ particle detectors in one hemisphere combined with imaging in the other (e.g., *Vo et al.* (1995)).

Ground all-sky cameras: The first optical study of the aurora in both hemispheres simultaneously was conducted by *DeWitt* (1962). Using all-sky cameras in Alaska and in the south Pacific, in geomagnetically conjugate regions, he confirmed what was expected at the time, that the auroral shapes, motion, and variation in intensity was similar in the two hemispheres. Further confirmation was provided that same year, by *Anderson et al.* (1962), who detected very similar coincident X-ray fluxes from balloons flying at approximately conjugate points.

The advantage of using all-sky imagers for conjugate studies is that they can have a much better time resolution than the cameras which so far have been carried by satellites. Their proximity to the auroral emissions also enables all-sky cameras to resolve fine structures in the aurora. A pair of stations in Syowa in Antarctica and in Iceland have been used for conjugate observations for three decades, utilizing both scanning photometers (*Sato et al.*, 1986) and all-sky imagers. Ground based auroral studies require darkness and clear skies, minimal interference from moon light, and of course, auroral activity. Because of these requirements, conjugate ground based observations can only be conducted during equinox, when both hemispheres are in darkness. Excellent conditions were prevalent during an event presented by *Sato et al.* (1998), enabling a number of detailed comparisons between the hemispheres. In the course of only a few minutes,

Authors	Northern hemisphere	Southern hemisphere
(*DeWitt*, 1962) (*Bond*, 1969) (*Sato et al.*, 1998, 2004, 2005) (*Fujii et al.*, 1987) (*Watanabe et al.*, 2007a) (*Minatoya et al.*, 1996)	Ground ASC	Ground ASC
(*Belon et al.*, 1969) (*Stenbaek-Nielsen et al.*, 1972, 1973) (*Stenbaek-Nielsen and Otto*, 1997)	Airplane ASC	Airplane ASC
(*Burns et al.*, 1990)	Ground ASC	Viking
(*Craven et al.*, 1991) (*Pulkkinen et al.*, 1995)	DE-1	Viking
(*Frey et al.*, 1999) (*Vorobjev et al.*, 2001)	Polar UVI	Ground ASC
(*Fillingim et al.*, 2010)	Ground ASC	Polar UVI
Frank and Sigwarth (2003)	Polar VIS Earth	Polar VIS Earth
Zhang et al. (2006)	TIMED/GUVI	IMAGE FUV
(*Østgaard et al.*, 2003, 2004, 2005c,a, 2007) (*Stubbs et al.*, 2005) (*Fillingim et al.*, 2005) Paper II, Paper III	IMAGE FUV	Polar VIS Earth/ UVI

Table 5.1: Overview of inter-hemispheric studies of the aurora, using imagers (ASC = all-sky camera). A few early studies in conference proceedings and monographs have been omitted.

they observed conjugate aurora with larger scales in the southern hemisphere (also when accounting for the IGRF mapping, shown in Figure 5.1), arcs showing excellent conjugacy, and a non-conjugate auroral breakup in the southern hemisphere. Comparing the travel times of Alfven waves between the hemispheres to their observation of a one minute delay between auroral breakups in the two hemispheres, they suggested that the triggering source of the auroral breakup was located near the ionosphere. A later study (*Sato et al.*, 2005) reported another event with very similar auroras observed in both hemispheres. In this event however, they observed corresponding auroral features to move relative to each other, $\sim$ 200 km in longitude and $\sim$ 50 km in latitude, during the course of one hour. This motion seemed to be independent of concurrent changes in the IMF. They suggested that the change in conjugate points was due to asymmetrical field aligned currents in the two hemispheres. In light of the discussion in Paper III, we may add that these asymmetries could also have been an effect of field aligned electric potential drops or by differences in ionospheric convection (see also Chapter 5.3.3).

All-sky cameras on airplanes: Conjugate observations of the aurora have also been conducted from airplanes. *Belon et al.* (1969) reported data from three flights along conjugate segments of a magnetic meridian in 1967. The conjugate images showed remarkable good dynamical and spatial conjugacy. Tracing arcs from the southern hemisphere, along a realistic model of the Earth's magnetic field (similar to the IGRF), onto the northern hemisphere, the arcs proved to be nearly coincident. These flights were done during magnetically quiet periods, but later flights, when the magnetic activity was high, contained events in which bright aurora appeared exclusively in one hemisphere. This disturbed event was later studied in more detail by *Stenbaek-Nielsen et al.* (1972), who concluded that the observations either implied that the aurora was non-conjugate, or that it was displaced from its counterpart in the opposite hemisphere by more than 1000 km. At auroral latitudes, this corresponds to a displacement of approximately 20° longitude, or $\approx$1.3 h MLT, which we now know is high, but not unheard of (*Østgaard et al.*, 2004).

A substantial result from the flight campaigns was that the inter-hemispheric displacement seemed to be stronger, and more transitory on higher latitudes. During the course of a substorm (as evidenced by a magnetic bay), the relative displacement was seen to change signs. On lower latitudes, *Stenbaek-Nielsen et al.* (1972) reported a stable arc system, which consistently showed a large degree of symmetry. *Stenbaek-Nielsen et al.* (1973) looked at intensity differences between hemispheres, and found that in all the flight observations, the equatorward arc was more intense in the northern hemisphere. At the longitude of the flight paths, the magnetic field is stronger in the southern hemisphere, which increases the mirror force there, preventing precipitation (see Section 5.3.4).

In a later paper, *Stenbaek-Nielsen and Otto* (1997) revisited the aircraft measurements, and argued that some of the observations, showing large inter-hemispheric variations in the distance between the equatorward arcs and poleward arcs, could not be reasonably explained in terms of spatial displacement. They offered a new interpretation, that some of the poleward arcs were in fact non-conjugate, and that the non-conjugacy arises from an

inter-hemispheric current. The inter-hemispheric current was suggested to be caused by a partial penetration of the IMF B_y component into the magnetotail. They argued that this penetration is lower at the earthward and tailward ends of the plasma sheet, implying two shears, or currents. For a positive B_y, the direction of these predicted inter-hemispheric currents would be from the northern to the southern hemisphere at the earthward end (presumably modifying the Region 2 current system) of the penetration region, and from the southern to the northern hemisphere at the tailward end (modifying the Region 1 current system). A more detailed description of inter-hemispheric currents is given in Chapter 5.3.4.

Ground all-sky and Viking: *Burns et al.* (1990) compared keograms derived from cameras at the South pole station, and from the Viking satellite in the northern hemisphere, during northern summer season. During the three events studied, they found that conjugate points were displaced up to 1.9 hours MLT, and 5.3° latitude. During one of the events, they observed the southern aurora to propagate further towards the pole than the aurora in the north. Given the similar seasonal conditions and type of asymmetry, this could represent the same kind of asymmetric poleward propagation reported in Paper III.

Viking and DE: The first purely space based opportunity for conjugate studies of the global aurora came with the launch of the Swedish satellite, Viking, in 1986. At this time, the Dynamics Explorer 1 was already in orbit, and both these satellites carried instruments which could observe the global UV aurora. Using these instruments, *Pulkkinen et al.* (1995) studied a substorm observed in both hemispheres. Similar features were observed, and the mapping of the auroral luminosity to the tail showed consistent results between the hemispheres.

Ground all-sky and Polar: *Frey et al.* (1999) and *Vorobjev et al.* (2001) used the UVI instrument on Polar together with ground all-sky cameras in Antarctica to study conjugate auroras. *Vorobjev et al.* (2001) focused on the poleward boundary of the aurora, in relation to the orientation of the IMF. They found that this boundary could be displaced by up to 5° latitude. Observations from 10 time intervals showed the night side aurora to be located on higher (lower) magnetic latitudes in the northern hemisphere when the IMF was in the $B_x < 0, B_y > 0$ ($B_x > 0, B_y < 0$) Parker spiral sector, compared to the aurora in the southern hemisphere. This displacement is in accordance with magnetic field perturbations on the nightside in the same direction as the IMF (see Figure 5.2).

Polar only: *Frank and Sigwarth* (2003) reported conjugate observations of a substorm using a single camera, VIS Earth on the Polar satellite. At the time of this event (1 November 2001), Polar had its apogee close to the geographic equatorial plane, on the nightside of Earth, and the large field of view of the VIS Earth camera enabled simultaneous observations of the nightside auroral zone in both hemispheres. The onset of the substorm occurred with a displacement of $\sim$ 40 minutes of local time. The intensity of the aurora

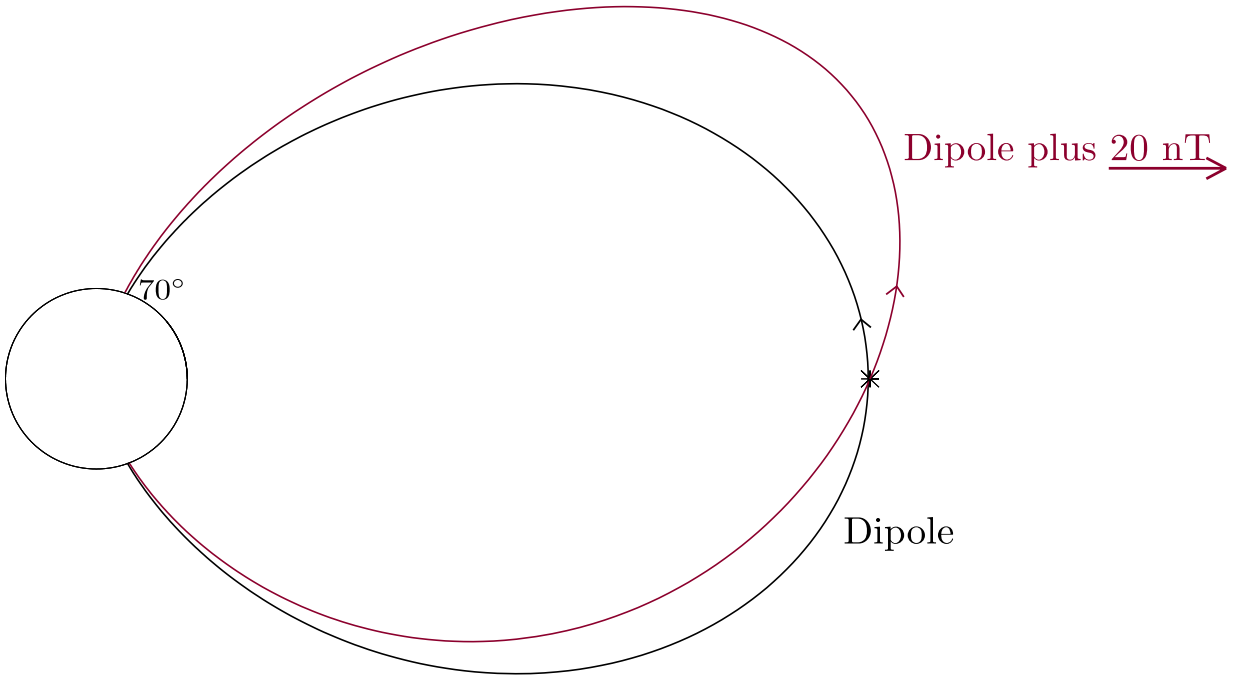

Figure 5.2: A dipole magnetic field line (using the dipole moment of the Earth, 8.05×10^{22} Am^2 (*Baumjohann and Treumann* (1997), p. 32)), intersecting the Earth surface at 70° latitude (black), and the field line of a dipole plus a 20 nT uniform field towards the right in the figure (red). The tracing of the "dipole plus uniform field"-line, using the Runge-Kutta-Fehlberg method with adaptive stepsize, was started at the crossing of the equatorial plane (asterisk). Its footpoints are displaced poleward in the northern hemisphere, and equatorward in the southern hemisphere, compared to the dipole field line. In this figure the dipole field line is in the same plane as the perturbation field, resulting in a latitudinal asymmetry. In all other cases, the perturbation field will produce a longitudinal asymmetry as well.

was also reported to be higher in the southern hemisphere. Since the southern hemisphere had summer in this event, the asymmetry in the intensity is opposite to what has been found in statistical studies of seasonal effects (described later in this chapter).

Polar and IMAGE: While still a scarce data set, the best global coverage of the aurora in the two hemispheres was provided by the IMAGE and Polar satellites. On a few occasions, primarily in the first years of the IMAGE mission, 2000 to 2002, the constellation of these satellites was such that both hemispheres could be observed simultaneously, at an unprecedented spatial and temporal resolution. *Østgaard et al.* (2004) identified 12 auroral forms which could be seen in both hemispheres, and showed that these forms were displaced in longitude, depending on the orientation of the IMF, in the yz plane. The displacement

was consistent with a partial penetration of the IMF to the magnetosphere. In a later study, *Østgaard et al.* (2005c) showed that the observed displacements were an order of magnitude stronger than the displacement predicted by the Tsyganenko 96 and 02 magnetic field models (*Tsyganenko* (2002) and references therein). They also found indication that the dipole tilt angle may act as a secondary controlling parameter, displacing the winter hemisphere footpoint dawnward compared to the summer hemisphere footpoint.

Stubbs et al. (2005) were the first to use these satellites to study the time development of inter-hemispheric asymmetries, during a 1.5 hours interval. Using a circle fitting technique, similar to *Holzworth and Meng* (1975), they found that the center of the ovals changed in step with concurrent variations in the IMF, in agreement with the "dipole plus uniform field" picture described by *Cowley et al.* (1991), shown in Figure 5.2. They also found that the ovals in both hemispheres were displaced towards dawn.

Fillingim et al. (2005) used the UVI camera on the Polar satellite, during an event when this camera observed the southern hemisphere afternoon sector while IMAGE WIC observed the northern hemisphere. They found that in the southern hemisphere, the afternoon aurora was structured, showing a "string of pearls" configuration, while the afternoon aurora in the northern hemisphere was more uniform. Because of modest seasonal differences, the most likely cause for this inter-hemispheric difference was the positive IMF B_y. Positive B_y is often associated with a crescent shaped convection cell at dusk in the southern hemisphere, and a round cell in the northern hemisphere. *Fillingim et al.* (2005) proposed that the stronger flow shears, and associated field-aligned currents in the southern dusk sector was the cause for the different morphology of the afternoon aurora.

5.2.2 Statistical studies of one hemisphere

Because of the relatively few truly conjugate observations, most of our quantitative knowledge about inter-hemispheric asymmetries stems from observations from only one hemisphere. To investigate the effect of IMF B_y, B_x and dipole tilt angle, such studies require a large number of measurements. To infer about instantaneous inter-hemispheric differences, one must assume that the northern hemisphere reacts to positive B_y, B_x and dipole tilt angle, in the same way as the southern hemisphere reacts to negative B_y, B_x and tilt, respectively.

Auroral morphology

Holzworth and Meng (1984) employed auroral images from a DMSP satellite to study the polar cap boundary in relation to the IMF orientation. Using a technique described by *Holzworth and Meng* (1975) they fitted the observed points of the polar cap boundary to a circle. The center of the circle in the southern hemisphere was observed to be displaced towards dusk (dawn) when $B_y > 0$ ($B_y < 0$), by 1° per 2 nT B_y. They also observed a shift of the center in this circle, independent of the IMF, $\sim 5°$ tailward, in agreement with an earlier study of a subset of the same data, by *Meng et al.* (1977). The B_y dependent dawn-dusk shift was later pointed out by *Cowley et al.* (1991) to be consistent with the

IMF creating a global perturbation to the geomagnetic field in the same direction as the IMF, the "dipole plus uniform field" picture. Figure 5.2 illustrates this concept, for a dipole field line in the same plane as the perturbation field.

Later, *Oznovich et al.* (1993) showed, using UV images of the aurora from the Polar BEAR satellite, that the auroral oval is displaced further towards midnight in the summer hemisphere, compared to the winter hemisphere. The displacement at noon and midnight was approximately 1° per 10° tilt angle.

Elphinstone et al. (1990) used images from the Viking satellite, in conjunction with IMP-8 measurements of the IMF, during times when the IMF B_z was positive. Among other findings, they concluded that in the northern hemisphere, the polar cap was displaced towards dawn (dusk) when $B_y > 0$ ($B_y < 0$), consistent with the findings of *Holzworth and Meng* (1984).

The repeating pattern of auroral substorms make them ideal for statistical studies of the IMF influence on the aurora. *Liou et al.* (2001) used observations of 648 substorm onsets, observed by Polar UVI, to study how their average location is affected by the IMF and by seasons. They found that the average location of substorm onsets is 22:30 MLT, and 67° magnetic latitude. When the IMF $B_x > 0$ ($B_x < 0$), the average onset location is shifted equatorward (poleward). They also found that, when keeping $|B_x| < 1$ nT, the average location of substorm onset is shifted westward in the northern hemisphere when the IMF $B_y > 0$, and eastward when $B_y < 0$. These shifts in average onset locations are consistent with a perturbation of the magnetic field in the direction of the IMF. They also found that when separating the substorms according to seasons, the onset on average occurs ~ 1.5 hours dawnward (closer to midnight) during winter, compared to summer.

Several similar studies have later been undertaken, based on observations of substorm onsets by the IMAGE FUV instrument. The basis for these studies is a list of more than 2400 onsets and their locations, compiled by *Frey et al.* (2004). Their list has later been expanded to include 4193 substorms from the whole IMAGE mission life time (*Frey and Mende*, 2006). Using this list, *Østgaard et al.* (2007) showed that the average MLT of the onset depends on the clock angle of the IMF. This result was contested by *Wang et al.* (2007), who, using the same data set, found that B_y exercises a better control over the onset location, than the clock angle. *Wang et al.* (2007) did not find evidence for inter-hemispheric differences in the latitude of the substorm onset, in contrast to the *Liou et al.* (2001) study. Not surprisingly, both *Liou et al.* (2001) and *Wang et al.* (2007) found that the latitude of the onset moves equatorward with an increasingly negative IMF B_z, presumably because this enhances the dayside reconnection, opening more flux, and consequently expanding the auroral oval. However, this effect is expected to occur in both hemispheres.

We note that the correlation between the onset MLT and the solar wind B_y is low: For the 4089 substorms in the *Frey et al.* (2004) list for which ACE data is available (see Paper IV), the correlation is only -0.19 (reversing the sign of B_y for observations in the southern hemisphere). This is significant (the chance of it being random is less than 0.01%), but it also means that a linear relation between B_y and the onset MLT can only explain $\approx 4\%$ of the variation in onset location. A natural interpretation of this, is that B_y does not

affect where in the magnetospheric equatorial plane the onset occurs, but it does affect the geometry of the field, and hence the mapping to the ionosphere.

Auroral intensity

The precipitation power, as well as the mean energy of the precipitating particles, can also vary between hemispheres. *Newell et al.* (1996) used DMSP electron energy spectra in a statistical study investigating the relation between electron energy and the degree of solar illumination in the ionosphere. They found that the probability of observing monoenergetic particle precipitation (electrons accelerated by parallel electric fields) is significantly higher when the ionosphere is in darkness. *Liou et al.* (2001), using the ratio of different passbands in the UVI imager, confirmed that on the night side, the average energy of precipitating electrons, as well as the total energy flux, is higher in darkness (by a factor of $\sim$ 3). On the dayside however, the energy flux is higher in summer, by a factor of $\sim$ 2. *Coumans et al.* (2004), using the SI-12 imager on IMAGE, showed that proton precipitation shows a generally opposite dependence on seasons, compared to electrons. The different dependencies for electrons and protons reflect the proposed seasonal dependence of parallel electric fields.

In a more recent study, by *Newell et al.* (2010), the seasonal dependencies of the various types of aurora were investigated. They found that the monoenergetic aurora is most dependent on seasons, with the nightside power being a factor of 1.7 higher during winter than during summer, when the solar wind driving (see Chapter 4.2.2) is strong. The Alfven wave aurora and diffuse electron aurora also show a seasonal dependence, with factors of of 1.26 and 1.3 higher power in winter. On the dayside, the electron auroral power is generally higher in the summer hemisphere, although the differences are less pronounced.

The orientation of the IMF has also been reported to cause intensity variations which are expected to be different in the two hemispheres. *Liou et al.* (1998) reported a significant increase in the auroral intensity, measured by Polar UVI, on the nightside in the northern hemisphere, when B_y is strongly negative. *Liou et al.* (1998) (and references therein) also showed that in the afternoon sector in the northern hemisphere, the aurora is more intense when B_x is negative. Later, *Shue et al.* (2002), who also used the UVI instrument, showed that the global auroral power in the northern hemisphere is higher when B_x is negative.

5.2.3 In-situ measurements of the magnetic field

In-situ observations confirm that the IMF B_y is associated with a similarly directed magnetic perturbation on closed field lines. *Wing et al.* (1995) studied 5 and 6 years of magnetic field measurements from the geo-stationary satellites GOES-6 and GOES-7 in relation to the y component of the IMF, measured by the IMP-8 satellite. B_y at geostationary orbit was found to be well correlated with B_y in the solar wind at all local times. Maximum correlation was found at midnight and at noon. However, the slope of the regression lines between the two variables had a minimum at midnight, increased towards the flanks and reached maximum on the dayside. This means that variations in the nightside magnetic

field are more sensitive to IMF B_y, but the amplitude of the variation, or degree of penetration, is stronger on the dayside.

The IMF orientation also affects the magnetic field further downtail (*Fairfield*, 1979). *Tsurutani et al.* (1984) showed that in the distant tail ($150 - 238R_E$ downtail), B_y in the north-dawn lobe and in the south-dusk lobe was correlated with positive IMF B_y. Practically no correlation with positive IMF B_y was found in the two other lobe quadrants. For negative IMF B_y, the situation was the opposite, with correlation between IMF B_y and north-dusk and south-dawn lobe B_y. The same asymmetry in lobe susceptibility to IMF B_y was later reported by *Khurana et al.* (1996), who presented measurements from $x = 43 - 87R_E$ downtail.

In a recent study, *Petrukovich* (2009) used 11 years of Geotail data from $|Y_{GSM}| < 15R_E$, and $-31R_E < X_{GSM} < -8R_E$, and found that B_y in the tail increases with increasing tilt angle (the northern hemisphere increasingly tilting towards the sun), after having subtracted the effect of the IMF. That is, during northern summer, the tail B_y is more positive, and during winter, it is more negative. This effect seemed to decrease with increasing distance from the Earth. One possible source for this is the shift in position of the sunlight terminator in the ionosphere. This connection is suggested by theoretical works (see Chapter 5.3.2), and studies of convection which are reviewed below.

5.2.4 Inter-hemispheric differences in ionospheric convection

As argued in Chapter 2, the magnetic field and plasma are frozen-in in most of the magnetosphere. Since the magnetic field moves with the plasma, we expect asymmetries in the the magnetic field to be closely linked to asymmetries in the convection. With very few exceptions, all empirical knowledge about global magnetospheric convection stems from measurements of the convection in the ionosphere. Here we briefly review some of the numerous studies that have been conducted of ionospheric convection patterns, and how these change with different seasons and the sign of IMF B_y when $B_z < 0$.

One of the first documented effects of different signs of the IMF B_y is that Hall currents in the high-latitude dayside ionosphere tend to flow eastward when $B_y > 0$ and westward when $B_y < 0$. This is called the Svalgaard-Mansurov effect (*Svalgaard*, 1968; *Mansurov*, 1969). Since the collision frequency with the neutral atmosphere is higher for ions than for electrons, ionospheric convection sets up these Hall currents in the opposite direction of the convection (the direction of the positive ions relative to the electrons). The Svalgaard-Mansurov effect is therefore equivalent to westward convection on the dayside during $B_y > 0$ and eastward convection when $B_y < 0$.

Statistical studies of average convection

Global convection patterns, derived from statistical ensembles of various measurements, have confirmed these findings. *Heppner and Maynard* (1987) and *Weimer* (1995) used DE-2 measurements, *Papitashvili and Rich* (2002) used DMSP, *Ruohoniemi and Greenwald* (2005) used SuperDARN radar measurements, and *Haaland et al.* (2007) used mea-

surements from Cluster. The latter study was the first to derive global convection patterns from measurements at magnetospheric altitudes. Figure 5.3 shows their measured convection patterns, mapped to the northern hemisphere ionosphere. Each pattern corresponds to a given sector of concurrent IMF clock angle.

This figure, which is in good agreement with the other studies cited above, reveals several details about how the IMF orientation affects the ionospheric and magnetospheric convection. When the IMF is southward, the convection is significantly stronger than when it is northward. This is in good agreement with what we expect from the discussion in Section 3.1. The convection is also profoundly affected by the sign of B_y. For positive B_y, the convection cell at dawn is crescent shaped, and the dusk convection cell is more circular. For negative B_y, we see the opposite configuration, although the differences between the convection cells are much less prominent. This lack of mirror symmetry between different signs of B_y appears in most large statistical studies of the convection.

Ruohoniemi and Greenwald (2005), who used SuperDARN measurements, found that the tendency that B_y sculpts a pair of crescent/round convection cells is reinforced for the combinations $B_y > 0$/summer and $B_y < 0$/winter. These combinations coincide with the findings of *Petrukovich* (2009), that the "penetration" of IMF B_y is reinforced in the tail for the same combinations. In Section 5.3, we discuss how this may be seen as an effect of ionospheric feedback on magnetospheric dynamics.

Conjugate measurements of convection

With the increased coverage of ground measurements in both hemispheres, it has become possible to study instantaneous inter-hemispheric differences in the convection. This was first done by *Lu et al.* (1994), using the Assimilative mapping of ionospheric electrodynamics (AMIE) technique. This technique employs all available measurements of convection (SuperDARN, DMSP, incoherent scatter radars), magnetic field (ground magnetometers) and ionospheric conductance (incoherent scatter radars, and indirectly from auroral imaging, ground magnetometers, and in-situ particle measurements from DMSP and NOAA), to produce an instantaneous map of ionospheric convection, conductance, and field-aligned current. *Lu et al.* (1994) studied a three days period, covering both positive and negative IMF B_z. When B_z was southward, they found a two-cell pattern which was largely mirror-symmetrical in the two hemispheres. When it was northward, lobe reconnection in one hemisphere produced significant differences in polar cap convection.

Later, studies of instantaneous conjugate ionospheric convection patterns have been undertaken, using SuperDARN (*Grocott et al.*, 2005; *Ambrosino et al.*, 2009). *Ambrosino et al.* (2009) looked at the response in the dayside convection patterns to rotations of the IMF. They found an immediate change in the convection in both hemispheres when the IMF turned southward, while the southern hemisphere response to a northward turning lagged behind the northern hemisphere by ~ 10 minutes. This time difference led them to suggest that the reconfiguration of the ionospheric convection was caused by lobe reconnection, and that this happened faster in the northern hemisphere because of the concurrent negative IMF B_x.

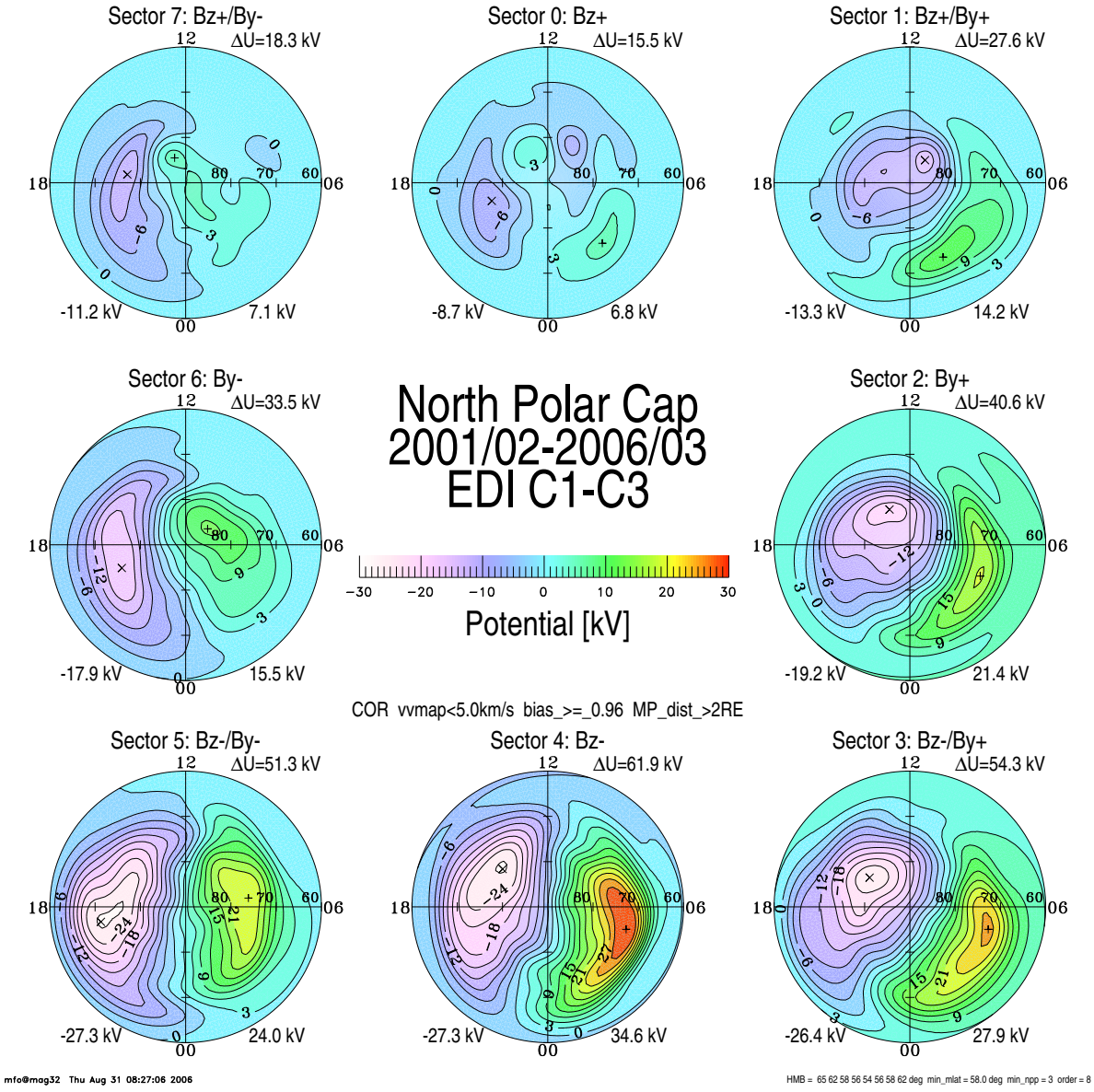

Figure 5.3: Electric potential, ϕ, composed by Cluster EDI measurements of plasma velocity, mapped to the ionosphere. The potential patterns are binned by concurrent IMF clock angle sectors. The convection velocities ($\mathbf{v} = -\nabla\phi \times \mathbf{B}/B^2$) are along the equipotential contours, clockwise where the potential is negative, and counter-clockwise where the potential is positive. Equipotential contours are spaced by 3 kV. From *Haaland et al.* (2007).

Ionospheric convection during substorms

Common to most of the statistical studies of convection is that they mainly considered steady IMF conditions, and relatively quiet conditions in the magnetosphere. Our observations in Papers III and IV are suggested to imply differences in the convection between the two hemispheres during a substorm expansion phase, which is not necessarily well represented by statistical studies based on steady conditions. A few studies exist which more directly observe the ionospheric convection following the onset of substorms. *Blanchard et al.* (1997) looked at the convection using the Sondrestrom incoherent scatter radar during 24 events. They found that on average, the convection does not increase until ~ 20 minutes after the substorm onset. Subsequent studies have employed the SuperDARN radars to study the convection during substorms statistically. Based on 67 events, *Provan et al.* (2004) reported a faster response than what was observed by *Blanchard et al.* (1997). *Bristow and Jensen* (2007), who looked at 10 substorms, observed an abrupt decrease in convection velocity, accompanied by a rotation from zonal to meridional, equatorward flow. A statistical study of 1979 isolated substorms, by *Grocott et al.* (2009), showed that during the expansion phase, the convection tends to increase for substorms with onset at high latitudes, and decrease for substorms with onset on lower latitudes. It should be noted that none of these studies focused on seasonal effects. Seasonal differences in ionospheric convection during substorms seem to be implied by the observations in Papers III and IV.

5.2.5 Summary

Before we look at some of the prevailing theories explaining the observed inter-hemispheric asymmetries, we briefly sum up the most important and established results reviewed above. Conjugate auroral imaging show that on high latitudes, the conjugate field line footprints can be significantly distorted from the IGRF predictions. Auroras on lower latitudes have been observed to be mostly symmetrical. From conjugate auroral imaging, statistical studies of the aurora, and in-situ magnetic field measurements, we know that on average, there is an effective "penetration" of the IMF to the magnetosphere, on both open and closed field lines. This is certainly true for the B_y component, while ambiguous results exist for the effect of B_x. However, this penetration is likely not uniform, as evidenced by the different correlations found in different sectors of the lobe (*Tsurutani et al.*, 1984), and at different local times (*Wing et al.*, 1995). The B_y component also has a very clear effect on the average ionospheric convection, sculpting a pair of crescent/round convection cells. Compelling evidence exists that the convection pattern for $B_y < 0$ is not the exact mirror image of $B_y > 0$ patterns, presumably due to a secondary seasonal control. Seasonal conditions is also a decisive factor for the acceleration of auroral electrons. In the dark hemisphere, electron spectra are harder than in the sunlit hemisphere. The auroral power is higher on the dayside during the summer, but it is lower on the nightside, compared to winter conditions.

5.3 Causes for inter-hemispheric asymmetries

In this section we look at theoretical works which address possible causes for inter-hemispheric asymmetries. We retain the approach from Section 2.3, focusing on the forces acting on the plasma fluid, causing the plasma to move such as to manifest the observations reviewed above. We start by considering asymmetrical forces acting on the magnetosphere from the solar wind, or forces which can be considered to directly generate the asymmetries. Then we look at how an asymmetrical ionospheric response between hemispheres may affect the global geometry of the magnetosphere. We also look at how regions where ideal MHD breaks down can contribute to the observed inter-hemispheric asymmetries. Finally we look at some proposed causes for inter-hemispheric differences in auroral intensity.

5.3.1 Asymmetric driving of the magnetosphere

In an effort to explain the Svalgaard-Mansurov effect, *Jørgensen et al.* (1972) proposed that newly opened magnetic field lines in the dayside magnetosphere are subject to magnetic stresses in the east-west direction. Figure 5.4a is copy from the paper by Jørgensen, showing that when the IMF has a positive B_y component, the geometry of newly opened field lines implies curvature forces, which produce flow from dusk to dawn in the northern hemisphere. For negative B_y, the effect is the opposite. This is in good agreement with observations of convection on the dayside, and this effect has been the starting point of many subsequent theories, addressing convection, ionospheric current systems, and inter-hemispheric asymmetries in field-line footpoints.

Although its influence is much less prominent in observations, a similar effect can be ascribed to the IMF B_x. Figure 5.4b is a copy from *Cowley* (1981a), who proposed the following scenario: If B_x is positive, the curvature on newly opened field lines will exert a stronger anti-sunward force in the northern hemisphere, causing an acceleration of the flow on these field lines, compared to the southern hemisphere. When the open field lines have crossed the dawn-dusk meridian, the curvature force will act in the opposite direction. These sunward forces will also be different in the two hemispheres, causing a more efficient breaking of the open field lines in the southern hemisphere, compared to the northern hemisphere. This implies that magnetic flux is added more rapidly to the northern lobe at any given distance along the GSM x axis, displacing the neutral sheet in the negative z direction. The result is equivalent to a net penetration of IMF B_x to the magnetosphere.

Based on earlier work by *Cowley* (1981b), their own observations and MHD modeling, and the observations by *Tsurutani et al.* (1984) of non-uniform (in tail cross-sections) correlation between the IMF B_y and lobe B_y, *Khurana et al.* (1996) presented a conceptual model to explain how the IMF affects the closed field line geometry in the magnetosphere. Figure 5.5a shows a cross section of the magnetotail. It illustrates an idea first suggested by *Cowley* (1981b), that in the presence of IMF $B_y \neq 0$, the normal component of the magnetic field through the magnetopause will be stronger in certain sectors. This is a consequence of the azimuthal B_y generated flow on newly opened field lines, which implies asymmetric addition of open flux to the lobe. For positive B_y, magnetic flux will enter the

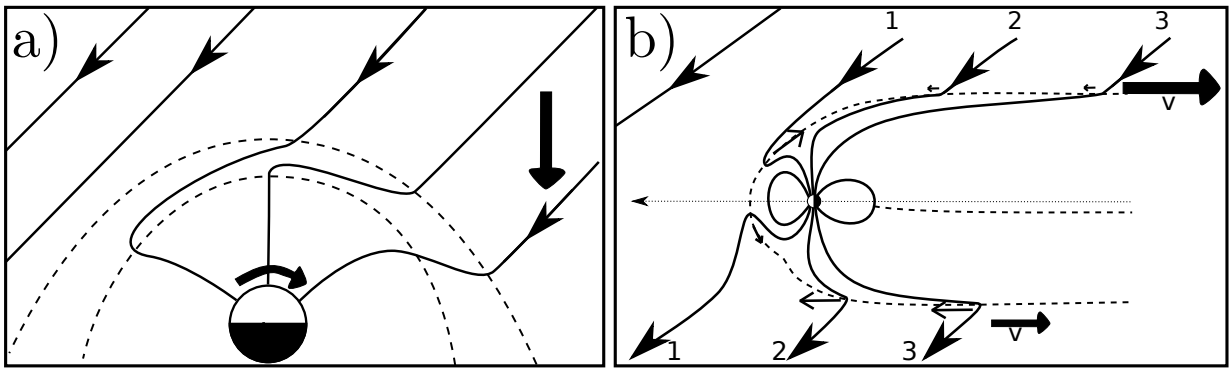

Figure 5.4: a) Equatorial cross section of the magnetosphere, in the presence of positive IMF B_y. Due to the orientation of the IMF, newly opened field lines are subject to curvature force, pulling them from dusk to dawn on the dayside. The resulting flow is shown with thick black arrows. After *Jørgensen et al.* (1972). b) Cross section of the magnetosphere in the xz-plane in the presence of positive IMF B_x. Curvature forces on open field lines (thin arrows) cause the field lines to move faster anti-sunward in the northern hemisphere. The thick arrows indicate flow velocity. The asymmetric velocity causes an increased pressure in the northern lobe, displacing the neutral sheet in the negative z direction. After *Cowley* (1981a).

magnetosphere preferentially in the northern dawn and southern dusk sectors of the tail, and the two remaining sectors will essentially be closed. *Khurana et al.* (1996) proposed that the asymmetric addition of flux and plasma which is implied by this geometry leads to a higher pressure, magnetic and thermal, in the open sectors. The pressure differences across the tail then induces a flow towards the closed parts of the lobe, indicated by the gray arrows in Figure 5.5a. Figure 5.5b shows the same phenomenon projected on the equatorial plane of the magnetosphere. The numbers indicate the progression of open field lines as they are being transported by the solar wind anti-sunward. In the northern lobe (left), magnetic flux enters through the magnetopause on the dawn flank (IMF $B_y > 0$). Since the magnetospheric part of the field line has been slowed down compared to its far end, the field lines gradually align with the x axis, decreasing B_y in the lobe. This process also implies convection from dawn to dusk in the northern lobe, and from dusk to dawn in the southern lobe (right part of Figure 5.5b). It also implies that in the closed sectors of the lobe, the magnetic field will be almost unaffected by the sign of the IMF B_y. The duskward convection in the northern lobe is also consistent with the duskward convection seen on the night-side of the polar cap, when $B_y > 0$ (see e.g., Figure 5.3), which takes over after the dawnward convection on the dayside which presumably is due to magnetic stresses on newly opened field lines (*Jørgensen et al.*, 1972).

The shear flow across the neutral sheet which is implied by this model is what *Khurana et al.* (1996) proposed leads to the observed "penetration B_y" on closed magnetic field

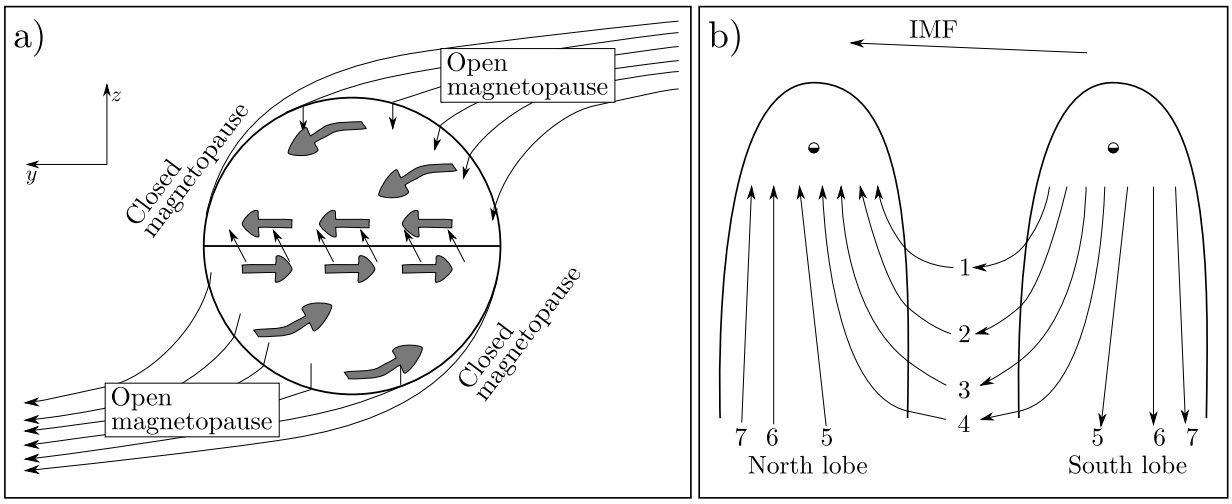

Figure 5.5: a) Cross section of the magnetotail in the presence of positive IMF B_y. Flux is added to the tail asymmetrically, causing the boundary to be essentially closed in the northern dawn and southern dusk. The asymmetrical addition of flux causes a gradient in the pressure, which induces the convection indicated by the thick gray arrows. b) Evolution of open magnetic field lines in the tail, in the northern (left) and southern (right) lobes. After *Khurana et al.* (1996).

lines. If this shear flow partially extends to closed field lines, it implies a perturbation B_y in the same direction as the IMF B_y. The resulting field is shown as tilted arrows at the demarcation line between the two lobes in Figure 5.5a. *Khurana et al.* (1996) suggested two ways that the shear flow in the lobes can be extended to closed field lines: 1) The inertia of the lobe will be partially preserved when field lines are closed, implying a continuation of the shear flow on closed field lines, and 2) viscous drag between the two regions, through a wave particle interaction mechanism.

5.3.2 Asymmetrical ionospheric feedback

In the above models, the ionosphere was regarded as passively (or at least symmetrically) complying with the imposed magnetospheric flow. This idealization has been prevalent in most theoretical works on inter-hemispheric asymmetries. However, as has become more and more clear from observations (including Papers III and IV), the ionosphere may play a key part in producing asymmetries at high latitudes.

In the time-dependent case, when the convection in the magnetosphere changes, the time it takes for the ionosphere to adapt to these changes depends, among other things, on the conductivity. The way that the magnetospheric flow, or the imposed stresses, are communicated to the ionosphere, is by shear Alfven waves (e.g. *Song et al.* (2009)). These waves can be reflected, depending on conditions in the ionosphere. *Scholer* (1970) derived

the following equation for the reflection of Alfven waves in the ionosphere:

$$R = \frac{E^{ref}}{E^{inc}} = \frac{\Sigma_A - \Sigma_P}{\Sigma_A + \Sigma_P} \tag{5.3}$$

where E is the electric field of the incident (superscript inc) and reflected (ref) Alfven wave, Σ_A is the Alfven conductance, $1/\mu_0 v_A$, and Σ_P is the ionospheric Pedersen conductance. Usually, $\Sigma_P > \Sigma_A$, and the ratio will be negative. This implies, considering a superposition of the incident and reflected wave, that the electric field is reduced. In the case that $\Sigma_P \gg \Sigma_A$, the coefficient is -1, and the ionosphere will remain unaffected by the wave, tying the magnetic field line footpoint to its current position (*Coroniti and Kennel*, 1973). Because of the frozen-in property, field-line tying will affect the geometry and convection in the magnetosphere.

The Pedersen conductance increases with the degree of solar illumination, and precipitation of particles, preferentially of low energies (*Robinson et al.*, 1987). Both of these parameters may well vary between hemispheres, causing the ionosphere to respond differently to magnetospheric convection in the two hemispheres. The different degrees of field-line tying will then lead to asymmetries between hemispheres, independent of whether or not the driving forces in the magnetosphere are symmetrical between hemispheres.

Atkinson and Hutchison (1978) showed that gradients in ionospheric conductance also can affect the convection in the steady state. Using a very simple model, they were able to qualitatively reproduce the observed accumulation of ionospheric convection equipotential contours at dawn, which seems to be implied by the lack of mirror symmetry between the average convection patterns when B_y is positive and when it is negative (e.g. *Heppner and Maynard* (1987); *Ruohoniemi and Greenwald* (2005)). Later, *Tanaka* (2001) confirmed, using an MHD model, that when the ionospheric conductance is uniform (Figure 5.6a), the convection is largely mirror symmetrical for different signs of B_y, and that deviations from mirror symmetry only appear when gradients in the ionospheric conductance are introduced (Figure 5.6b). The lack of mirror symmetry also affects the convection at magnetospheric altitudes, as demonstrated by *Haaland et al.* (2007), who used in-situ measurements from the Cluster spacecraft to derive average convection patterns, mapped to the ionosphere. Combined, these results strongly indicate that the ionosphere affects magnetospheric convection (and hence the geometry) also in the average, time-independent case.

Based on this, we may interpret the findings of *Ruohoniemi and Greenwald* (2005), that the sculpting of crescent/round convection cells are reinforced with the combination of summer/positive B_y and winter/negative B_y, to be an effect of the changing average location of the sunlight terminator with seasons. The same can be said about the findings by *Petrukovich* (2009), that the "penetration" B_y is more strongly positive (negative) in the magnetosphere during northern summer (winter), as well as the secondary influence of dipole tilt angle on the inter-hemispheric longitudinal displacement of the aurora, indicated by the data set of *Østgaard et al.* (2005c).

In summary, there is clear evidence, both from theoretical considerations and from observations, that the ionosphere significantly modulates the magnetospheric dynamics,

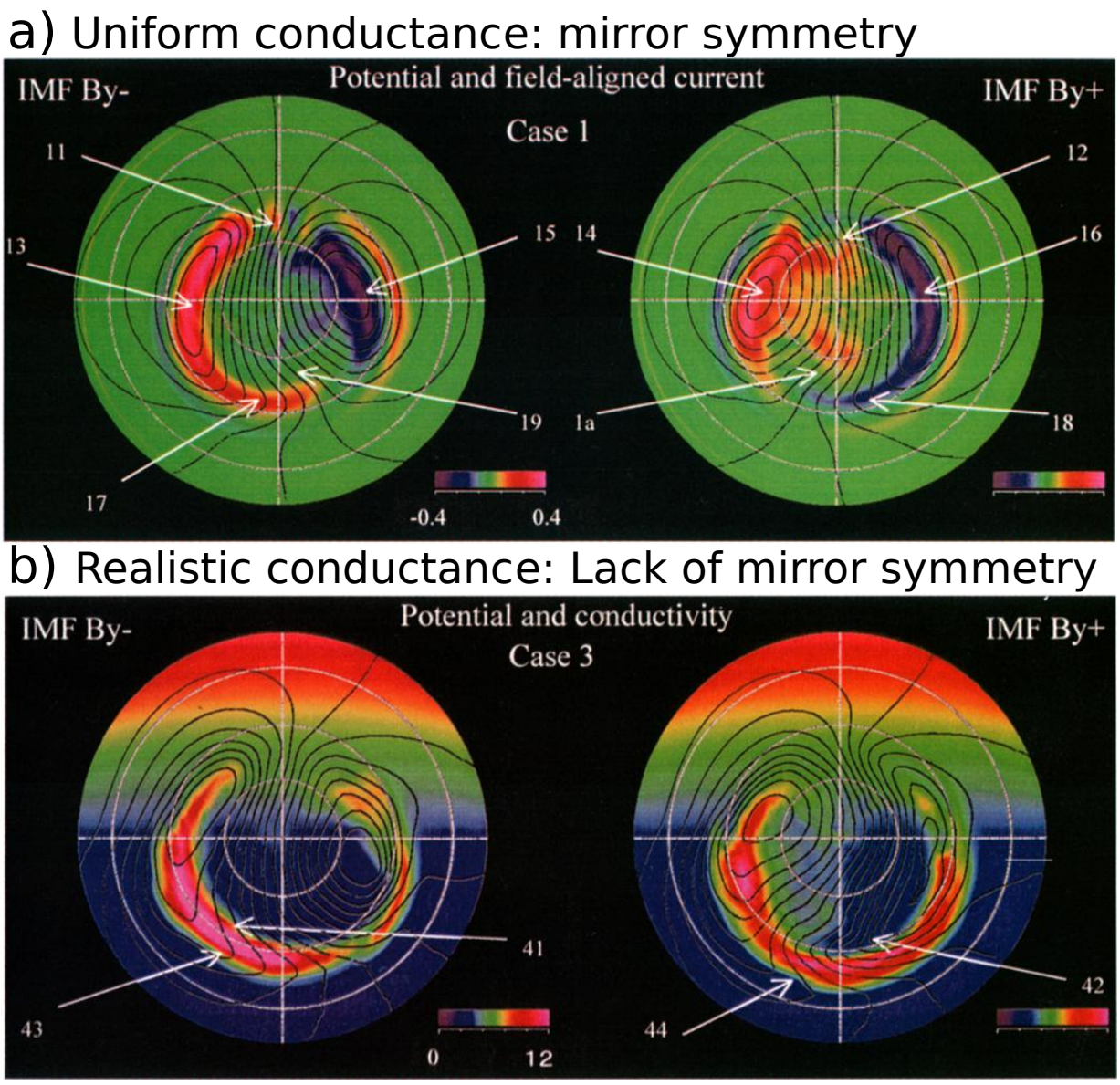

Figure 5.6: Figures 1 (a) and 4 (b) from *Tanaka* (2001). The contours show ionospheric convection patterns (in steps of 6 kV) in the northern hemisphere, from MHD modeling with uniform (a) and realistic (b) ionospheric conductance, and otherwise similar conditions ($B_y = \pm 2.5$ nT and $B_z = -4.3$ nT). Note that the colors show different quantities in the two figures (FACs [μAm^{-2}] in the upper figure, and conductance [mhos] in the lower figure). The numbered arrows refer to the text in Tanaka's paper.

and because the ionospheric conditions rarely are symmetrical between hemispheres, inter-hemispheric asymmetries are expected to arise in the magnetosphere. This is true, not only in the average, steady state magnetosphere, but also in the highly dynamical case.

5.3.3 Effects of breakdown of ideal MHD

Two effects which arise only when ideal MHD breaks down may also affect the geometry of the magnetosphere more directly than what has been discussed in the preceding section: Reconnection and parallel electric fields.

Reconnection

Terrestrial magnetic field lines which connect to the interplanetary magnetic field on the dayside of the magnetosphere are being transported tailward with the solar wind. Newly opened field lines will have a curvature which depends on the orientation of the IMF. In the southern hemisphere, the curvature will cause these field lines to be pulled in the same direction as the orientation of the IMF (in the y direction), and in the northern hemisphere, the field lines will be pulled the other way. The effect that this has on nightside reconnection is illustrated in Figure 5.7, where the solid open magnetic field lines, which have symmetrical ionospheric footpoints, do not meet in the tail. Instead, reconnection takes place between the dashed open field lines, which have asymmetrical footpoints. The result is equivalent to a net penetration of the IMF to closed field lines (*Østgaard et al.*, 2004).

Parallel electric fields

As discussed in Chapter 2.3.3, any electric field which appears in the frame of reference of the plasma will affect the time development of the magnetic field, in such a way that the frozen-in condition will break down. One way to look at the frozen-in property is to consider the field lines to be labeled by the plasma to which it is frozen. A temporal breakdown of the frozen-in condition can therefore be seen as a period when the magnetic field lines are "re-defined"; plasma which used to be attached to field lines mapping to one region of the ionosphere becomes frozen to field lines mapping to a different region in the ionosphere, without the field lines or the plasma convecting (field line motion may not even be a meaningful concept without some way to label them).

The effect of parallel electric fields was discussed quantitatively in Paper III, for the open/closed boundary, based on an approach by *Vasyliunas* (1984). We repeat the argument here, to emphasize that it holds for arbitrary closed field lines.

Faraday's law (Eq. 2.3) can be written:

$$\oint (\mathbf{E} + \mathbf{u} \times \mathbf{B}) \cdot d\mathbf{l} = -\frac{\partial \Phi}{\partial t} \tag{5.4}$$

where we used Equation 2.13 to transform the electric field in the frame of reference of the integration path to the frame of reference in which the path moves at velocity $\mathbf{u}$.

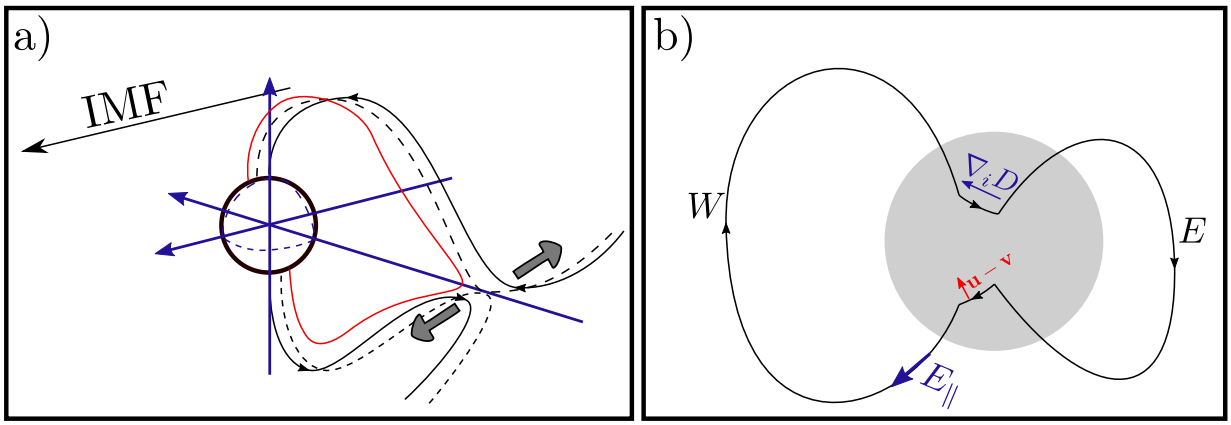

Figure 5.7: a) Reconnection of magnetic field lines with asymmetrical footpoints. The open solid field lines have symmetrical footpoints, but curvature forces (gray arrows) pull the southern hemisphere field lines in the direction of the IMF, and the northern hemisphere field line the other way, preventing them from reconnecting in the tail. Instead, reconnection takes place between the dashed field lines, which have asymmetrical footpoints (after *Østgaard et al.* (2004)). b) Illustration of the effect of gradients in net magnetic field aligned potential drops between hemispheres. For simplicity, we only consider a parallel electric field in the southern hemisphere in the western magnetic field line. This implies a westward gradient in net field aligned potential drop.

We choose the integration path sketched in Figure 5.7b, which we can divide into four parts: Two segments which run along magnetic field lines between hemispheres, and two segments connecting these field lines in the ionosphere. We place the ionospheric paths at sufficiently high altitude so that we can assume that $\mathbf{E} = -\mathbf{v} \times \mathbf{B}$, where $\mathbf{v}$ is the ionospheric convection (*Østgaard et al.*, 2005b). Integrating 5.4 along magnetic field lines yield the net electric potential drop along the field lines. We define $D = \int E_\parallel d\mathbf{l}$, with $E_\parallel$ positive in the direction parallel to the magnetic field. With this convention, the integrals along magnetic field lines in Figure 5.7b can be written $D_W - D_E$, where W and E indicate west and east, respectively. The different signs arise from the opposite directions of the integration paths. Now we choose the integration path such that it encloses a surface which is never threaded by magnetic field lines. Then, the right hand side of Equation 5.4 is constantly 0, and therefore vanishes. The result is

$$D_W - D_E + \int_{C_{iN}} (\mathbf{u} - \mathbf{v}) \times \mathbf{B} \cdot d\mathbf{l} + \int_{C_{iS}} (\mathbf{u} - \mathbf{v}) \times \mathbf{B} \cdot d\mathbf{l} = 0 \tag{5.5}$$

where C_{iS} and C_{iN} denote the ionospheric segments of the integration path in the southern and northern hemispheres, respectively. The $\mathbf{u} - \mathbf{v}$ terms denote plasma flow across the integration path. This flow is zero in ideal MHD, but Equation 5.5 shows that the presence of non-canceling parallel electric fields causes this to break down. A change in the $\mathbf{u}$'s,

relative to the **v**'s is equivalent to "re-defining" the flux tubes, by changing the plasma by which they are labeled. This equation can be used to address the magnitude of the plasma flow across magnetic field lines: We approximate the ionospheric integration paths by introducing average magnetic fields, assumed to point in the perpendicular direction, $B_\perp$, average **v** and **u**, positive in the equatorward direction, and perpendicular to the integration path, and the length of the integration path, L. $LB_\perp$ is, for simplicity, assumed to be equal in magnitude in the two hemispheres (although the signs are different, because of the different magnetic field directions in the two hemispheres). We also assume that the equatorward ionospheric convection is equal in the two hemispheres. Then Equation 5.5 reduces to:

$$u_N - u_S = \frac{D_E - D_W}{L|B_\perp|} = \frac{\nabla_i D}{|B_\perp|} \tag{5.6}$$

where ∇_i denotes the gradient along the ionospheric integration paths, positive in the eastward direction.

As illustrated in Figure 5.7b, a westward gradient in net potential drop is associated with an equatorward motion of the southern hemisphere footpoint relative to the northern hemisphere footpoint. *Sato et al.* (2005), who used ground all-sky imagers, reported a relative difference in the motions of conjugate magnetic field line footpoints between 43 and 66 m/s. According to Eq. 5.6, for a difference in u of 50 m/s, magnetic field strength of 50,000 nT, and a scale of 300 km (the field of view of the all-sky camera at 110 km is 1000 km), the corresponding difference in net field aligned potential drop along the western magnetic field lines is 750 V (assuming $D_E = 0$). With typical auroral arc electron energies of a few keV, this does not seem unreasonable.

We note that even though the plasma and magnetic field are not frozen-in at regions with parallel electric fields, the frozen-in condition holds in the surrounding regions, and the ideal-MHD conditions prevent plasma from entering these regions. The flow of plasma across field lines is confined to regions in which ideal MHD breaks down. This suggests that the effect of $\nabla_i D$ could be much less straightforward than what has been described above.

5.3.4 Causes for differences in auroral intensities

Differences in auroral energy spectra and power between hemispheres have been reported both statistically and from conjugate measurements. To gain a complete understanding of the cause for many of the reported differences, it is likely that we have to consider a long chain of events, involving global magnetospheric dynamics, ionospheric feedback, reconnection, and current sheets in the auroral acceleration region.

A simplified approach is therefore often used, in which the auroral precipitation is thought to reflect the field-aligned current (FAC) density. This is partly motivated by the Knight relation (*Knight*, 1973), $j_\parallel = KE_\parallel$, where $j_\parallel$ is the FAC density, $E_\parallel$ is the parallel electric field, and K is a proportionality constant, which depends on the electron density. While this is a useful relation, which may provide explanations to many reported auroral phenomena, it may also be misleading.

Ohtani et al. (2009) studied the relation between FAC density and precipitating particles in the pre midnight sector, using DMSP satellites. Contrary to what we would expect from a naive inspection of the Knight relation, the mean energy did not show a clear dependence on the current density. However, the precipitating electron energy flux was observed to increase with increasing FAC density. The latter finding lends some credence to the use of FACs to explain variations in auroral luminosity. A surprising conclusion from their study was that field-aligned currents (region 1 and region 2) in the pre-midnight sector were generally stronger in the dark hemisphere, compared to the sunlit hemisphere. They suggested that this difference is because the Pedersen conductance, which they calculated using their measurements of electron precipitation, and the *Robinson et al.* (1987) formula, is actually higher in the dark hemisphere than in the sunlit hemisphere. This is because of increased precipitation in the dark hemisphere, creating an "over-reaction" to the absence of sunlight induced conductance. These results demonstrate the complexity of explaining variations in auroral intensity.

In the following, we have grouped some of the proposed explanations for inter-hemispheric differences in auroral intensity into three mechanisms which are linked to FACs (these are necessarily simplifications), and one which is an effect of the magnetic field geometry.

Solar illumination

As described in Section 4.2.1, increasing parallel electric fields can be understood as a way of compensating for the inability of ambient electric charges to carry the current implied by $\nabla \times \mathbf{B}$ (Eq. 2.5). Intuitively, we would therefore expect parallel electric fields to be stronger in regions with fewer charges, i.e. in the dark hemisphere. This is in agreement with the observed seasonal differences on the nightside. A similar conclusion can be drawn from the Knight relation: with equal current densities, stronger electric fields are needed when the electron density is low.

Seasonal differences in the ionosphere also affect the formation of the feedback instability (*Lysak*, 1991). This instability, which may be responsible for small-scale structuring of auroral arcs, arises from the modification of the ionospheric conductivity which is due to precipitating particles.

Effects of the IMF

The IMF has also been suggested to cause intensity asymmetries more directly. For instance, the observations by *Shue et al.* (2002), that the intensity in the northern hemisphere is higher when $B_x < 0$ were suggested to be caused by a change in the prevailing convection pattern, modifying the flow shears and consequently FACs and auroral intensity. The change in convection was proposed as a direct consequence of the partial penetration of the B_x component, changing the geometry of the field (*Cowley*, 1981a).

Inter-hemispheric field aligned currents

As mentioned in Section 5.2.1, *Stenbaek-Nielsen and Otto* (1997) proposed that inter-hemispheric field aligned currents could explain the observations obtained by the all-sky cameras mounted on airplanes two decades earlier. This inter-hemispheric current (or rather, an inter-hemispheric component to the existing FAC) was suggested to be due to partial, but non-uniform penetration of the IMF B_y to the magnetosphere, which by Ampere's law implies currents. A similar explanation was suggested by *Liou et al.* (1998), to explain their observations of increased intensity in the northern hemisphere when IMF $B_y < 0$.

In Paper II, we observe large asymmetries in the auroral intensity, during times with concurrent negligible IMF B_y. During this event, however, the dipole tilt angle was large, so that the northern hemisphere was sunlit, and the southern hemisphere auroral zone was in darkness. These conditions are favorable for inter-hemispheric currents, according to a model by *Benkevich et al.* (2000). Their proposed inter-hemispheric FACs would go up (i.e., electrons going down) approximately at the locations where the aurora was most intense in the event studied in Paper II, making inter-hemispheric FACs a possible explanation for the observations. According to *Benkevich et al.* (2000), these currents arise in the following way: High latitude FACs (region 1) are stronger in the sunlit hemisphere, than they are in the dark hemisphere. The strong FACs close partly across the polar cap, and partly via the opposite hemisphere. The current crosses from one hemisphere to the other at regions with sharp gradients in the conductance, the conductance being calculated as the sum of the conductance in the two hemispheres at conjugate points. These gradients could be located at the sunlight terminator (in one of the hemispheres).

The *Benkevich et al.* (2000) model uses imposed FACs, with a realistic conductance pattern, current continuity, and Ohm's law. As we argued in Chapter 2, this technique presupposes a stable equilibrium. In the 12 May 2001 event (Papers II and III), the non-conjugate spots appeared during a very active time, a period of very strong imbalance between nightside and dayside reconnection. In the northern hemisphere, the non-conjugate spot was visible for only ≈ 10 minutes. The dynamic nature of the event, and the stable equilibrium required for the *Benkevich et al.* (2000) model poses a paradox. An alternative approach would be to treat the problem in terms of $\mathbf{v}$ and $\mathbf{B}$, and take into account large-scale dynamics. Using the observations of asymmetries in the poleward boundary of the aurora (assumed to coincide with the OCB), the appearance of inter-hemispheric current could be explained in the following (highly qualitative) way:

Figure 5.8 shows a flux tube connected to the ionosphere in both hemispheres at three different times. First, magnetospheric convection is excited, marked by the bold gray circle in the middle of the flux tube. For simplicity, the convection is depicted as circular, equivalent to a closed convection path in the magnetosphere. Since the ionosphere does not immediately catch up with this convection, it holds the magnetic field back, and the flux tube becomes twisted. Symmetrical (between hemispheres) currents then appear according to Ampere's law (a similar figure and explanation for FACs is given by *Paschmann et al.* (2003), p. 63). The magnetic stress associated with this twist excites ionospheric convec-

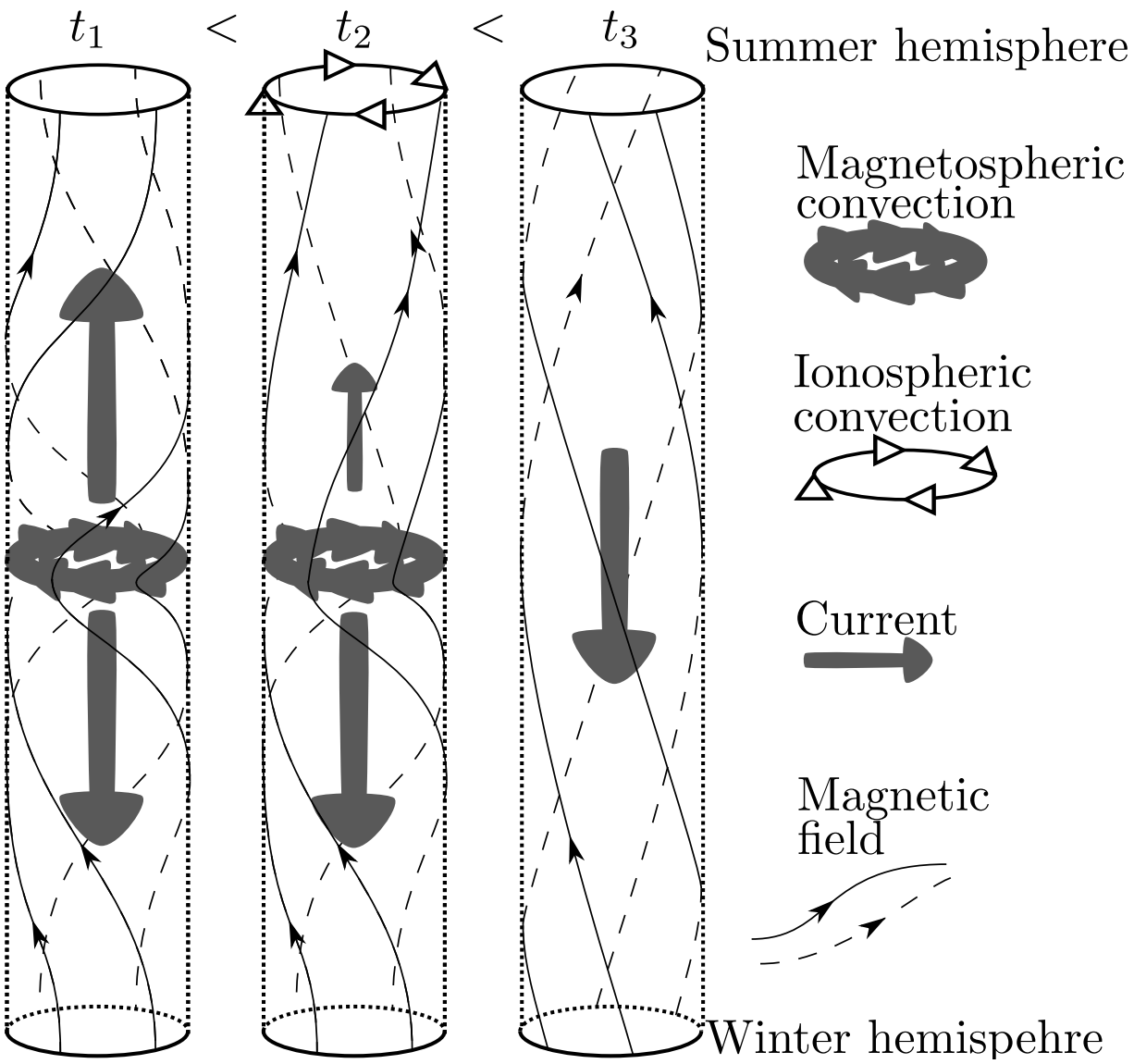

Figure 5.8: Cartoon showing how the asymmetrical ionospheric convection can set up a temporary inter-hemispheric current. We show magnetic flux tubes, extending from the winter hemisphere (in this case the southern hemisphere, since the magnetic field goes upwards) to the summer hemisphere. At $t = t_1$, convection is excited in the magnetosphere. Since the ionosphere does not immediately catch up with this convection, the flux tube is twisted, producing a symmetrical pair of currents. At $t = t_2$, the summer ionosphere has begun convecting, unwinding the twist, and reducing the current. At $t = t_3$, the driving of magnetospheric convection has been reduced, and the remaining twist of the magnetic field implies an inter-hemispheric current, which lasts until convection in the two hemispheres has unwound the twist. For simplicity, the figure shows only closed field lines, but the same mechanism may apply if part of the ionospheric loops connect to open field lines.

tion. At time t_2, the ionospheric convection has begun rectifying the twist in the summer hemisphere (in this case, the northern hemisphere), but not in the winter hemisphere. If the magnetospheric convection at some point is reduced, we are left with the situation seen at $t = t_3$, in which a twist remains in the flux tube connecting the two hemispheres. This implies an inter-hemispheric current. With convection differences similar to those suggested by the observations in Paper III and Paper IV, the direction of the current is from the sunlit hemisphere to the winter hemisphere at the dawnside of the onset local time (or the local time with the strongest earthward magnetospheric convection), and from the winter hemisphere to the summer hemisphere on the duskside. These locations are consistent with the observations of differences in auroral intensity (Paper II), and the timing is in qualitative agreement with the proposed difference in ionospheric convection following magnetotail reconnection (Paper III). This chain of events may also explain why the non-conjugate aurora appeared only for approximately 10 minutes at the dawn side, since the curl in the magnetic field represents magnetic stress which will have a finite duration.

It should be noted that the situation at $t = t_2$ in Figure 5.8 seems unrealistic, with stronger FACs flowing to the dark, and presumably (but not necessarily, according to *Ohtani et al.* (2009)) low-conductance, hemisphere. This happens because the ionospheric convection is depicted to take place first in the summer hemisphere, in accordance with the simplest representation of the ionospheric convection differences suggested in Papers III and IV. The net effect of this difference in convection is that the magnetic field geometry changes, such as to imply an inter-hemispheric current.

Low altitude magnetic field strength

As charged particles, spiraling around magnetic field lines, approach regions of converging magnetic field strength, they experience a net force (mirror force), repelling them in the opposite direction. This means that, at low altitudes in the auroral zone, particles either collide with the atmosphere, and may thereby contribute to the aurora, or they bounce back along the magnetic field lines. Since the magnetic field of the earth is not perfectly symmetrical between the hemispheres, the altitude at which they mirror can be different, and thus also the probability that particles collide with the atmosphere. Figure 5.9 shows the ratio of magnetic field strength at low altitude, in apex coordinates at 21:45 UT. From this figure, we see that in certain regions, the difference can be quite large, almost up to a factor of 2. The largest differences are found at 21 MLT in Figure 5.9, which in the southern hemisphere corresponds to the region south of South America, where the magnetic field has a minimum, called the South Atlantic anomaly.

Since accelerated auroral particles attain a highly field-aligned distribution, the resulting auroral intensity is not expected to be notably affected by the differences in mirror force. However, it may be important to the diffuse aurora, as reported by *Stenbaek-Nielsen et al.* (1973). These authors showed that, if the pitch angle distribution of the particles is isotropic, the ratio of the precipitating particle flux to the two hemispheres depends on the ratio of magnetic field strength in the two hemispheres. They also considered theoretically how parallel electric fields and weak diffusion may amplify or reduce this effect. In the case

of weak diffusion, the particles slowly fill the loss cone. If the bounce periods of the particles are shorter than the diffusion time, the particle precipitation could be much stronger to the hemisphere with the weakest magnetic field, even with relatively small differences between hemispheres.

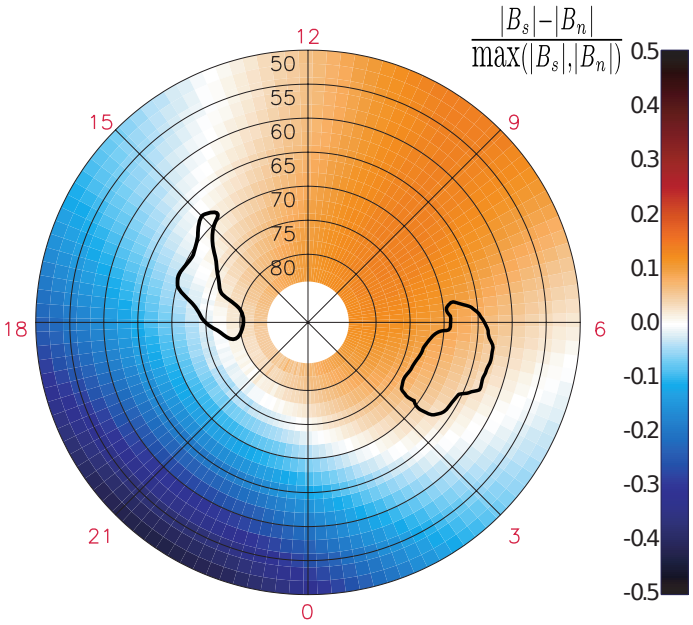

Figure 5.9: The ratio of low-altitude magnetic field strength at conjugate points in the two hemispheres. The ratio is plotted in magnetic coordinates, for 21:45 UT, at the time of the images in Figure 1 in Paper II. Contours of the non-conjugate aurora are also shown.

Also shown in Figure 5.9 are contours of the non-conjugate auroral spots reported in Paper II. These spots were seen in regions where the magnetic field strength differed by less than 10%. We can conclude that the differences in intensity were not caused by asymmetries in the mirror force in this case.

5.4 Future potential of inter-hemispheric measurements in the study of magnetosphere-ionosphere coupling

Ionospheric dynamics on high latitudes can be seen as the result of magnetospheric dynamics, and the various coupling mechanisms between the ionosphere and the magnetosphere (Figure 3.2). Since the ionosphere is much more accessible than the high altitude magnetosphere, most of our knowledge about magnetospheric dynamics is indirectly inferred from

ionospheric measurements, assuming the coupling processes to be known. This assumption may however be problematic during the most dynamical events, such as substorms. Since conjugate measurements in the ionosphere provide two independent pictures of the ionosphere, which correspond to the exact same magnetospheric state, such measurements have the potential to increase our understanding of the coupling processes.

The incoherent results of the studies of ionospheric convection during substorm expansion phase (Section 5.2.4) may be seen as an example of how the unknown coupling prevents a consistent picture of the magnetospheric state during certain periods. If the response time in the two hemispheres is different, as was discussed in Section 5.3.2 and in Papers III and IV, conjugate measurements of the convection could provide new insight to the nature of the coupling.

Several current projects provide conjugate measurements which can be used to continue and expand on studies of the magnetosphere-ionosphere coupling. SuperDARN[2], which consists of several radars measuring ionospheric convection, can be used to generate global instantaneous maps of the convection in both hemispheres (although the signal tends to disappear when auroral activity is high, excluding many of the perhaps most interesting events). SuperMAG[3] is an initiative to collect as many ground magnetometer measurements as possible in one place, using a common coordinate system, and a user friendly interface. With knowledge, or assumptions, on the conductance, these measurements can also be used to assess global convection in both hemispheres. The Assimilative Mapping of Ionospheric Electrodynamics (AMIE) technique is a technique to combine all relevant measurements to produce instantaneous maps of ionospheric parameters (conductance, currents, convection). The Iridium satellites, whose primary purpose is to facilitate satellite phone communication, are also equipped with magnetometers. Magnetic field measurements from the ~ 70 satellites have recently been used to generate maps of field-aligned currents in both hemispheres with relatively high time resolution[4].

The Chinese KuaFu project is a future initiatives to monitor the aurora in both hemispheres. It is thought to encompass two IMAGE like satellites and one solar wind monitor in orbit around the L1 point. The schedule for the KuaFu project, if it is realized, is not known[5].

[2]SuperDARN web page: http://superdarn.jhuapl.edu/

[3]SuperMAG web page: http://supermag.jhuapl.edu/

[4]FAC patterns derived from Iridium: http://sd-www.jhuapl.edu/UPOS/FAC/

[5]Information about KuaFu at Chinese Academy of Science web pages: http://english.cssar.cas.cn/ic/ip/200909/t20090917_38746.html

Chapter 6

Instrumentation and data processing

In this chapter we give an overview of the instruments which have been used to collect the data for the papers in this thesis. The key instruments are the FUV cameras on IMAGE, and the VIS Earth camera on Polar, and we therefore place particular emphasis on the quantitative interpretation of these measurements. We also mention the solar wind monitors that we have used, as well as low altitude satellites which provide in-situ measurements of particle precipitation.

Please consult page 86 for definitions of acronyms which are not given in this chapter.

6.1 Global UV imaging

The IMAGE satellite (*Burch*, 2000) was launched in March 2000. It carried several instruments designed to remotely image the magnetosphere, observing extreme ultraviolet (EUV), radio (RPI) and far ultraviolet (FUV) wavelengths, as well as neutral atoms (HENA/MENA). The FUV instrument consisted of three photometers (GEO), and three imagers, WIC, SI-12, and SI-13. In the next section, we look in more detail at the FUV instrument.

The Polar satellite was launched in February 1996. It carried twelve scientific instruments, three of which could be used for auroral imaging: The Polar Ionospheric X-ray Imaging Experiment (PIXIE) measured X-ray emissions, or bremsstrahlung, emitted as energetic electrons are decelerated in the atmosphere. The Ultraviolet Imager (UVI) provided UV images of the aurora, at different wavelengths. The Visible Imaging System (VIS) was designed to provide images of the aurora at visible wavelengths. This instrument also included a UV camera, VIS Earth Camera, which was intended as a monitor of the direction of the field of view of the visible wavelength cameras with respect to the sunlit Earth. However, the VIS Earth camera has produced a much more extensive scientific outcome than the visible wavelength cameras. The VIS Earth camera has also been used in the present thesis, preferred over UVI because of its large field of view (20°, compared to 8° in UVI), enabling observations of the global aurora from relatively low altitudes.

Table 6.1 shows an overview of the satellite orbits, and some parameters for the cameras

Satellite	IMAGE			Polar
Apogee			7.2 R_E	9 R_E
Perigee			1000 km	1.8 R_E
Orbit period			14.2 h	17.5 h
Life time			Mar. 2000 - Dec. 2005	Feb. 1996 - Apr. 2008
Camera	WIC	SI-13	SI-12	VIS Earth
Integration time	10 s	5 s	5 s	32.5 s
Cadence	123 s	123 s	123 s	54 s
Wavelength range	140 − 190 nm	135.6 nm	126.8 nm	124 − 149 nm
Resolution	256×256	128×128	128×128	256×256
Field of view	17° × 17°	16° × 16°	17° × 17°	20° × 20°
PSF FWHM	3 pixels	1.5 pixels	2 pixels	–

Table 6.1: Overview of the IMAGE and Polar satellites, and the UV cameras which were used in this thesis (*Mende et al.*, 2000a; *Frank et al.*, 1995). PSF FWHM is the full width at half maximum of the point spread function, taken from *Hubert et al.* (2002). Both IMAGE and Polar were NASA missions.

that we have used.

6.1.1 IMAGE FUV

The FUV instrument (*Mende et al.*, 2000a) encompasses three imagers: The Wideband Imaging Camera (WIC) (*Mende et al.*, 2000b) and the Spectrographic Imager (SI) (*Mende et al.*, 2000c), which produces two narrow pass-band images, centered at two wavelengths: 135.6 nm (SI-13) and 121.8 nm (SI-12). The WIC camera has a wide passband, observing aurora primarily in the range from 140 nm to 190 nm, which is dominated by N_2 emissions in the Lyman-Birge-Hopfield (LBH) band and a few strong N lines. The dynamic range and the resolution of WIC is higher than that of SI (*Frey et al.*, 2003a).

The narrow passbands of the SI channels are accomplished by an advanced system of slit grills, leading the light at the two wavelength regions to two different detectors. The 135.6 nm line, which is observed by SI-13, is emitted by atomic oxygen. Compared to the more intense OI line at 130.4 nm, the 135.6 nm line is scattered in the atmosphere only to a limited degree, and therefore the intensity in this camera is believed to be a good proxy for the total precipitation. The SI-12 camera has a peak sensitivity at ≈ 0.2 nm longer wavelengths than the Lyman-α emissions from hydrogen (121.56 nm). Blocking the Ly-α, the SI-12 signal is dominated by Doppler shifted emissions from receding hydrogen atoms. These atoms are primarily caused by precipitating protons, which charge exchange with the neutral atmosphere, and pick up an electron. Subsequently the hydrogen atom may be ionized again, and the process can be repeated hundreds to ten thousands of times, depending on the initial energy of the proton precipitation. The SI-12 transmission function is sinusoidal, with peaks every 0.5 nm, decreasing in amplitude away from 121.8

nm and reaching 0% transmission at $\lambda < 117$ nm and $\lambda > 126.5$ nm (*Immel et al.*, 2003).

The FUV cameras were mounted radially out from the spinning satellite, the spin period being 123 s. As the Earth swept through the field of view of the cameras every 123 s, the WIC camera, which had a CCD detector, took approximately 300 video frames. These frames were subsequently integrated to produce one image, using an on-board processing system called Time Delay Integration (TDI). A similar technique was applied to produce the SI images, which had a different detector (crossed delay-line) which recorded the position of every single photon, rather than instant 2D images.

Since the cameras observed the Earth from an altitude of several R_E, even very small inaccuracies in the pointing would lead to large errors in the interpretation of the images. To pin down the orientation of the cameras, we adjust the images so that the stars in the camera field of view (FOV) match the predicted (known) locations of the stars. With the software we have used[1], the pointing adjustment could be done with a one day resolution (i.e., the adjustment for one image is common to all images from the same day). During the course of the IMAGE mission, different pieces of the RPI instrument fell off[2], changing the satellite spin axis, causing it to wobble. This wobbling motion introduced a rhythmic (period of ~ 10 minutes) pointing error in the FUV images which could not be corrected using the star adjustment method. Data from the last part of the IMAGE mission should therefore be treated with some caution. For the 12 May 2001 event (Papers II and III) the wobbling is hardly noticeable.

In Section 6.1.4, we discuss the quantitative interpretation of images taken by the FUV cameras.

6.1.2 Polar VIS Earth

The VIS Earth Camera (*Frank et al.*, 1995) was mounted on a despun platform on the Polar satellite. It was equipped with a broad-band filter with a FWHM of 25 nm. The passband was $124 - 149$ nm, which includes the intense OI emissions at 130.4 nm and 135.6 nm, but also part of the LBH-band. For the periods when Polar and IMAGE produced conjugate images, the integration time of the camera was 32.5 seconds. The nominal cadence was 54 seconds, although when the other channels of the VIS instrument were in use, it was longer. Its sensitivity was 4.32 counts/kR.

The quantitative interpretation of VIS Earth images is discussed in Section 6.1.4.

6.1.3 Background removal

The images taken by VIS Earth and the FUV cameras have been corrected for varying sensitivity between pixels (flatfield correction), and for time dependent variations in sensitivity. This means that the intensity at different regions of the image, and images from different times, can be directly compared. The source of the measured intensity can then

[1] We have used Fuview, developed by the FUV team at Space Sciences Laboratory, Berkeley, USA

[2] according to the log file of FUV operations, at
http://sprg.ssl.berkeley.edu/image/wic_summary/0_fuv_operations.log

roughly be divided into three components: Auroral emissions, a relatively uniform (Poisson distributed) but slowly time-varying background, and sunlight induced emissions from the dayside ionosphere. The two latter components can represent a significant fraction of the total observed intensity, making auroral studies difficult, unless these emissions are subtracted from the image. In the following we describe two background/dayglow removal techniques which have been used in this thesis.

Geometric

In Paper I, we used geometric considerations to remove the dayglow in the SI-12 camera. The background proved to be fairly well modeled in two steps: First subtract a constant background, and then a crescent-shaped dayglow. The dayglow was modeled by assuming that the intensity in a given pixel depends on the area which is observed in the pixel, roughly proportional to $1/\cos\theta_{DZA}$, $\theta_{DZA} \leq 80°$, where θ_{DZA} is the satellite zenith angle (0° when viewing from nadir), and the inverse of the area over which the sunlight is distributed, which is approximately proportional to $1/\cos\theta_{SZA}$, where θ_{SZA} is the solar zenith angle. In a pixel with given $(\theta_{SZA}, \theta_{DZA})$, the intensity is then:

$$I(\theta_{SZA}, \theta_{DZA}) = I_0 \frac{\cos(\theta_{SZA})}{\cos(\theta_{DZA})}, \qquad \theta_{DZA} \leq 80° \tag{6.1}$$

The proportionality constant I_0 was determined by assuming that the pixels on the dayside, equatorward from the oval, were entirely due to sunlight. This constant was determined individually for each image, and the resulting background was then removed.

This method did not produce good results in WIC, because WIC is sensitive to dayglow scattered beyond the sunlight terminator. Another problem with this method is the assumption of linearity between the pixel intensity and the fraction on the right hand side. With more background pixel samples, more advanced relations can be considered. Figure 6.1a shows the intensity in background WIC pixels, plotted against $\cos\theta_{SZA}/\cos\theta_{DZA}$, from images taken during the substorm event shown in Figure 3.3. The background sample pixels were obtained from the region between 45° and 53°, and above 85° magnetic latitude, which was well outside the auroral oval. The red curve is a fitted function, in this case a constant for low values of the fraction (pixels which were in darkness, and dominated by the constant background), and a polynomial for higher values. The correlation between the observed intensity, and the intensity predicted by the fitted function is 0.997, which means that practically all variation in background intensity is captured by the functional fit. This method requires that the time variation in the background is slow. To include time variation in the model background, time could be included as a parameter, in addition to the angles.

Figure 6.1b shows this method applied on the WIC image from 13:29 UT on 23 June 2000. Panel b1 shows the original image, b2 shows the modeled background, and b3 shows the difference between the original image and the background. This is one of the auroral images which are shown in Figure 3.3. Since the background is roughly Poisson distributed, subtraction of a smooth background leaves significant noise in sunlit parts of the image.

This is clearly seen in Figure 6.1b3. Figure 6.1b4 shows a smoothed version of Figure 6.1b3.

Another variation of this method has also been used, in which the background is assumed to depend only on solar zenith angle. Then one can collect pixels which are outside the auroral oval, and interpolate to get the function $I_{BG} = f(\theta_{SZA})$. This method supposes that the background is independent of viewing angle, and is therefore less accurate. However, it is often seen to work quite well, and it is useful in double-checking the validity of other methods.

Polynomial fit

In Papers II and III, we fitted a 2D polynomial to the background and dayglow in the WIC images, and then subtracted this polynomial from the image. The method is illustrated in Figure 6.2. To fit the polynomial to the background, we had to define a region which was believed to be void of auroral emissions. This region was identified in two steps: First, we guessed at the location of the aurora, and fitted a polynomial to the remaining pixels. The fitting was done using the IDL routine, SFIT. Then we subtracted the resulting polynomial (Figure 6.2b) from the image (Figure 6.2a), thus removing most of the background. This corrected image (Figure 6.2c) was then used to identify the pixels containing aurora, using the IDL search routine SEARCH2D. The pixels which did not contain aurora, were assumed to contain background. These background pixels were then used in a second iteration of the polynomial fit. The resulting polynomial (Figure 6.2d) was subtracted from the original image, producing the final corrected images (Figure 6.2e). This method is inspired by the method used by *Hubert et al.* (2002) (personal communication).

While the polynomial fit effectively removed the background in the 12 May 2001 event, studied in Papers II and III, it did not prove to be universally applicable. To use this method, we had to carefully choose the input parameters, such as the order of the polynomial, the threshold in the SEARCH2D routine, and the initial guess at oval location. In many cases we did not succeed in getting a good polynomial fit to the background. The method seems more likely to succeed if the oval is completely embedded in sunlight, as in the 12 May 2001 event. It is less likely to succeed in imagers with lower resolution (e.g. the SI), since the fitted function will be based on fewer pixels. The same is true if the FOV does not cover much of the background. In the case of 12 May 2001, similar results were obtained using both the polynomial fit, and the geometric technique described in the previous section.

6.1.4 Quantitative interpretation

As the unit of the FUV images we have generally chosen instrument counts, rather than something physically more meaningful. Ideally, we would like to convert counts to energy flux (mW/m^2), so that the measurements can be compared with those from other instruments, or serve as input in models. However, without complementary measurements, which can assess the mean energy of the precipitating particles, such a conversion necessitates

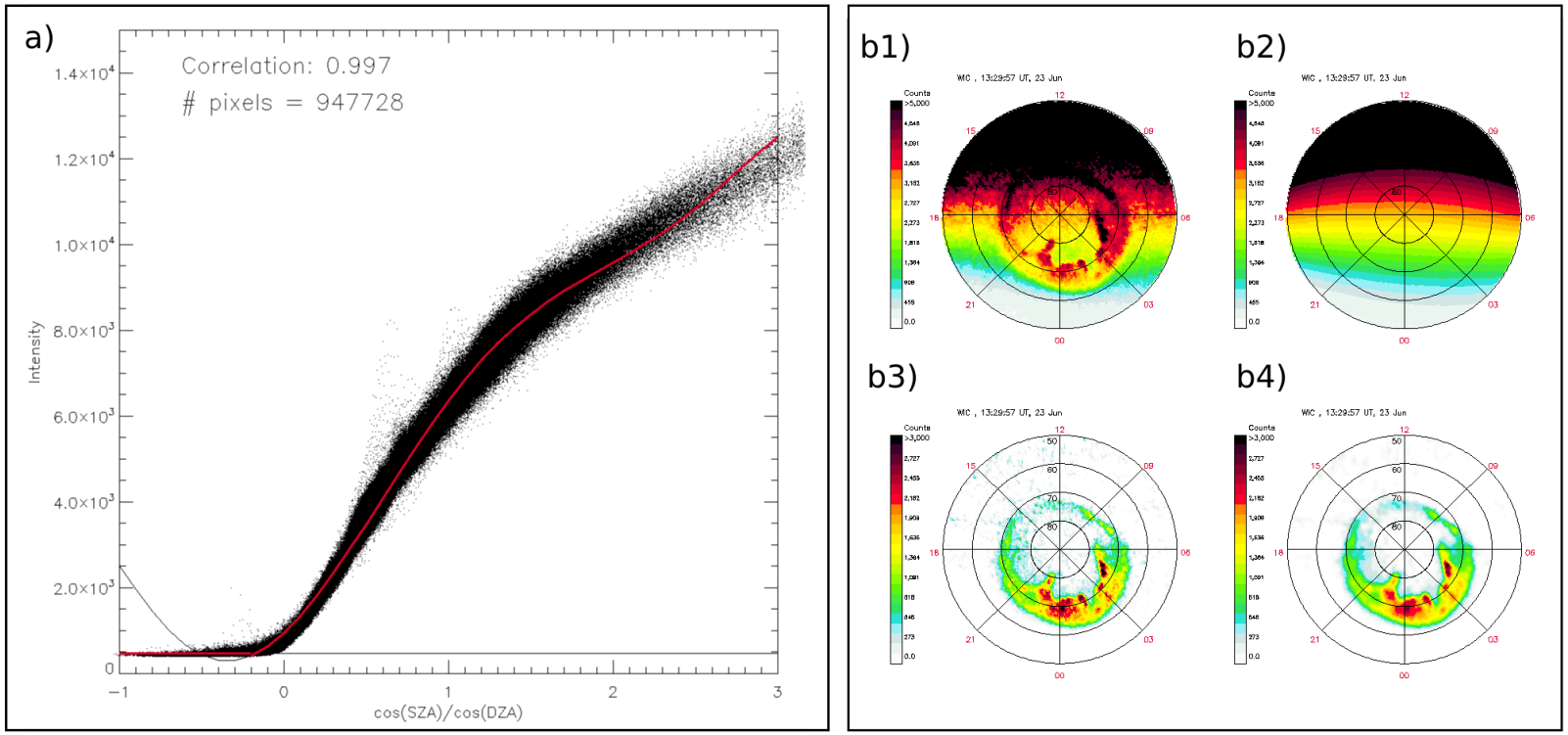

Figure 6.1: a) Intensity in background (non-aurora) pixels, plotted against $\frac{\cos\theta_{SZA}}{\cos\theta_{DZA}}$ (947728 pixels in total), for images taken on 23 June 2000, by IMAGE WIC. The correlation between the observed intensity, and the intensity according the functional fit, shown in red, is 0.997. b) The technique applied on the image taken at 13:29 UT at 23 June 2000. The figure shows the original image (b1), the modeled background (b2), the corrected image (b3), and a smoothed version of the corrected image (b4).

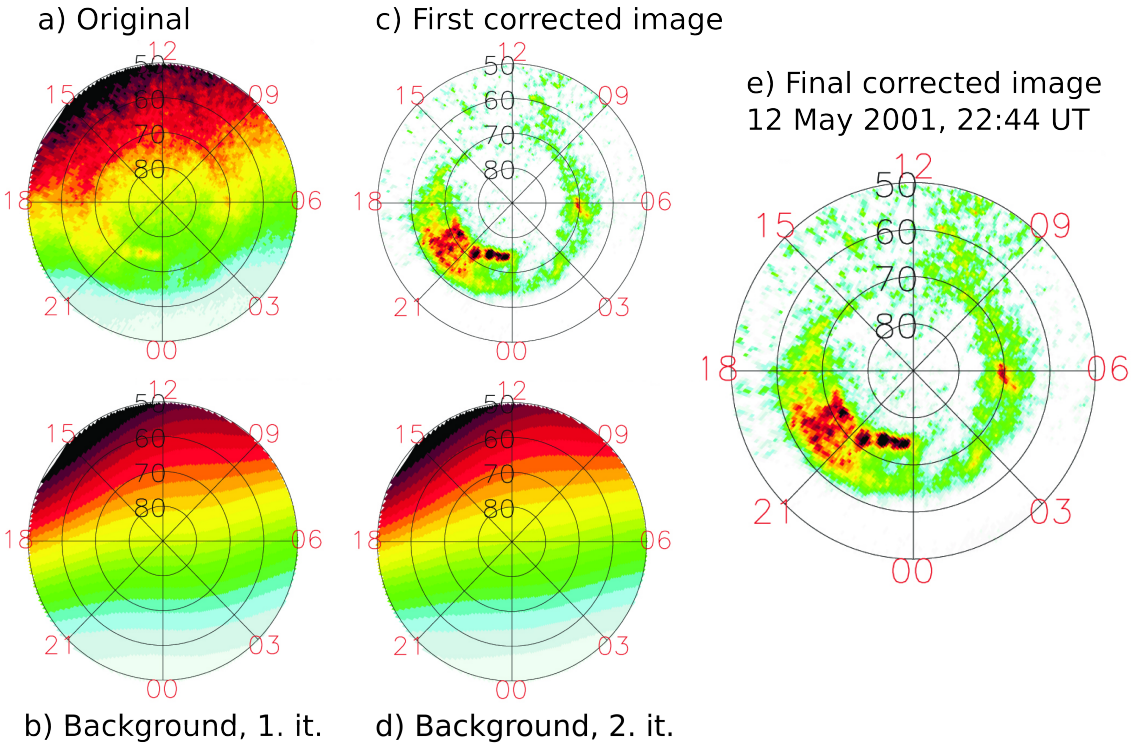

Figure 6.2: Example of polynomial fit method. a) Original image, b) fitted polynomial in first iteration, c) corrected image after first iteration, d) fitted polynomial in second iteration, e) final image. The color scales range from 0 to 2000 counts in the corrected images (c and e) and from 0 to 8000 in the original and the background images (a, b, and d)

crude assumptions, which almost certainly introduce errors. When energy flux is not required to answer a particular scientific question, we have chosen to process the data as little as possible, thus minimizing the black box between the instrument signal and the printed figures. The reduced black box does however come at the expense of an increased gap between the measurements and the real world. In this section we try to fill in part of this gap, by describing what the instrument counts correspond to in terms of energy flux.

Mathematically, the instrument counts in a given pixel at a given time is

$$I_{\text{counts}} = \int_0^\infty f(\lambda)g(\lambda)d\lambda \tag{6.2}$$

where λ is the wavelength, $g(\lambda)$ is the (known) passband of the camera, and $f(\lambda)$ is the spectrum of the photons reaching the camera. The latter function contains the atmospheric response to particle precipitation, and thus depends on several unknowns: The atmospheric composition, and the energy, mass, charge and direction of each of the precipitating particles. With the aid of atmospheric models and knowledge about (or assumptions on) the

distribution of the precipitating particles, it is possible to reduce the number of unknowns down to energy flux, so that Equation 6.2 relates energy flux and instrument counts.

We also note that the uncertainty in $f(\lambda)$ also affects the conversion from counts to kR. Since kR is a measure of the number of photons/sm^{2}sr, and since $g(\lambda)$ is not uniform, the same number of photons (same kR) does not always produce the same number of counts. In the conversion to kR used in Figure 1 of Paper II, we have used the value at the peak sensitivity of the WIC camera (*Frey et al.*, 2003a).

IMAGE FUV

Frey et al. (2003a) presented a comprehensive summary of some of the efforts (e.g., *Gérard et al.* (2000, 2001); *Hubert et al.* (2001, 2002)), that have been made to determine Equation 6.2 for the FUV instrument. The description given here relies heavily on this paper. The IMAGE FUV instrument package was designed so that energy flux could be determined with a fair accuracy without additional measurements. The procedure is as follows:

1. First, we have to measure, or assume a mean energy for the proton precipitation. When this is done, we use a modeled atmospheric response (assuming the protons have a kappa distribution) to find the energy flux associated with the observed intensity in SI-12. The modeled response is shown in Figure 6.3a.

2. A given proton energy flux, at the assumed mean proton energy, produces a known (also from the model) intensity in the SI-13 and WIC images. Figures 6.3b and c can be used to subtract the proton contribution, so that the SI-13 and WIC counts are solely due to electron precipitation.

3. After having subtracted the proton contribution, the WIC/SI-13 ratio depends on the mean electron energy (Figure 6.3f). An increasing mean electron energy increases the WIC/SI-13 ratio. This division produces a map of the mean electron energy.

4. Finally, Figure 6.3d (6.3e) can be used to find the energy flux which produces the observed WIC (SI-13) counts.

Application of the above technique requires that the images have been accurately corrected for background and dayglow. To perform the proton subtraction and the WIC/SI-13 division, it is also necessary that the pixels are mapped to the same grid. A slight error in background subtraction in SI-13, or in the pointing of one of the cameras with respect to the other, can produce dramatic results, since the SI-13 count rate is low compared to WIC. Another complication is that the atmospheric composition is assumed fixed in the atmospheric models used to derive the relations in Figure 6.3. It has been shown that the O/N_2 atmospheric density ratio is reduced during high geomagnetic activity, and the difference between the regions of O/N_2 depletion and the undisturbed regions can be more than a factor of two (*Strickland et al.*, 1998). Since WIC is sensitive to emissions from N_2, and SI-13 responds to emissions from atomic oxygen, the WIC/SI-13 ratio will increase in the O/N_2 depletion region, independent of the mean electron energy.

Even though the application of the method may be problematic, the modeled relations between counts and energy flux are informative. Figure 6.3a shows that the SI-12 camera efficiency increases very rapidly at low energies, just above its threshold at ≈ 1 keV. It reaches a maximum at $2-3$ keV, before it decreases rapidly. At high energies, the decrease is slow, and so we would expect SI-12 to respond to all energies of some significance in the magnetosphere (the ring current energy is is mainly carried by $10-200$ keV protons). These protons would also produce a significant signal in WIC and SI-13 (Figures 6.3b and 6.3c).

Figure 6.3d shows that, for constant energy flux, the WIC signal is rapidly reduced when the electron energy increases: At 20 keV the signal is ≈ 20 % of that at 1 keV. This trend is even more severe for SI-13 (Figure 6.3e) which is reduced to ≈ 10 % when the mean energy is increased from 1 to 20 keV.

Polar VIS Earth

We are not aware of any published study systematically addressing the response of the VIS Earth camera to electron energy flux, and its dependence on precipitation at different energies. However, Figure 6.4a shows an empirically determined relationship between observed VIS Earth intensity and electron energy flux (J. Sigwarth, personal communication). The upper and lower curves represent the 75th and 25th percentiles, respectively.

The VIS Earth response to varying electron energy can also be described using the information in *Frey et al.* (2003a). Polar VIS Earth observes emissions in the energy range of $124-149$ nm. This range encompasses the two intense atomic oxygen lines at 130.4 nm and 135.6 nm (which SI-13 observes), as well as some of the LBH band. In quiet conditions (pure dayglow emissions and very low-energy precipitation), *Frank and Sigwarth* (2003) report that 83 % of the VIS Earth signal is due to the 130.4 nm line. Figure 6.4b compares the intensity of the 130.4 nm and 135.6 nm lines for 1 $\mathrm{mW/m^2}$ electron flux at various mean energies. The numbers are based on Table 6.2, which is a reproduction of Table II in *Frey et al.* (2003a). The figure shows that, if the intensity in the VIS Earth camera is dominated by the 130.4 nm line, we can expect it to behave approximately like SI-13.

However, for higher energies, the relative importance of the various emission lines will change. According to Table 6.2, at low electron energies (0.2 keV) the 130.4 nm line represents 54% of the collective intensity escaping from the atmosphere from the LBH band and the two OI lines at 135.6 nm and 130.4 nm (in the nadir direction). At high energies (25.0 keV), the ratio is 10%. The relative importance of the 130.4 nm line in the VIS Earth signal thus decreases with increasing energy. We can therefore expect that as the energy increases, the relative importance of LBH emissions will increase in the VIS Earth signal. Based on this, the WIC/VIS ratio is expected to be approximately proportional to the WIC/SI-13 ratio (Figure 6.3f) at low energies, but to increase more slowly at higher energies.

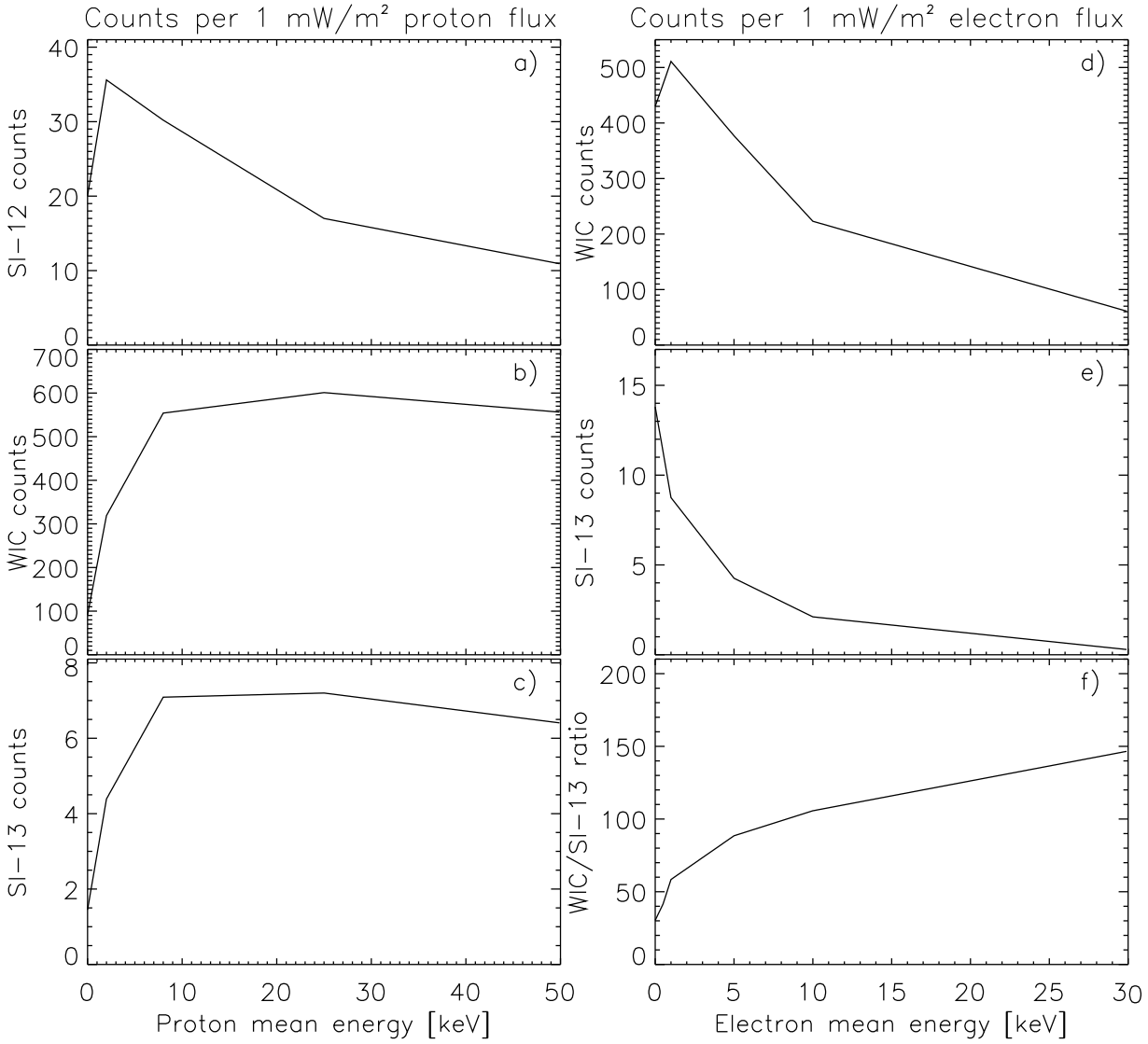

Figure 6.3: The response of SI-12 (a), WIC (b) and SI-13 (c), in corrected instrument counts, to a 1 mW/m^2 proton flux. Protons are assumed to be kappa distributed, with mean energy given at the x axis. The response of WIC (d) and SI-13 (e) to Maxwell distributed electron precipitation is also shown (mean energy at the x axis). Panel f shows the ratio of WIC and SI-13 counts. Figures a, b, c, d, and e are after Tables III, V, VIII, IV, and VII in *Frey et al.* (2003a), respectively.

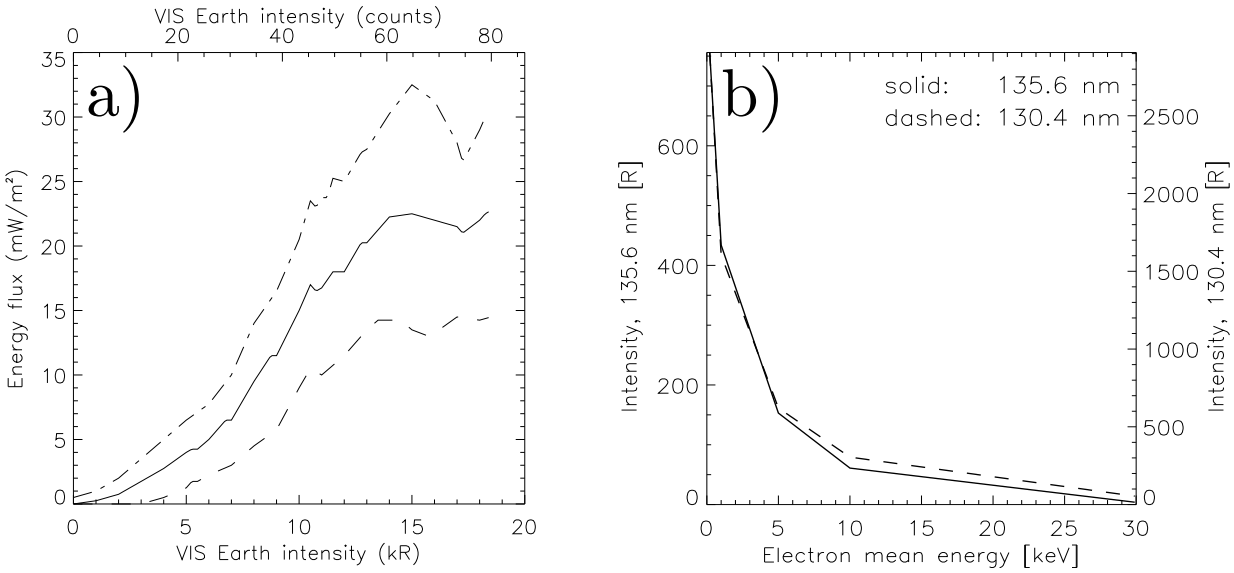

Figure 6.4: a) Empirically determined relationship between precipitating energy flux, and auroral intensity observed in the VIS Earth camera (J. Sigwarth, personal communication). b) Atmospheric response to electron precipitation. The solid curve shows intensity (in Rayleigh) of the 135.6 nm emission line (left y axis), and the dashed curve shows intensity of the 130.4 nm emission line (right y axis). The intensities are shown as function of mean electron energy. After Table 6.2

6.2 Solar wind monitors

Since high-latitude geomagnetic activity is mainly a consequence of interaction with the solar wind and the interplanetary magnetic field, it is essential to have measurements of the upstream solar wind conditions. In the papers in the present thesis, we have used ACE and Geotail for this purpose. Other spacecraft which are commonly located outside the magnetosphere is Wind and Cluster, which are still operational, and IMP-8, which was in operation from 1973 to 2001.

6.2.1 ACE

In all the papers in this thesis, we use measurements from the Advanced Composition Explorer (ACE) (*Stone et al.*, 1998). ACE was launched in August 1997, into an orbit about the Earth-Sun Lagrangian point, $\sim 240\ R_E$ sunward of the Earth. The Solar Wind Electron, Proton and Alpha Monitor (SWEPAM) (*McComas et al.*, 1998) is used to measure solar wind density and velocity. The solar wind dynamic pressure can then be calculated as $m_p n v^2$. Here, n is the proton density, v is the bulk velocity, and m_p is the proton mass. The formula should be adjusted if the fraction of heavier ions, mainly He^{2+}, becomes significant. The magnetic field is measured by a magnetometer, MAG (*Smith et al.*,

< E > [keV]	LBH	LBH abs.	1356	1356 abs.	1304	1304 abs.
0.20	1630	1629	757	756	2908	2907
0.50	1940	1910	638	635	2420	2420
1.00	2450	2390	440	434	1607	1606
5.00	3070	2320	194	153	630	621
10.0	3194	1738	124	61	321	305
25.0	3170	1010	85	18	132	116

Table 6.2: Atmospheric response to 1 mW/sm^2 electron precipitation, from Table II of *Frey et al.* (2003a). The table shows intensity [Rayleigh] of LBH, 135.6 nm and 130.4 nm emissions at different average electron energies, with and without atmospheric absorption.

1998).

Since the instruments on ACE record changes in the solar wind before these changes reach Earth, we need to time shift the measurements to relate solar wind parameters to observations in the magnetosphere, typically by $\sim$ 30 minutes to 1 h. We have employed two methods to do this time shift. 1) Variations in solar wind dynamic pressure often have clear effects on the magnetic field at ground, in near-Earth space, and on the auroral luminosity. ACE measurements can therefore be time shifted to match the solar wind variations with the observed effects in the magnetosphere. 2) The plane in which the solar wind has the least variation can be identified, using a statistical technique (*Weimer*, 2004). This plane represents the orientation of solar wind "phase fronts", which may be tilted with respect to the direction of the solar wind propagation. When the phase front orientation is known, this can be used to determine the distance which the measured solar wind variations travel before reaching the magnetosphere, and the time shift can be calculated as this distance, divided by the velocity. Time shift methods are discussed further in the papers.

6.2.2 Geotail

The Geotail satellite was launched in 1992, to an elliptical orbit with perigee and apogee at 8 R_E and 210 R_E, respectively. Its primary purpose was to investigate tail dynamics. In the 12 May 2001 event, which is studied in Papers II and III, it was located in the solar wind. We use measurements of the magnetic field, from the MGF instrument (*Kokubun et al.*, 1994). The advantage of using Geotail measurements in this event was that the time shift from the Geotail position to the magnetosphere was very small, or negligible.

6.3 Low altitude satellite measurements

Low altitude (< 1000 km) satellite measurements were used in Paper III to accurately determine the poleward boundary of the precipitation using in-situ particle measurements.

These measurements have also been useful in comparison with FUV images in the other studies, as an assurance that the camera pointing is correct.

6.3.1 DMSP

The Defense Meteorological Satellite Program (DMSP) comprises a series of satellites equipped with instruments measuring, among other things, the flux of charged particles, magnetic field, auroral emissions, and convection velocity. Their orbit is circular, at an altitude of ~ 840 km, with a period of approximately 100 minutes.

In Paper III, we used particle measurements from the Special Sensor Precipitating Plasma Monitor (SSJ) instrument, from the DMSP F12, F13, F14, and F15 satellites. The SSJ/4 instrument consists of four electrostatic analyzers that record the flux of precipitating electrons or ions in 19 logarithmically spaced energy intervals from 30 eV to 30 keV. The detected particles originate from within $\sim 3^\circ$ of the vertical, which means that the particles are well within the loss cone.

The data we have used have been downloaded from, or plotted at the Johns Hopkins University Applied Physics Laboratory web site[3].

6.3.2 NOAA POES

Another series of low-altitude satellites also carry instruments that measure the flux of precipitating particles: The National Oceanic and Atmospheric Administration (NOAA) Polar Orbiting Environmental Satellites (POES). We used measurements from the Total Energy Detector (TED) to determine the point at which the the satellites (NOAA-15 and NOAA-16) crossed the polar cap boundary. TED measures electrons and ions in 16 different energy intervals, from 50 eV to 20 keV. The quantities which are telemetered to the Earth is the integrated flux from these channels, as well as the flux in four different channels, and in the channel with the highest flux.[4]

The NOAA data was provided by NOAA's National Geophysical Data Center (NGDC).

6.4 Magnetic indices

Ground magnetometers have long been the primary method of measuring geomagnetic disturbances on Earth. Several magnetic indices exist, which facilitate studies of long-term trends, and comparisons between different events. Here we briefly describe the AE indices, and the SYM-H and Dst indices.

[3] http://seegar.jhuapl.edu/dmsp/

[4] Documentation on the NOAA TED can be found at
http://poes.ngdc.noaa.gov/docs/sem2_docs/2006/SEM2v2.0.pdf

6.4.1 AE

The AE (Auroral Electrojet) indices are composed of measurements from up to 12 stations distributed around the northern auroral zone at typical auroral latitudes. The AL index is defined as the minimum horizontal component measured by these stations (a baseline is subtracted first from each station). Since the horizontal component points northward in the northern hemisphere, AL is a measure of the maximum southward perturbation of the magnetic field in the auroral zone. This southward perturbation is often interpreted as an effect of westward ionospheric currents. AU is defined as the maximum horizontal component among the 12 stations, presumably associated with eastward currents. The AE index is the difference between AL and AU, and AO is the mean.

Substorms are often associated with a sharp drop in the AL index and a slow (~ 1 hour) recovery. However, if the substorm is located on different latitudes than the AE stations, its magnitude may be misrepresented, or it may not even be noticed in the indices.

6.4.2 SYM-H

The SYM-H index, which was used in Paper I, is derived from six magnetometers at low latitudes (although they are $\approx 20^\circ$ away from the magnetic equator). The index is constructed as a normalized average of the southward component measured at these stations. The SYM-H index can be seen as a high resolution (1 minute) version of the Dst index, which has existed since the 1950s. These indices are used to define the periods called geomagnetic storms, when SYM-H drops below a certain value (*Gonzalez et al.* (1994) define Dst < -30 nT to be a small storm, and Dst < -100 nT is an intense storm). As in substorms, storms are seen as a sharp drop, and a slow recovery in the magnetic indices, but in the case of storms, the time scales are longer: The main phase (sharp drop in SYM-H) lasts for typically a few hours, and the recovery can last for several days.

The SYM-H and Dst indices have a fairly clear physical meaning. The total kinetic energy of all charged particles trapped in a dipole magnetic field is proportional to the deflection of the magnetic field at the center of the dipole. This elegant relation was derived by *Dessler and Parker* (1959) and *Sckopke* (1966). Since SYM-H and Dst are derived from low-latitude magnetic field measurements, they are therefore often interpreted as a direct measure of ring current energy. However, in the real magnetosphere, other regions also contribute to the measurements comprising SYM-H and Dst, most notably the magnetopause, the magnetotail, the ionosphere, and induced currents in the ground.

Chapter 7

Summary of papers

This chapter contains brief summaries and corrections to the papers.

Paper I: Persistent global proton aurora caused by high solar wind dynamic pressure

In this paper, we report observations of a very clear response in the global proton aurora to changes in the solar wind dynamic pressure: Examples from five events with concurrent IMF $B_z > 0$ (low solar wind driving) show that the proton aurora, observed by the SI-12 camera on IMAGE, increases (decreases) as the solar wind dynamic pressure increases (decreases). This is true, even for low pressure and small changes (from 1 nPa to only 2 nPa in one event), and for both gradual and transient changes in the pressure.

An important distinction between our observations (and the similar observations of electron dominated aurora reported by *Liou et al.* (2007)) and previous studies of auroral response to changes in solar wind dynamic pressure (e.g., *Meurant et al.* (2004)), is that our observations show a persistent change in intensity when the pressure changes; the intensity seems to depend on the magnitude of the pressure, rather than (or in addition to) the rate of change of the pressure.

We also present observations from two events, classified as storm main phases. A storm main phase is a time of very high geomagnetic activity, and intense auroras are expected. In the two events presented, the drop in SYM-H, which is a way of quantifying the intensity of the main phase, was comparable, but the intensity of the proton aurora was very different. The dynamic pressure in the solar wind was also very different in these events, which might suggest that the solar wind dynamic pressure plays an important modulating role, also for the most intense proton auroras, with intense auroras occurring only when the geomagnetic activity and the solar wind pressure are high simultaneously.

Correction

In the abstract, it says that the time delay between changes in ground magnetic signatures and changes in the proton aurora intensity is short, "-2 minutes". This should be "$\lesssim$ 2 minutes".

Paper II: Asymmetric auroral intensities in the Earth's Northern and Southern hemispheres

In this paper, we show that the global auroral intensity distributions can be completely different in the two hemispheres. A pair of simultaneous images, taken by IMAGE WIC in the northern hemisphere and Polar VIS Earth in the southern hemisphere, showed a much higher intensity at dawn in the northern hemisphere, compared to dusk, and a much higher intensity at dusk in the southern hemisphere, compared to dawn.

The inter-hemispheric intensity asymmetries which are reported in this paper are important for at least two reasons: i) They confirm a proposed pattern of inter-hemispheric currents (assuming the currents go up where the electrons precipitate), which arises because of differences in the ionospheric conductance (*Benkevich et al.*, 2000). ii) Auroral intensifications in global images are often interpreted as "tv screen" images of activations of the magnetospheric regions to which they map. The observations show that such an interpretation could lead to very different conclusions, depending on which hemisphere we look at.

See Section 5.3.4 for a more detailed discussion of inter-hemispheric currents, and different causes for inter-hemispheric differences in auroral intensity.

Paper III: Inter-hemispheric observations of emerging polar cap asymmetries

In this paper, we take a new look at the same event as in Paper II, but change focus to inter-hemispheric asymmetries in the poleward boundary of the aurora. This boundary was visible at all local times for approximately 50 minutes, during the expansion phase of a strong substorm. We show that the poleward boundaries propagated poleward at different rates in the two hemispheres, leading to large inter-hemispheric asymmetries in the polar caps. We also show that the magnetic flux encircled by the boundaries in the two hemispheres was similar, in both the 12 May 2001 event and in another event, on 23 October 2002 (studied in more detail by *Stubbs et al.* (2005)). This is expected if the boundary coincides with the open/closed magnetic field line boundary.

From the argument discussed in Section 5.3.3, we know, with certain assumptions about parallel electric fields, that the open/closed magnetic field line boundary moves in response to either magnetic reconnection (the boundary is re-defined), or plasma convection (the boundary is physically moved) (*Cowley and Lockwood*, 1992). The emerging

inter-hemispheric asymmetries in the polar cap boundary which were observed in Paper III can therefore be explained if the ionospheric response to increased magnetospheric convection was different in the two hemispheres. Specifically, the emerging asymmetries were consistent with earthward convection in the magnetosphere, excited by tail reconnection (*Cowley and Lockwood*, 1992), being communicated faster to the northern summer hemisphere than to the southern winter hemisphere. A detailed investigation of the [18, 21] MLT region indicated a difference in equatorward convection of $\sim$ 500 m/s, lasting for $\sim$ 10 minutes.

This study shows that substorm signatures can be different in the two hemispheres, emphasizing the need for further parameterization in statistical studies of substorms.

Correction

In Equation 8 in the paper, positive ΔV corresponds to an electric field in the same direction as the magnetic field somewhere along the eastern integration paths (C_{p2} and C_{p4}). This is in contrast to the statement in parentheses preceding the equation.

Paper IV: Seasonal and IMF dependent nightside polar cap contraction during substorm expansion phase

Paper IV follows up on the finding from Paper III, that during the expansion phase of a substorm, the polar caps can contract differently in the two hemispheres. The aim of this paper is to investigate statistically the effect of different seasons, and different orientations of the IMF, on the location of the open/closed boundary during substorm expansion phase. The observational basis for this paper is images from the 3943 substorms identified by *Frey et al.* (2004) in the IMAGE WIC data set. These images were used to construct a large database of OCBs at different substorm epoch times.

Among the main results from this study was that the average OCB in substorms observed in the winter (dark) hemisphere has a more pronounced bulge, compared to summer hemisphere substorms. In the summer hemisphere, the OCB is more smooth. At substorm onset, the summer hemisphere OCB is, on average, closer to equator, compared to the winter hemisphere. During the expansion phase, this asymmetry reduces. One possible scenario leading to this asymmetry is an overall larger increase in equatorward convection in the winter ionosphere, except for in the bulge region, where the winter hemisphere ionospheric convection is more suppressed than in the summer hemisphere.

With the statistical data set developed for this study, we can investigate the average effect of the seasonal and IMF conditions in the 12 May 2001 event, studied in Papers II and III. During this event, $B_x > 0$, $B_y \approx 0$, and the tilt angle was quite strongly positive (summer in the north). Figure 7.1a compares the average boundaries for substorms in the northern hemisphere with $B_x > 3$ nT and tilt angle $> 15^\circ$ (dashed) with substorms with $B_x < -3$ nT and tilt angle less than -15° (solid). These groups resemble the conditions in the 12 May 2001 event in the northern and southern hemisphere, respectively. Figure 7.1b

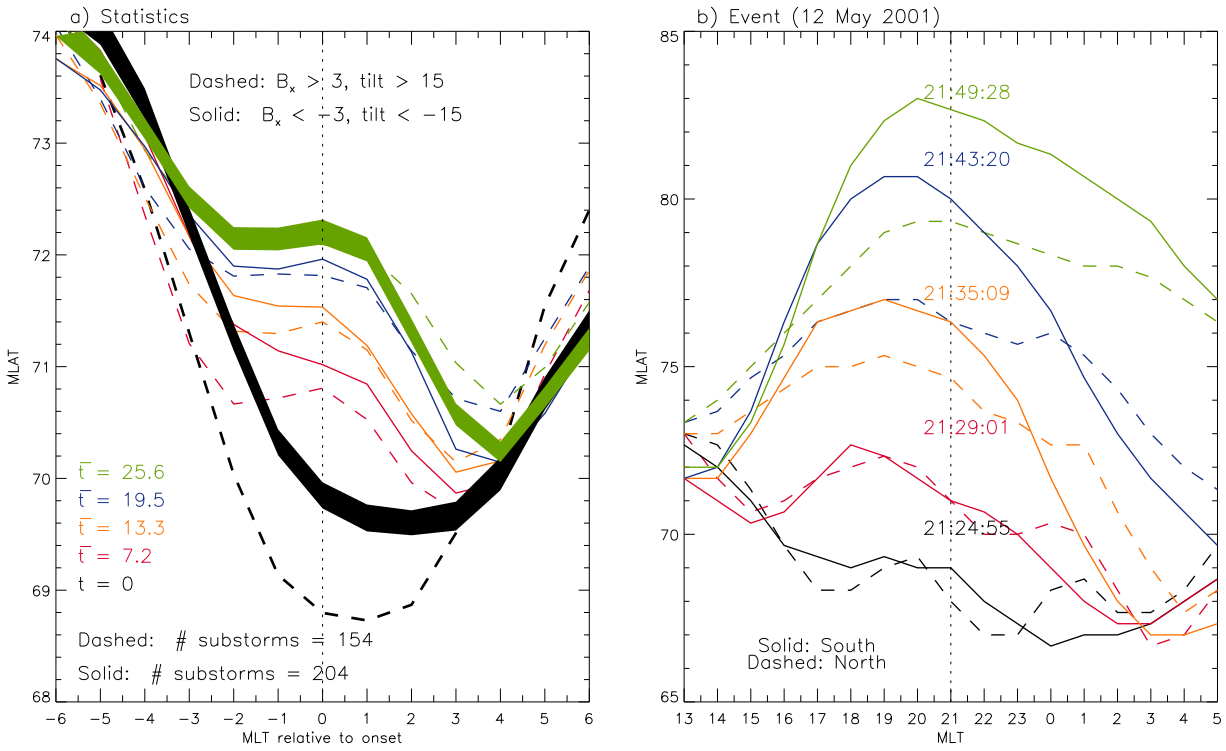

Figure 7.1: Comparison between the OCB in the 12 May 2001 event, discussed in Paper III (b), and the statistical average OCB for substorms with similar conditions (a). Note that the x and y axis have a much higher range for the single event boundaries. Figure a shows the average poleward boundary of the aurora at times relative to substorm onset, indicated by different colors. The x axis shows distance in MLT from the substorm onset location. The width of the black and green curves spans the standard error of mean. In contrast to the figures in Paper IV, these statistical distributions have been smoothed, using a boxcar average of width 3. Figure b is composed from simultaneous, or close to simultaneous, images taken in the two hemispheres at the times indicated in the figure (the times refer to the center of exposure in WIC). The dashed vertical bar in this figure denotes the approximate location of the onset (21 MLT).

shows the OCBs in the event, plotted in a similar format as the statistics. Dashed curves are from the northern hemisphere, and solid curves are from the southern hemisphere (to be compared with the corresponding line styles in Figure 7.1a).

A big difference between these plots is that the single event encompassed a larger dynamic range: The y axis spans 20°, compared to 6° in the statistics, and the x axis spans 16 hours, compared to 12 hours in the statistics. Further, there is only small, or no asymmetries at onset in the single event. During the expansion phase, a more pronounced bulge is formed in the winter hemisphere in the event, but this is not very prominent in the statistical plot. However, this feature corresponds well to what we observe when dividing substorms into summer and winter, independent of the orientation of the IMF (Figure 2 in Paper IV). In fact, the statistical plot and the case study look more similar for the reverse signs of B_x (this is not shown). Based on this, and on the conclusions from Paper IV, we conclude that the differences in the case study were due to seasonal effects, and that B_x may actually have had a dampening effect on the asymmetries in the May 2001 event. We can also conclude that instantaneous asymmetries between hemispheres can be much stronger than what is suggested by the modest (but statistically significant) asymmetries in the statistical study.

LoA List of Acronyms

AACGM	Altitude Adjusted Corrected Geomagnetic
ACE	Advanced Composition Explorer
AE	Auroral Electrojet
AMIE	Assimilative Mapping of Ionospheric Electrodynamics
ASC	All-Sky Camera
CCD	Charge-coupled device
DMSP	Defense Meteorological Satellite Program
EDI	Electron Drift Instrument
ENA	Energetic Neutral Atom
EUV	Extreme Ultraviolet
IGRF	International Geomagnetic Reference Field
IMAGE	Imager for Magnetopause-to-Auroral Global Exploration
IMP	Interplanetary Monitoring Platform
FAC	Field-aligned current
FOV	Field of view
FUV	Far Ultraviolet
FWHM	Full Width at Half Maximum
GEO	Geocorona Photometer
GSM	Geocentric Solar Magnetospheric Coordinate System
GUVI	Global UltraViolet Imager
HENA	High Energy Neutral Atom Imager
IMF	Interplanetary Magnetic Field
LBH	Lyman-Birge-Hopfield
LLBL	Low Latitude Boundary Layer
MENA	Medium Energy Neutral Atom Imager
MHD	MagnetoHydroDynamics
MLT	Magnetic Local Time
NENL	Near-Earth Neutral Line
NOAA	National Oceanic and Atmospheric Administration
OCB	Open/Closed Boundary
PIXIE	Polar Ionospheric X-ray Imaging Experiment
POES	Polar Orbiting Environmental Satellites
Polar BEAR	Polar Beacon Experiment and Auroral Research (satellite)
PSBL	Plasma Sheet Boundary Layer
PSF	Point Spread Function
R_E	Earth Radius (approximately 6370 km)
RPI	Radio Plasma Imager
SI	Spectrographic Imager
SuperDARN	Super Dual Auroral Radar Network
SWEPAM	Solar Wind Electron, Proton and Alpha Monitor
TED	Total Energy Detector
TIMED	Thermosphere Ionosphere Mesosphere Energetics and Dynamics (satellite)
UT	Universal Time
UV	UltraViolet
UVI	UltraViolet Imager
VIS	Visible Imaging System
WIC	Wideband Imaging Camera

Bibliography

Akasofu, S.-I., The development of the auroral substorm, *Planet. Space Sci.*, *12*, 273–282, 1964.

Ambrosino, D., E. Amata, M. F. Marcucci, I. Coco, W. Bristow, and P. Dyson, Different responses of northern and southern high latitude ionospheric convection to IMF rotations: A case study based on SuperDARN observations, *Ann. Geophys.*, *27*, 2423–2438, 2009.

Anderson, K. A., C. D. Anger, R. R. Brown, and D. S. Evans, Simultaneous electron precipitation in the northern and southern auroral zones, *J. Geophys. Res.*, *67*, 4076–4077, 1962.

Angelopoulos, V., et al., Statistical characteristics of bursty bulk flow events, *J. Geophys. Res.*, *99*, 21,257–21,280, 1994.

Atkinson, G., and D. Hutchison, Effect of the day night ionospheric conductivity gradient on polar cap convective flow, *J. Geophys. Res.*, *83*, 725–729, 1978.

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

Baker, K. B., and S. Wing, A new magnetic coordinate system for conjugate studies at high latitudes, *J. Geophys. Res.*, *94*, 9139–9143, 1989.

Baumjohann, W., and R. A. Treumann, *Basic space plasma physics*, Imperial College Press, 1997.

Belon, A. E., J. E. Maggs, T. N. Davis, K. B. Mather, N. W. Glass, and G. F. Hughes, Conjugacy of visual auroras during magnetically quiet periods, *J. Geophys. Res.*, *74*, 1–27, 1969.

Benkevich, L., W. Lyatsky, and L. L. Cogger, Field-aligned currents between conjugate hemispheres, *J. Geophys. Res.*, *105*, 27,727–27,737, 2000.

Blanchard, G. T., L. R. Lyons, and O. de la Beaujardire, Magnetotail reconnection rate during magnetospheic substorms, *J. Geophys. Res.*, *102*, 14,303–24,312, 1997.

Boakes, P. D., S. E. Milan, G. A. Abel, M. P. Freeman, G. Chisham, B. Hubert, and T. Sotirelis, On the use of IMAGE FUV for estimating the latitude of the open/closed magnetic field line boundary in the ionosphere, *Ann. Geophys.*, pp. 2759 – 2769, 2008.

Boakes, P. D., S. E. Milan, G. A. Abel, M. P. Freeman, G. Chisham, and B. Hubert, A statistical study of the open magnetic flux content of the magnetosphere at the time of substorm onset, *Geophys. Res. Lett.*, doi:10.1029/2008GL037059, 2009.

Bond, F. R., Auroral morphological similarities at two magnetically conjugate stations: Buckles Bay and Kotzebue, *Aust. J. Phys.*, *22*, 421–433, 1969.

Bristow, W. A., and P. Jensen, A superposed epoch study of SuperDARN convection observations during substorms, *J. Geophys. Res.*, *112*, doi:10.1029/2006JA012049, 2007.

Burch, J. L., IMAGE Mission Overview, *Space Sci. Rev.*, *91*, 1–14, 2000.

Burch, J. L., J. Goldstein, and B. R. Sandel, Cause of plasmasphere corotation lag, *Geophys. Res. Lett.*, *31*, doi:10.1029/2003GL019164, 2004.

Burns, G. B., D. J. McEwen, R. A. Eather, F. T. Berkey, and J. S. Murphree, Optical auroral conjugacy: Viking UV imager - South Pole Station ground data, *J. Geophys. Res.*, *95*, 5781–5790, 1990.

Carbary, J. F., T. Sotirelis, P. T. Newell, and C.-I. Meng, Auroral boundary correlations between UVI and DMSP, *J. Geophys. Res.*, *108*, doi:10.1029/2002JA009378, 2003.

Chaston, C. C., J. W. Bonnell, C. W. Carlson, J. P. McFadden, R. E. Ergun, and R. J. Strangeway, Properties of small-scale Alfven waves and accelerated electrons from FAST, *J. Geophys. Res.*, *108*, doi:10.1029/2002JA009420, 2003.

Coroniti, F. V., and C. F. Kennel, Can the ionosphere regulate magnetospheric convection?, *J. Geophys. Res.*, *78*, 2837–2851, 1973.

Coumans, V., J.-C. Gérard, B. Hubert, S. B. Mende, and S. W. H. Cowley, Morphology and seasonal variations of global auroral proton precipitation observed by IMAGE-FUV, *J. Geophys. Res.*, *109*, doi:10.1029/2003JA010348, 2004.

Coumans, V., J.-C. Gérard, B. Hubert, and M. Meurant, Global auroral proton precipitation observed by IMAGE-FUV: Noon and midnight brightness dependence on solar wind characteristics and IMF orientation, *J. Geophys. Res.*, *111*, doi:10.1029/2005JA011317, 2006.

Cowley, S. W. H., Asymmetry effects associated with the x-component of the IMF in a magnetically open magnetosphere, *Planet. Space Sci.*, *29*(8), 809–818, 1981a.

Cowley, S. W. H., Magnetospheric asymmetries associated with the y-component of the IMF, *Planet. Space Sci.*, *29*(1), 79–96, 1981b.

Cowley, S. W. H., and M. Lockwood, Excitation and decay of solar wind-driven flows in the magnetosphere-ionosphere system, *Ann. Geophys.*, pp. 103–115, 1992.

Cowley, S. W. H., J. P. Morelli, and M. Lockwood, Dependence of convective flows and particle precipitation in the high-latitude dayside ionosphere on the x and y components of the interplanetary magnetic field, *J. Geophys. Res.*, *96*, 5557–5564, 1991.

Craven, J. D., J. S. M. L. A. Frank, and L. L. Cogger, Simultaneous optical observations of transpolar arcs in the two polar caps, *Geophys. Res. Lett.*, *18*, 2297 – 2300, 1991.

Crooker, N. U., Reverse convection, *J. Geophys. Res.*, pp. 19,363–19,372, 1992.

Dessler, A. J., and E. N. Parker, Hydromagnetic theory of geomagnetic storms, *J. Geophys. Res.*, *64*, 2239–2252, 1959.

DeWitt, R. N., The occurrence of aurora in geomagnetically conjugate areas, *J. Geophys. Res.*, *62*, 1347–1352, 1962.

Dungey, J. W., Interplanetary magnetic field and the auroral zones, *Phys. Rev. Lett.*, *6*, 47–48, 1961.

Elphinstone, R. D., K. Jankowska, J. S. Murphree, and L. L. Cogger, The configuration of the auroral distribution for interplanetary magnetic field B_z northward 1. IMF B_x and B_y dependencies as observed by the Viking satellite, *J. Geophys. Res.*, *95*, 5791–5804, 1990.

Evans, D. S., Precipitating electron fluxes formed by a magnetic field aligned potential difference, *J. Geophys. Res.*, *79*, 2853–2858, 1973.

Fairfield, D. H., On the average configuration of the geomagnetic tail, *J. Geophys. Res.*, *84*, 1950–1958, 1979.

Fillingim, M. O., G. K. Parks, H. U. Frey, T. J. Immel, and S. B. Mende, Hemispheric asymmetry of the afternoon electron aurora, *Geophys. Res. Lett.*, *32*, doi: 10.1029/2004GL021635, 2005.

Fillingim, M. O., J. P. Eastwood, G. K. Parks, V. Angelopoulos, I. R. Mann, S. B. Mende, and A. T. Weatherwax, Polar UVI and THEMIS GMAG observations of the ionospheric response to a hot flow anomaly, *J. Atmos. Terr. Phys.*, doi: doi:10.1016/j.jastp.2010.03.001, 2010.

Frank, L. A., and J. B. Sigwarth, Simultaneous images of the northern and southern auroras from the Polar spacecraft: An auroral substorm, *J. Geophys. Res.*, *108*, doi: 10.1029/2002JA009356, 2003.

Frank, L. A., J. B. Sigwarth, J. D. Craven, J. P. Cravens, J. S. Dolan, M. R. Dvorsky, P. K. Hardebeck, J. D. Harvey, and D. W. Muller, The visible imaging system VIS for the Polar spacecraft, *Space Sci. Rev.*, *71*, 297–328, 1995.

Frey, H. U., Localized aurora beyond the auroral oval, *Reviews of Geophysics*, *45*, 2007.

Frey, H. U., and S. B. Mende, Substorm onsets as observed by IMAGE-FUV, in *Substorms VIII: Proc. of the 8th International Conference on Substorms*, pp. 71–76, Univ. of Calgary Press, Calgary, 2006.

Frey, H. U., S. B. Mende, H. B. Vo, M. Brittnacher, and G. K. Parks, Conjugate observation of optical aurora with Polar satellite and ground-based cameras, *Advances in Space Research*, *23*, 1647–1652, 1999.

Frey, H. U., S. B. Mende, C. W. Carlson, J.-C. Gérard, B. . Hubert, J. Spann, R. Gladstone, and T. J. Immel, The electron and proton aurora as seen by IMAGE-FUV and FAST, *Geophys. Res. Lett.*, *28*, 1135–1138, 2001.

Frey, H. U., T. D. Phan, S. A. Fuselier, and S. B. Mende, Continuous magnetic reconnection at Earth's magnetopause, *Nature*, *426*, 533–537, 2003a.

Frey, H. U., S. B. Mende, V. Angelopoulos, and E. F. Donovon, Substorm onset observations by IMAGE-FUV, *J. Geophys. Res.*, *109*, doi:10.1029/2004JA010607, 2004.

Frey, H. U., et al., Summary of quantitative interpretation of IMAGE far ultraviolet auroral data, *Space Sci. Rev.*, *109*, 255–283, 2003b.

Fujii, R., N. Sato, T. Ono, H. Fukunishi, T. Hirasawa, S. Kokubun, T. Araki, and T. Saemundsson, Conjugacies of pulsating auroras by all-sky TV observations, *Geophys. Res. Lett.*, *14*, 115–118, 1987.

Galand, M., and D. Lummerzheim, Contribution of proton precipitation to space-based auroral FUV observations, *J. Geophys. Res.*, *109*, doi:10.1029/2003JA010321, 2004.

Gérard, J., B. Hubert, M. Meurant, V. I. Shematovich, D. V. Bisikalo, H. Frey, S. Mende, G. R. Gladstone, and C. W. Carlson, Observation of the proton aurora with IMAGE FUV imager and simultaneous ion flux in situ measurments, *J. Geophys. Res.*, *106*, 28,939–28,948, 2001.

Gérard, J.-C., B. Hubert, D. V. Bisikalo, and V. I. Shematovich, A model of the Lyman-α line profile in the proton aurora, *J. Geophys. Res.*, *105*, 15,795–15,805, 2000.

Gjerloev, J. W., R. A. Hoffman, J. B. Sigwarth, L. A. Frank, and J. B. H. Baker, Typical auroral substorm: A bifurcated oval, *J. Geophys. Res.*, *113*, doi:10.1029/2007JA012431, 2008.

Gonzalez, W., J. Joselyn, Y. Kamide, H. Kroehl, G. Rostoker, B. Tsurutani, and V. Vasyliunas, What is a geomagnetic storm?, *J. Geophys. Res.*, *99*, 5771–5792, 1994.

Griffiths, D. J., *Introduction to electrodynamics*, 3.rd ed., Prentice Hall, 1981.

Grocott, A., T. K. Yeoman, S. E. Milan, and S. W. H. Cowley, Interhemispheric observations of the ionospheric signature of tail reconnection during IMF-northward non-substorm intervals, *Ann. Geophys.*, *23*, 1763–1770, 2005.

Grocott, A., J. A. Wild, S. E. Milan, and T. K. Yeoman, Superposed epoch analysis of the ionospheric convection evolution during substorms: Onset latitude dependence, *Ann. Geophys.*, *27*, 591–600, 2009.

Haaland, S. E., G. Paschmann, M. Forster, J. M. Quinn, R. B. Torbert, C. E. McIlwain, H. Vaith, P. A. Puhl-Quinn, and C. A. Kletzing, High-latitude plasma convection from cluster EDI measurements: Method and IMF-dependence, *Ann. Geophys.*, pp. 239–253, 2007.

Hargreaves, J. K., Conjugate and closely-spaced observations of auroral radio absorption - I, *Planet. Space Sci.*, *17*, 1459–1484, 1969.

Heppner, J. P., and N. C. Maynard, Empirical high-latitude electric field models, *J. Geophys. Res.*, *92*, 4467–4489, 1987.

Holzworth, R. H., and C.-I. Meng, Mathematical representation of the auroral oval, *Geophys. Res. Lett.*, *2*, 377–380, 1975.

Holzworth, R. H., and C.-I. Meng, Auroral boundary variations and the interplanetary magnetic field, *Planet. Space Sci.*, *32*, 25–29, 1984.

Hubert, B., J.-C. Gérard, D. V. Bisikalo, V. I. Shemantovich, and S. C. Solomon, The role of proton precipitation in the excitation of auroral FUV emissions, *J. Geophys. Res.*, *106*, 21,475–21,494, 2001.

Hubert, B., J.-C. Gérard, D. S. Evans, M. Meurant, S. B. Mende, H. U. Frey, and T. J. Immel, Total electron and proton energy input during auroral substorms: Remote sensing with IMAGE-FUV, *J. Geophys. Res.*, *107*, doi:10.1029/2001JA009229, 2002.

Hubert, B., S. E. Milan, A. Grocott, C. Blockx, S. W. H. Cowley, and J. C. Gérard, Dayside and nightside reconnection rates inferred from IMAGE-FUV and Super Dual Auroral Radar Network data, *J. Geophys. Res.*, *111*, doi:10.1029/2005JA011140, 2006.

Imber, S. M., S. E. Milan, and B. Hubert, Observations of significant flux closure by dual lobe reconnection, *Ann. Geophys.*, *25*, 1617–1627, 2006.

Immel, T., S. Mende, H. Frey, N. Østgaard, and G. Gladstone, Effects of the 14 July 2000 solar flare on Earth's FUV emissions, *J. Geophys. Res.*, *108*, doi:10.1029/2001JA009060, 2003.

Jørgensen, T. S., E. Friis-Christensen, and J. Wilhjelm, Interplanetary magnetic-field directions and high-latitude ionospheric currents, *J. Geophys. Res.*, *77*, 1976–1977, 1972.

Khurana, K. K., R. J. Walker, and T. Ogino, Magnetospheric convection in the presence of interplanetary magnetic field B_y: A conceptual model and simulations, *J. Geophys. Res.*, *101*, 4907–4916, 1996.

Kivelson, M. G., and C. T. Russell, *Introduction to space physics*, Cambridge University Press, Cambridge, UK, 1995.

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

Kokubun, S., T. Yamamoto, M. H. Acuna, K. Hayashi, K. Shiokawa, and H. Kawano, The geotail magnetic field experiment, *Journal of Geomagnetism and Geoelectricity*, *46*, 7–21, 1994.

Liou, K., P. T. Newell, C.-I. Meng, M. Brittnacher, and G. Parks, Characteristics of the solar wind controlled auroral emissions, *J. Geophys. Res.*, *103*, 17,543–17,557, 1998.

Liou, K., P. T. Newell, and C.-I. Meng, Seasonal effects on auroral particle acceleration and precipitation, *J. Geophys. Res.*, *106*, 5531–5542, 2001.

Liou, K., C.-I. Meng, A. T. Y. Lui, P. T. Newell, and S. Wing, Magnetic dipolarization with substorm expansion onset, *J. Geophys. Res.*, *107*, doi:10.1029/2001JA000179, 2002.

Liou, K., P. T. Newell, J.-H. Shue, C.-I. Meng, Y. Miyashita, H. Kojima, and H. Matsumoto, "Compression aurora": Particle precipitation driven by long-duration high solar wind ram pressure, *J. Geophys. Res.*, *112*, doi:10.1029/2007JA012443, 2007.

Lockwood, M., S. W. H. Cowley, and M. P. Freeman, The excitation of plasma convection in the high-latitude ionosphere, *J. Geophys. Res.*, *95*, 7961–7972, 1990.

Lu, G., et al., Interhemispheric asymmetry of the high-latitude ionospheric convection pattern, *J. Geophys. Res.*, *99*, 6491–6510, 1994.

Lysak, R. L., Feedback instability of the ionospheric resonant cavity, *J. Geophys. Res.*, *96*, 1553–1568, 1991.

Mansurov, S. M., New evidence of a relationship between magnetic fields in space and on Earth, *Geomagn. Aeron.*, *9*, 622, 1969.

McComas, D., S. Bame, P. Barker, W. Feldman, J. Phillips, P. Riley, and J. Griffee, Solarwind Electron Proton Alphamonitor (SWEPAM) for the Advanced Composotion Explorer, *Space Sci. Rev.*, *86*, 563–612, 1998.

Mende, S., et al., Far ultraviolet imaging from the IMAGE spacecraft. 1. System design, *Space Sci. Rev.*, *91*, 243–270, 2000a.

Mende, S. B., H. U. Frey, T. J. Immel, J.-C. Gérard, B. Hubert, and S. A. Fuselier, Global imaging of proton and electron aurorae in the far ultraviolet, *Space Sci. Rev.*, *109*, 211–254, 2003a.

Mende, S. B., H. U. Frey, B. J. Morsony, and T. J. Immel, Statistical behavior of proton and electron auroras during substorms, *J. Geophys. Res.*, *108*, doi:10.1029/2002JA009751, 2003b.

Mende, S. B., et al., Far ultraviolet imaging from the IMAGE spacecraft. 3. Spectral imaging of Lyman-α and OI 135.6 nm, *Space Sci. Rev.*, *91*, 287–318, 2000b.

Mende, S. B., et al., Far ultraviolet imaging from the IMAGE spacecraft. 2. wideband FUV imaging, *Space Sci. Rev.*, *91*, 271–285, 2000c.

Meng, C.-I., R. H. Holzworth, and S.-I. Akasofu, Auroral circle - delineating the poleward boundary of the quiet auroral belt, *J. Geophys. Res.*, *82*, 164–172, 1977.

Meurant, M., J.-C. Gérard, C. Blockx, B. Hubert, and V. Coumans, Propagation of electron and proton shock-induced aurora and the role of the interplanetary magnetic field and solar wind, *J. Geophys. Res.*, *109*, doi:10.1029/2004JA010453, 2004.

Milan, S. E., G. Provan, and B. Hubert, Magnetic flux transport in the dungey cycle: A survey of dayside and nightside reconnection rates, *J. Geophys. Res.*, *112*, doi:10.1029/2006JA011642, 2007.

Milan, S. E., A. Grocott, C. Forsyth, S. M. Imber, P. D. Boakes, and B. Hubert, A superposed epoch analysis of auroral evolution during substorm growth, onset and recovery: Open magnetic flux control of substorm intensity, *Ann. Geophys.*, pp. 659–668, 2009a.

Milan, S. E., J. Hutchinson, P. D. Boakes, and B. Hubert, Influences on the radius of the auroral oval, *Ann. Geophys.*, *27*, 2913–2924, 2009b.

Minatoya, H., N. Sato, T. Saemundsson, and T. Yoshino, Large displacements of conjugate auroras in the midnight sector, *J. Geomag. Geoelectr.*, *48*, 967–975, 1996.

Newell, P. T., C.-I. Meng, and K. M. Lyons, Suppression of discrete aurorae by sunlight, *Nature*, *381*, 766–767, 1996.

Newell, P. T., J. M. Ruohoniemi, and C.-I. Meng, Maps of precipitation by source region, binned by IMF, with inertial convection streamlines, *J. Geophys. Res.*, *109*, doi:10.1029/2004JA010499, 2004.

Newell, P. T., T. Sotirelis, and S. Wing, Diffuse, monoenergetic, and broadband aurora: The global precipitation budget, *J. Geophys. Res.*, *114*, doi:10.1029/2009JA014326, 2009.

Newell, P. T., T. Sotirelis, and S. Wing, Seasonal variations in diffuse, monoenergetic, and broadband aurora, *J. Geophys. Res.*, *115*, doi:10.1029/2009JA014805, 2010.

Ohtani, S., S. Wing, G. Ueno, and T. Higuchi, Dependence of premidnight field-aligned currents and particle precipiation on solar illumination, *J. Geophys. Res.*, *114*, doi:10.1029/2009JA014115, 2009.

Østgaard, N., S. B. Mende, H. U. Frey, L. A. Frank, and J. B. Sigwarth, Observations of non-conjugate theta aurora, *Geophys. Res. Lett.*, *30*, doi:10.1029/2003GL017914, 2003.

Østgaard, N., S. B. Mende, H. U. Frey, T. J. Immel, L. A. Frank, J. B. Sigwarth, and T. J. Stubbs, Interplanetary magnetic field control of the location substorm onset and auroral features in the conjugate hemisphere, *J. Geophys. Res.*, *109*, doi:10.1029/2003JA010370, 2004.

Østgaard, N., S. B. Mende, H. U. Frey, and J. B. Sigwarth, Simultaneous imaging of the reconnection spot in the opposite hemispheres during northward IMF, *Geophys. Res. Lett.*, *32*, doi:10.1029/2005GL024491, 2005a.

Østgaard, N., J. Moen, S. B. Mende, H. U. Frey, T. J. Immel, P. Gallop, K. Oksavik, and M. Fujimoto, Estimates of magnetotail reconnecton rate based on IMAGE FUV and EISCAT measurements, *Ann. Geophys.*, *23*, 123–134, 2005b.

Østgaard, N., N. A. Tsyganenko, S. B. Mende, H. U. Frey, T. J. Immel, M. Fillingim, L. A. Frank, and J. B. Sigwarth, Observations and model predictions of substorm auroral asymmetries in the conjugate hemispheres, *Geophys. Res. Lett.*, *32*, doi:10.1029/2004GL022166, 2005c.

Østgaard, N., S. B. Mende, H. U. Frey, J. B. Sigwarth, A. Aasnes, and J. Weygand, Auroral conjugacy studies based on global imaging, *J. Atmos. Terr. Phys.*, *69*, 249–255, 2007.

Oznovich, I., R. W. Eastes, R. E. Hoffman, M. Tur, and I. Glaser, The aurora at quiet magnetospheric conditions: Repeatability and dipole tilt angle dependence, *J. Geophys. Res.*, *98*, 3789–3797, 1993.

Papitashvili, V. O., and F. J. Rich, High-latitude ionospheric convection models derived from Defense Meteorological Satellite Program ion drift observations and parameterized by the interplanetary magnetic field strength and direction, *J. Geophys. Res.*, *107*, doi:10.1029/2001JA000264, 2002.

Parker, E. N., The alternative paradigm for magnetospheric physics, *J. Geophys. Res.*, *101*, 10,587–10,625, 1996.

Parker, E. N., Reply, *J. Geophys. Res.*, *102*, 9657–9658, 1997.

Parker, E. N., *Newton, Maxwell, and magnetospheric physics*, *AGU Geophysical monograph*, vol. 118, pp. 1–10, Washington D.C., 2000.

Parker, E. N., *Conversations on electric and magnetic fields in the cosmos*, Princeton University Press, Princeton and Oxford, 2007.

Paschmann, G., S. Haaland, and R. Treumann, *Auroral Physics*, Kluwer Academic, Dordrecht, 2003.

Petrukovich, A. A., Dipole tilt effects in plasma sheet B_y : Statistical model and extreme values, *Ann. Geophys.*, pp. 1343–1352, 2009.

Provan, G., M. Lester, S. B. Mende, and S. E. Milan, Statistical study of high-latitude plasma flow during magnetospheric substorms, *Ann. Geophys.*, *22*, 3607–3624, 2004.

Pulkkinen, T. I., D. N. Baker, R. J. Pellinen, J. S. Murphree, and L. A. Frank, Mapping of the auroral oval and individual arcs during substorms, *J. Geophys. Res.*, *100*, 21,987–21,994, 1995.

Richmond, A. D., Ionospheric electrodynamics using magnetic apex coordinates, *J. Geomag. Geoelectr.*, *47*, 191–212, 1995.

Robinson, R. M., R. R. Vondrak, K. Miller, T. Dabbs, and D. Hardy, On calculating ionospheric conductances from the flux and energy of precipitating electrons, *J. Geophys. Res.*, *92*, 2565–2569, 1987.

Ruohoniemi, J. M., and R. A. Greenwald, Dependencies of high-latitude plasma convection: Consideration of interplanetary magnetic field, seasonal, and universal time factors in statistical patterns, *J. Geophys. Res.*, *110*, doi:10.1029/2004JA010815, 2005.

Sato, N., R. Fujii, T. Ono, H. Fukunishi, T. Hirasawa, T. Araki, S. Kokubun, K. Makita, and T. Saemundsson, Conjugacy of proton and electron auroras observed near $L = 6.1$, *Geophys. Res. Lett.*, *13*, 1368–1371, 1986.

Sato, N., T. Nagaoka, K. Hashimoto, and T. Saemundsson, Conjugacy of isolated auroral arcs and nonconjugate auroral breakups, *J. Geophys. Res.*, *103*, 11,641–11,652, 1998.

Sato, N., D. M. Wright, C. W. Carlson, Y. Ebihara, M. Sato, T. Saemundsson, S. E. Milan, and M. Lester, Generation region of pulsating aurora obtained simultaneously by the FAST satellite and a Syowa-Iceland conjugate pair of observatories, *J. Geophys. Res.*, *109*, doi:10.1029/2004JA010419, 2004.

Sato, N., A. Kadokura, Y. Ebihara, H. Deguchi, and T. Saemundsson, Tracing geomagnetic conjugate points using exceptionally similar synchronous auroras, *Geophys. Res. Lett.*, *32*, doi:10.1029/2005GL023710, 2005.

Scholer, M., On the motion of artificial ion clouds in the magnetosphere, *Planet. Space Sci.*, *18*, 977, 1970.

Sckopke, N., A general relation between the energy of trapped particles and the disurbance field near the earth, *J. Geophys. Res.*, *71*, 3125–3130, 1966.

Sergeev, V. A., E. M. Sazhina, N. A. Tsyganenko, J. A. Lundblad, and F. Søraas, Pitch-angle scattering of energetic protons in the magnetotail current sheet as the dominant source of their isotropic precipitation into the nightside ionosphere, *Planet. Space Sci.*, *31*, 1147–1155, 1983.

Sergeev, V. A., K. Liou, C.-I. Meng, P. T. Newell, M. Brittnacher, G. Parks, and G. D. Reeves, Development of auroral streamers in association with localized impulsive injections to the inner magnetotail, *Geophys. Res. Lett.*, *26*, 417–420, 1999.

Shay, M. A., and M. Swisdak, Three-species collisionless reconnection: Effect of O+ on magnetotail reconnection, *Phys. Rev. Lett.*, *93*, doi:10.1103/PhysRevLett.93.175001, 2004.

Shue, J.-H., P. T. Newell, K. Liou, C.-I. Meng, and S. W. H. Cowley, Interplanetary magnetic field B_x asymmetry effect on auroral brightness, *J. Geophys. Res.*, *107*, doi: 10.1029/2001JA000229, 2002.

Siscoe, G. L., and T. S. Huang, Polar cap inflation and deflation, *J. Geophys. Res.*, *90*, 543–547, 1985.

Smith, C., J. L'Heureux, N. Ness, M. Acuña, L. Burlaga, and J. Scheifele, The ACE magnetic fields experiment, *Space Sci. Rev.*, *86*, 613–632, 1998.

Snekvik, K., R. Nakamura, N. Østgaard, S. Haaland, and A. Retino, The hall current system revealed as a statistical significant pattern during fast flows, *Ann. Geophys.*, *26*, 3429–3437, 2008.

Song, P., V. M. Vasyliunas, and X.-Z. Zhou, Magnetosphere-ionosphere/thermosphere coupling: Self-consistent solutions for a one-dimensional stratified ionosphere in three-fluid theory, *J. Geophys. Res.*, *114*, doi:10.1029/2008JA013629, 2009.

Stenbaek-Nielsen, H. C., and A. Otto, Conjugate auroras and the interplanetary magnetic field, *J. Geophys. Res.*, *102*, 2223–2232, 1997.

Stenbaek-Nielsen, H. C., T. N. Davis, and N. W. Glass, Relative motion of auroral conjugate points during substorms, *J. Geophys. Res.*, *77*, 1844–1858, 1972.

Stenbaek-Nielsen, H. C., E. M. Wescott, T. N. Davis, and R. W. Peterson, Differences in auroral intensity at conjugate points, *J. Geophys. Res.*, *78*, 659–671, 1973.

Stone, E. C., A. M. Frandsen, R. A. Mewaldt, E. R. Christian, D. Margolies, J. F. Ormes, and F. Snow, The Advanced Composotion Explorer, *Space Sci. Rev.*, *86*, 1–22, 1998.

Strickland, D. J., R. J. Cox, R. R. Meier, and D. P. Drob, Global O/N_2 derived from DE 1 FUV dayglow data: Technique and examples from two storm periods, *J. Geophys. Res.*, *104*, 4251–4266, 1998.

Stubbs, T. J., R. R. Vondrak, N. Østgaard, J. B. Sigwarth, and L. A. Frank, Simultaneous observations of the auroral ovals in both hemispheres under varying conditions, *Geophys. Res. Lett.*, *32*, doi:10.1029/2004GL021199, 2005.

Svalgaard, L., *Sector structure of the interplanetary magnetic field and daily variation of the geomagnetic field at high latitudes*, Geophys Pap. R-6, Danish Meterol. Inst., Charlottenlund, 1968.

Tanaka, T., Interplanetary magnetic field B_y and auroral conductance effects on high-latitude ionospheric convection patterns, *J. Geophys. Res.*, *106*, 24,505–24,516, 2001.

Tsurutani, B. T., D. E. Jones, R. P. Lepping, E. J. Smith, and D. G. Sibeck, The relationship beteween the IMF B_y and the distant tail $(150 - 238R_E)$ lobe and plasmasheet B_y fields, *Geophys. Res. Lett.*, *11*, 1082–1085, 1984.

Tsyganenko, N. A., A model of the near magnetosphere with a dawn-dusk asymmetry 1. Mathematical structure, *J. Geophys. Res.*, *107*, doi:10.1029/2001JA000219, 2002.

VanZandt, T. E., W. L. Clark, and J. W. Warnock, Magnetic Apex Coordinates: A Magnetic Coordinate System for the Ionospheric F2 Layer, *J. Geophys. Res.*, *77*, 2406–2411, 1972.

Vasyliunas, V. M., *Steady state aspects of magnetic field line merging*, *Geophys. Mongr. Ser.*, vol. 30, pp. 25–31, AGU, Washington D.C., 1984.

Vasyliunas, V. M., Electric field and plasma flow: What drives what?, *Geophys. Res. Lett.*, *28*, 2177–2180, 2001.

Vasyliunas, V. M., Relation between magnetic fields and electric currents in plasmas, *Ann. Geophys.*, *23*, 2589–2597, 2005a.

Vasyliunas, V. M., Time evolution of electric fields and currents and the generalized Ohm's law, *Ann. Geophys.*, *23*, 1347–1354, 2005b.

Vasyliunas, V. M., and P. Song, Meaning of ionospheric Joule heating, *J. Geophys. Res.*, *110*, doi:10.1029/2004JA010615, 2005.

Vo, H. B., J. S. Muphree, D. Hearn, P. T. Newell, and C.-I. Meng, A satellite study of dayside auroral conjugacy, *Ann. Geophys.*, *13*, 1134–1143, 1995.

Vorobjev, V. G., O. I. Yagodkina, D. Sibeck, K. Liou, and C.-I. Meng, Aurora conjugacy during substorms: Coordinated Antarctic ground and Polar Ultraviolet observations, *J. Geophys. Res.*, *106*, 24,579–24,591, 2001.

Wang, H., H. Lühr, S. Y. Ma, and H. U. Frey, Interhemispheric comparison of average substorm onset locations: Evidence for deviation from conjugacy, *Ann. Geophys.*, *25*, 989–999, 2007.

Watanabe, M., A. Kadokura, N. Sato, and T. Saemundsson, Absence of geomagnetic conjugacy in pulsating auroras, *Geophys. Res. Lett.*, *34*, doi:10.1029/2007GL030469, 2007a.

Watanabe, M., G. J. Sofko, K. Kabin, R. Rankin, A. J. Ridley, C. R. Clauer, and T. I. Gombosi, Origin of the interhemispheric potential mismatch of merging cells for interplanetary magnetic field B_y-dominated periods, *J. Geophys. Res.*, *112*, doi: 10.1029/2006JA012179, 2007b.

Weimer, D. R., Models of high-latitude electric potentials derived with a least error fit of spherical harmonic coefficients, *J. Geophys. Res.*, *100*, 19,595–19,607, 1995.

Weimer, D. R., Correction to "Predicting interplanetary magnetic field (IMF) propagation delay times using the minimum variance technique", *J. Geophys. Res.*, *109*, doi: 10.1029/2004JA010691, 2004.

Wescott, E. M., Magnetoconjugate phenomena, *Space Sci. Rev.*, *5*, 507–561, 1966.

Wing, S., P. T. Newell, D. G. Sibeck, and K. B. Baker, A large statistical study of the entry of interplanetary magnetic field y-component into the magnetosphere, *Geophys. Res. Lett.*, *22*, 2083–2086, 1995.

Zhang, Y., L. J. Paxton, and A. T. Y. Lui, An unusual nightside distortion of the auroral oval: TIMED/GUVI and IMAGE/FUV observations, *J. Geophys. Res.*, *111*, doi: 10.1029/2005JA011217, 2006.

Paper I

K. M. Laundal and N. Østgaard (2008)

Persistent global proton aurora caused by high solar wind dynamic pressure

Journal of Geophysical Research Vol. 113, doi:10.1029/2008JA013147

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 113, A08231, doi:10.1029/2008JA013147, 2008

Persistent global proton aurora caused by high solar wind dynamic pressure

K. M. Laundal[1] and N. Østgaard[1]

Received 6 March 2008; revised 1 May 2008; accepted 9 May 2008; published 28 August 2008.

[1] Global images of the proton aurora taken with the SI-12 camera onboard the IMAGE satellite reveal a very direct relationship between the solar wind dynamic pressure and the intensity of the global proton aurora. We show that an increase in dynamic pressure leads to an immediate and persistent increase in proton precipitation, also when the increase is slow. When the dynamic pressure decreases, the proton aurora diminishes. Five events during geomagnetic quiet times, with mostly northward IMF, have been selected in order to characterize the proton aurora caused exclusively by high dynamic pressure and establish important criteria that the dynamic pressure-induced precipitation mechanism(s) must satisfy. We also present measurements during southward IMF and show that the combined effect of high solar wind dynamic pressure and southward IMF produces intense global proton aurora. Some of the characteristics are: (1) The aurora is global, with peak intensities at midnight and flanks. (2) A dawn/dusk asymmetry shows that the precipitation originates from magnetospheric protons that have undergone gradient/curvature drift. (3) The time delay between ground magnetic signatures of a change in the solar wind dynamic pressure and a change in global proton aurora is short (−2 minutes). Our observations indicate that the precipitation mechanism(s) behind the proton aurora during high dynamic pressure is directly connected to the compression of the magnetosphere, both at the flanks and nightside.

Citation: Laundal, K. M., and N. Østgaard (2008), Persistent global proton aurora caused by high solar wind dynamic pressure, *J. Geophys. Res.*, *113*, A08231, doi:10.1029/2008JA013147.

1. Introduction

[2] The solar wind dynamic pressure is an important parameter for controlling the shape and size of the magnetosphere. A sudden increase in dynamic pressure can cause global disturbances in the magnetosphere, which are accompanied by an increase of the intensity of the aurora [*Vorobyev*, 1974; *Craven et al.*, 1986; *Spann et al.*, 1998; *Zhou and Tsurutani*, 1999; *Zhou et al.*, 2003]. Recent studies, using global images from space, have shown that the auroral intensifications appear first at the dayside and propagate to the nightside at an ionospheric speed consistent with the speed of the solar wind discontinuity [*Zhou and Tsurutani*, 1999], at least when the dynamic pressure pulse is preceded by a period with northward IMF [*Boudouridis et al.*, 2003]. These transient intensifications are often referred to as shock auroras, since they are believed to be caused by the sudden magnetospheric reconfiguration due to the sudden increase of solar wind pressure. *Liou et al.* [2006] showed that a negative sudden impulse in solar wind dynamic pressure leads to a fast, global reduction in auroral intensity.

[3] The effect of a long-lasting high solar wind dynamic pressure on the aurora has been much less studied. *Zhou and Tsurutani* [2003] showed that the auroral intensity at dawn and dusk, seen by the UVI camera on the Polar satellite, increased (decreased) during gradually increasing (decreasing) dynamic pressure. Using the same camera, *Liou et al.* [2007] showed that the global auroral luminosity is higher when the magnetosphere is compressed, leading to the term "compression aurora". They also found, using in situ DMSP measurements, that most of the auroral emissions during compression were due to diffuse electron precipitation.

[4] The production mechanism for the persistent aurora during high dynamic pressure is believed to differ from the transient mechanisms behind the shock aurora. *Zhou and Tsurutani* [2003] suggested that the dawn and dusk aurora could be due to Kelvin–Helmholtz waves on the magnetopause. *Liou et al.* [2007] suggested that the increased intensity is due to a larger loss cone caused by a decrease in the mirror ratio, B_m/B_{eq}, since the equatorial magnetic field strength, B_{eq} increases more than at the mirror points, B_m, during compression.

[5] The first, and so far only camera with the ability to look at the global aurora solely produced by protons, is the SI-12 camera [*Mende et al.*, 2000] on board the IMAGE satellite. Using this ability, *Meurant et al.* [2003, 2004] showed that a sudden increase in the solar wind dynamic pressure affects both electron and proton precipitation however with some differences in timing and distribution. *Coumans et al.* [2006] showed in a statistical study, that the

[1]Department of Physics and Technology, University of Bergen, Bergen, Norway.

0148-0227/08/2008JA013147$09.00

total power of the proton precipitation was correlated with the solar wind dynamic pressure. The slope in the regression line between dynamic pressure and proton precipitation energy flux was highest near midnight and higher in the summer than in the winter.

[6] In this paper we look at the lasting effect of a change in solar wind dynamic pressure on the global proton aurora, seen by the SI-12 camera. We show that high dynamic pressure leads to a persistent intense proton aurora. We look at five events during geomagnetic quiet times, in order to establish some characteristics of the dynamic pressure induced aurora, that a production mechanism must account for. The quiet times are chosen to avoid contributions from other processes, such as substorms and convection caused by southward IMF, and the magnetosphere can be assumed to be in a quasisteady state. Although the effect of the dynamic pressure is clearly seen in the quiet time measurements, the aurora is faint. Two events with southward IMF are also studied, and we show that a combination of southward IMF and high solar wind dynamic pressure may be decisive parameters for generation of intense proton precipitation.

[7] In section 2, the observations are presented. Section 3 contains a discussion of possible production mechanisms induced by the high solar wind dynamic pressure, in relation to our observations. Section 4 contains the conclusions.

2. Observations

[8] The SI-12 camera [*Mende et al.*, 2000] has a narrow passband with peak sensitivity at 121.8 nm and block-out of the Hydrogen emission line at 121.567 nm (Ly-α). Hence only Doppler shifted Ly-α emissions are detected. Such emissions are produced by Hydrogen descending at a sufficiently high speed along the line of sight. This Hydrogen is in turn produced by proton precipitation which charge exchange with atmospheric constituents. Only protons with energy above 1 keV can produce detectable emissions [*Mende et al.*, 2003]. Above this threshold, the sensitivity of the SI-12 camera is highest at 3–4 keV for an isotropic distribution of protons and decreases with increasing energy [*Frey et al.*, 2003].

[9] The relation between an instrument count in SI-12 and a physical quantity such as energy flux or number flux requires that, e.g., the mean energy of the precipitation is known [*Gérard et al.*, 2001; *Frey et al.*, 2003]. Since global measurements of the mean energy are not available, statistical maps [*Hardy et al.*, 1989] are often used. However, in this paper, we focus on proton aurora produced by a mechanism induced by high solar wind dynamic pressure, which may produce different energy spectra than other mechanisms. Since assumptions based on statistical maps of proton precipitation may introduce artificial inaccuracies, we will avoid the use of energy flux and present the SI-12 data in corrected instrument counts (instrument calibration is described in *Frey et al.* [2003]). We use the terms "instrument counts" and "proton aurora intensity" interchangeably. A constant background of 4 counts/pixel, and a variable crescent shaped dayglow, has been subtracted from each image. The time resolution of the SI-12 camera is approximately 2 minutes (one satellite spin period), and the integration time is 5 seconds.

[10] For measurements of the solar wind and IMF, we use the ACE satellite, located near the Lagrange point at $x_{GSM} \sim 250\ R_E$. To determine the timing of the solar wind arrival to the magnetopause (taken as 10 R_E), we time-shift each data point by s/v, where s is the distance along the GSM x direction to the ACE satellite. We then adjust the time-shift so that solar wind dynamic pressure discontinuities coincide with abrupt changes in the SYM-H index. The SYM-H index is derived from low latitude ground magnetometer measurements, and can be seen mainly as a one minute equivalent to the Dst index [*Wanliss and Showalter*, 2006], and is sensitive to changes in the ring current, but also has a contribution from the magnetopause current [e.g., *Burton et al.*, 1975]. Thus we have an accurate time, within the one minute time resolution of SYM-H and the ACE measurements, for the arrival of the discontinuity at the magnetopause, while the timing for other solar wind data points may be slightly dislocated in time.

2.1. Five Quiet Time Dynamic Pressure Events

[11] In Figures 1a–1e, we show five events with quiet geomagnetic conditions, to avoid any superimposed effects of processes independent of the solar wind dynamic pressure. All events in this study are from 2001, when the $\sim$7 R_E apogee of the IMAGE satellite was close to the north pole, which enabled continuous observation of the area above 50° magnetic latitude for several hours per 14 hour orbit.

2.1.1. 21–22 April 2001

[12] Figure 1a shows the end of a long period with high dynamic pressure. The positive direction of the IMF B_z (bottom) indicates low geomagnetic activity. The AE indices (not shown) also showed low activity at this time. The third panel from the top shows that the solar wind dynamic pressure first increases at 23:35 UT, and then quickly decreases just before 00:40 UT. The dynamic pressure reduction is accompanied by a fast drop in proton aurora intensity. In the more than 1 hour long period with SI-12 data prior to the drop in dynamic pressure, the MLT keogram (top) shows two steady maxima in aurora intensity at dusk, centered approximately at 18 MLT, and at dawn, centered at approximately 4 MLT. The dusk aurora is more intense than at dawn.

[13] The fourth panel from the top shows the solar wind velocity (black) and density. When the dynamic pressure decreases, at 00:40 UT, the solar wind velocity increases. The drop in dynamic pressure is caused by a large decrease in density, from approximately 40 to 10 cm^{-3}. The relative changes in density and velocity are important for various proposed production mechanisms, such as Kelvin–Helmholtz (K–H) waves on the magnetopause [*Zhou and Tsurutani*, 2003], which will be discussed in section 3.

[14] In the top panel of Figure 2a, we have plotted the correlation coefficients between the proton aurora intensity and solar wind dynamic pressure at different magnetic local times for the 21–22 April event. The intensity of the proton aurora depends on the state of the magnetosphere, which may have a highly delayed response to changes in the solar wind and the IMF. Hence consecutive measurements by the SI-12 camera, spaced by only 2 minutes, are not statistically independent. This means that even though the correlations presented in Figure 2 are based on a large number of

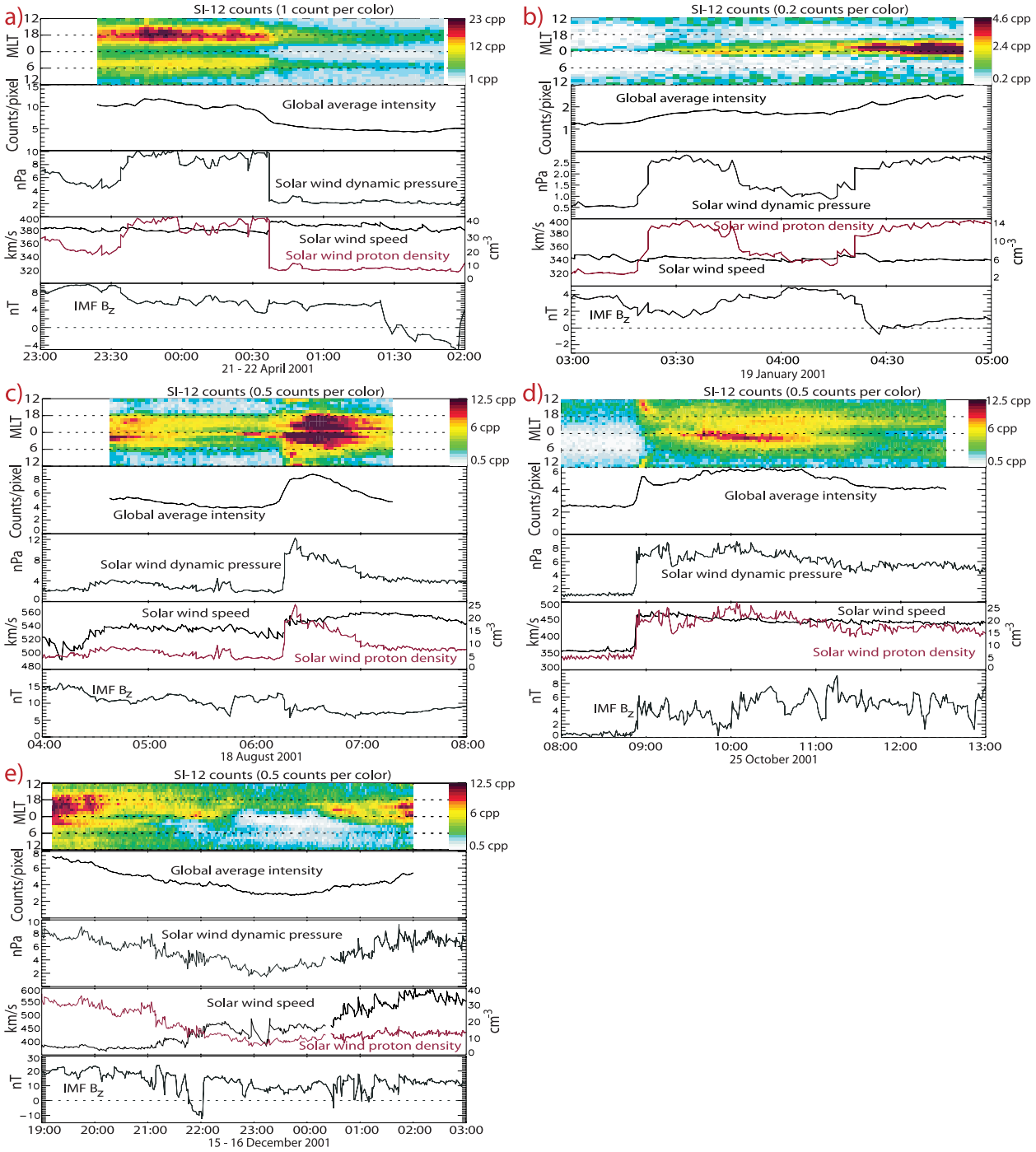

Figure 1. The panels show, from top to bottom: (1) MLT keogram. Each pixel represents 1 hour magnetic local time, integrated from 60 to 85° magnetic latitude. The unit on the color bars is counts per pixel (cpp). (2) Global average proton aurora intensity, measured in corrected instrument counts/pixel in the area between 60 and 85 degrees magnetic latitude. (3) Time-shifted solar wind dynamic pressure. (4) Solar wind velocity (black) and density (red). (5) IMF B_z. Note that the scales vary.

samples, one should be careful in concluding about the true correlation, since the significance of the correlation coefficients is much less than they would be if the data were obtained with larger time intervals. However we shall use the calculated correlation coefficients only as a measure of how well the data in each particular event vary together, and look at the differences between magnetic local times. The top panel of Figure 2a shows a clear pattern, with higher correlation at dawn and dusk, than at midnight and noon. In the bottom panel, we have used data points only from the period prior to the dynamic pressure decrease at 00:40 UT, to look at the common variation without the effect of the

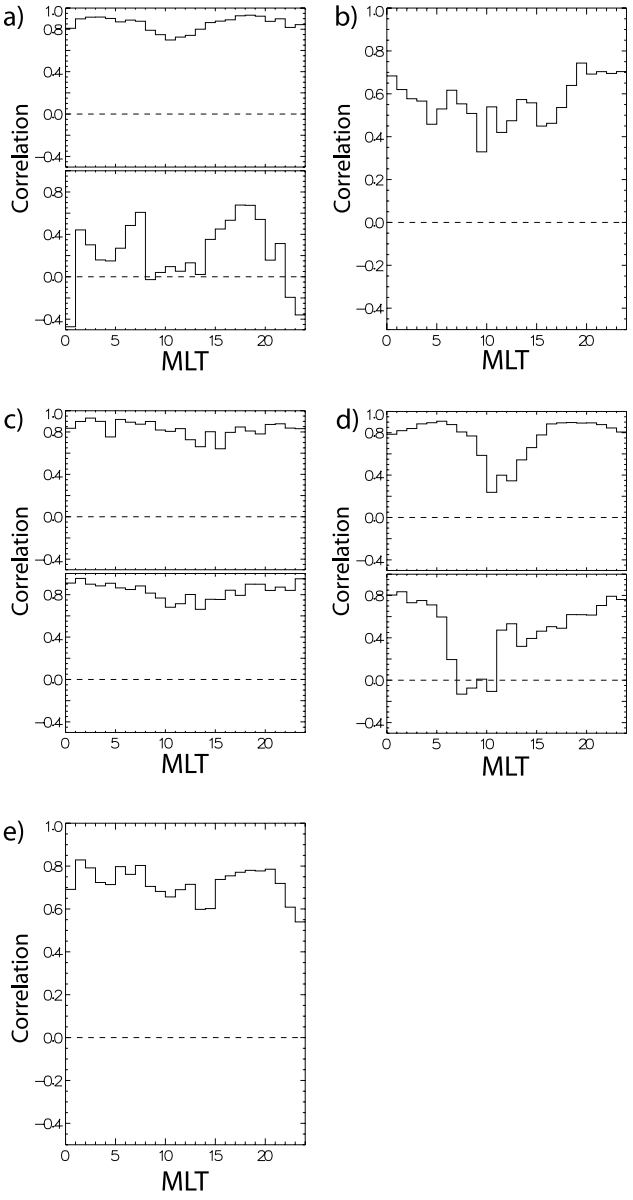

Figure 2. The calculated correlation coefficients between solar wind pressure and proton aurora intensity at different MLTs for the events in Figure 1. (a) 21–22 April, (b) 19 January, (c) 18 August, (d) 25 October, (e) 15–16 December, all from 2001. The bottom panels show the correlation without pressure discontinuity.

large dynamic pressure drop. We see that the correlation is lower at all local times, as expected, but the >0.6 correlation seen on the dusk side suggests that also the small changes in solar wind dynamic pressure affect the proton aurora. This is confirmed by the MLT keogram, which shows an increase in the dusk aurora intensity in response to the dynamic pressure increase at 23:35 UT.

2.1.2. 19 January 2001

[15] Figure 1b shows data from 19 January 2001. A dynamic pressure increase from 0.5 to 2.5 nPa is seen at 03:20 UT. The MLT keogram shows a small increase in intensity mainly at the nightside (note the greatly reduced color scale from the previous event). Between 03:50 and 04:20 the dynamic pressure is down to 1 nPa, and then it rises again to 2–2.5 nPa. At the time of this second dynamic pressure increase, the auroral intensity increases again, to a higher value than in the first increase. The IMF B_z was northward (≈3–4 nT) for approximately two hours prior to the second dynamic pressure increase, when it became ≈0 nT. This turning of the IMF may be the reason for the higher proton aurora intensity at this time.

[16] This example illustrates that the solar wind dynamic pressure has a clear effect on the proton aurora, also for relatively low values. Whether or not these observations are a consequence of the abrupt increase, or an effect of persistent high dynamic pressure, is unclear, due to the relatively short duration of the first dynamic pressure pulse.

[17] The aurora was very faint at all MLTs in this event, only a few counts higher than the background. However, compared to other MLTs, the correlation between dynamic pressure and aurora intensity was high on the pre midnight MLTs (Figure 2b).

2.1.3. 18 August 2001

[18] Figure 1c shows proton aurora measurements from 18 August 2001. This is the late recovery phase of a geomagnetic storm, but the geomagnetic conditions are otherwise quiet: no substorms are detected, and the IMF B_z is strictly positive. At approximately 06:15 UT we see a sudden increase in dynamic pressure, which results in an intensification of the proton aurora at all magnetic local times. The most intense proton aurora is found on the nightside. During the subsequent hour, both the dynamic pressure and the proton aurora intensity decreases.

[19] The top panel of Figure 2c shows the correlation between solar wind dynamic pressure and proton aurora intensity at different magnetic local times. The bottom panel shows the correlation only in the period starting 20 minutes after the increase in solar wind dynamic pressure. The high correlation seen in both cases indicates that the dynamic pressure controls the proton aurora at all times.

2.1.4. 25 October 2001

[20] Figure 1d shows the response in proton aurora intensity to a sudden increase in solar wind dynamic pressure, followed by a 3.5 hours long period with slowly varying, high dynamic pressure. In the ∼20 minutes following the sudden increase in dynamic pressure, the proton aurora intensifies on the dayside, and travels anti-sunward, mainly along the dusk flank. This is consistent with the description of the proton aurora caused by a sudden increase in dynamic pressure described by *Meurant et al.* [2004]. According to *Frey et al.* [2004], there was a substorm onset at 09:09:56 UT, located at 23:26 MLT. The local intensification seen at this time and location in the MLT keogram may be due to this possibly pressure triggered substorm. It should be mentioned that we have seen no substorm signatures in ground magnetometer data in this event. In the last ∼2 hours with SI-12 data, the aurora is located mainly at the flanks, with a similar configuration as the 22 April event (Figure 1a).

[21] Figure 2d shows the correlation between proton aurora intensity and solar wind dynamic pressure at different local times for the whole time interval (top) and the period starting 20 minutes after the solar wind dynamic pressure increase (bottom). As discussed above, the inten-

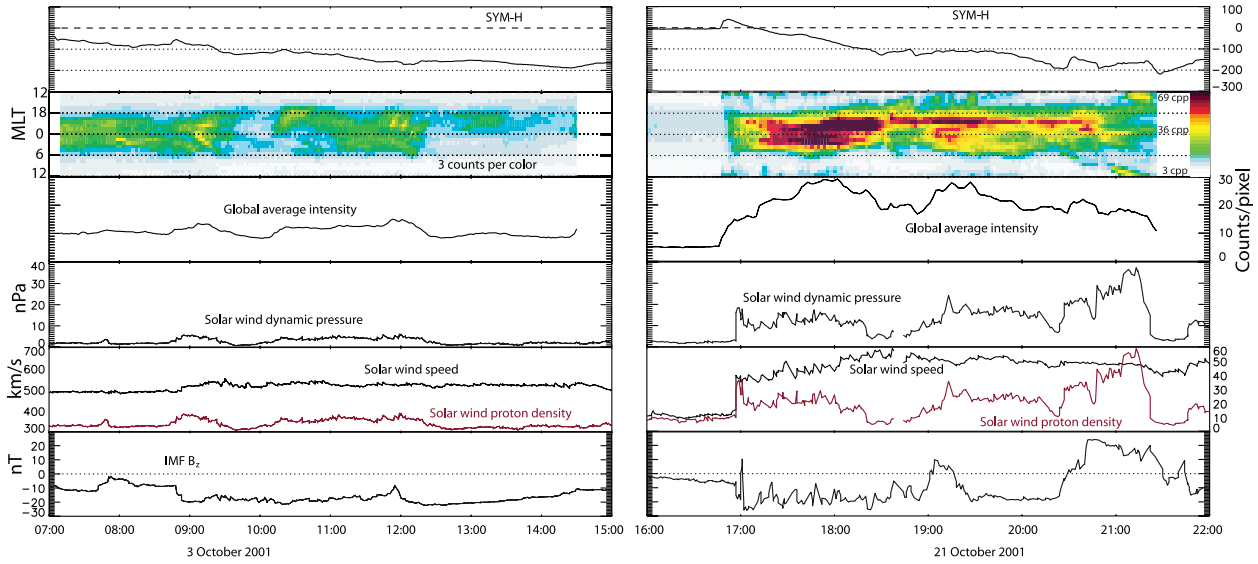

Figure 3. 3 October (left) and 21 October (right), 2001. The panel show, from top to bottom: (1) SYM-H [nT], a similar index as the Dst, with 1 minute resolution. (2) MLT keogram (note the crude color scale, due to the high intensity in these events). (3) Global average proton aurora intensity ($\geq$50 magnetic latitude). (4) Solar wind dynamic pressure. (5) Solar wind velocity (black) and density (red). (6) panel: IMF B_z.

sity in this event may have been affected by a substorm. However it is not believed that the substorm was equally important at all local times, and we see a similar pattern here as in the event from April 2001, with higher correlation on the flanks, and lower correlation on the dayside. The bottom panel shows that the correlation is lower without the transient effect of the sudden increase in dynamic pressure, but it is still high at the nightside and flanks, compared to the dayside.

2.1.5. 15–16 December 2001

[22] Figure 1e shows an event from 15 December 2001 when the dynamic pressure decreased slowly from $\approx$9 to $\approx$2 nPa over 4 hours, and then increased again. This was also the case for the proton aurora intensity, which started at an average of 7 counts, decreased to 3 counts, and increased to 5 counts again. In the first 70 minutes, the proton aurora was primarily located in an $\approx$3 hours wide MLT sector in the post midnight sector, and a much wider area at dusk. The dusk aurora is visible throughout the period. The IMF B_z was positive or close to zero during the whole event, except for a period of $\approx$15 minutes around 22:00 UT, when it turned southward. This is probably the reason for the seen in the MLT keogram near midnight brief intensification at this time.

[23] Due to the absence of large discontinuities in the solar wind, this event demonstrates one of our main points: High solar wind dynamic pressure leads to an enhanced proton aurora intensity, also when the solar wind dynamic pressure changes slowly (or not at all).

[24] The correlation between solar wind dynamic pressure and proton aurora intensity (Figure 2e) is quite high at all magnetic local times in this event, including the dayside. The minimum seen between 23 and midnight can probably be explained by the fact that the intensity at this local time was low, except for the 15 minutes with southward IMF.

2.2. Solar Wind Dynamic Pressure Control of High Intensity Proton Aurora

[25] So far we have looked only at events when the IMF was mostly northward. Many previous studies show that changes in solar wind dynamic pressure have much more dramatic consequences when it is combined with southward IMF [e.g., *Meurant et al.*, 2004; *Lee and Lyons*, 2004; *Boudouridis et al.*, 2004; *Lee et al.*, 2007], and we want to see if this holds also for the proton aurora intensity.

[26] Figure 3 shows data from two geomagnetic storm main phases, characterized by steep decreases in SYM-H, a signature of significant injection of energy into the ring current. It is believed that an increase of ring current energy is associated with injection of protons of isotropic pitch angle distribution, which is a major cause for precipitation [*Søraas et al.*, 2002; *Mende et al.*, 2002]. The similarity in the magnitude of the drop in SYM-H suggests that the events in Figure 3 are to some extent comparable in terms of injection of isotropic protons.

[27] There are however big differences in the proton aurora intensity: On 21 October 2001, the intensity (Third panel from the top) was approximately a factor of 2 higher than 3 October. Note that the color scale in the MLT keogram (second panel) has been adjusted to 3 counts per color due to the high intensity, and that we look at the entire region poleward of 50° magnetic latitude, due to the expansion of the auroral oval. The axis and color scales are the same for both events in Figure 3.

[28] The solar wind data shows that the IMF B_z (bottom panel) was negative, and had a similar magnitude in the two events. The dynamic pressure (fourth panel from the top) was however much higher on 21 October than it was on 3 October. This indicates that the difference in dynamic pressure is the reason for the large difference in proton aurora intensity. A one-to-one relationship between solar

wind dynamic pressure and proton aurora intensity cannot be claimed when so many processes are present that can cause precipitation. However we also see that the periods with the most intense aurora in both events coincided with the periods when the dynamic pressure was highest. An exception to this is found at the end of the 21 October event, when the dynamic pressure reached very high values (>30 nPa), with a relatively faint nightside aurora. This can probably be explained by the fact that the IMF B_z had turned northward. The keogram shows that most of the intensity in this period is located on the dayside.

[29] Figure 3 indicates that high solar wind dynamic pressure or southward IMF alone are not able to produce proton aurora with very high intensity. A combination of high dynamic pressure and southward IMF may however be a necessary and sufficient criterion for high intensity proton aurora. This is in accordance with the statistical study by *Coumans et al.* [2006], who found that the slope of the linear regression line for proton aurora power and solar wind dynamic pressure was higher during southward IMF.

3. Discussion

[30] The observations presented in this paper show that a change in solar wind dynamic pressure is reflected in the global proton aurora intensity. The mechanism(s) responsible for this relationship must account for several characteristics: (1) Persistent high dynamic pressure leads to a persistent high proton aurora intensity. This is the reason that we believe the mechanism(s) behind the proton aurora observed in this study differ from the mechanisms behind the transient shock aurora. (2) The time delay between a change in ground magnetic field signatures of a change in solar wind dynamic pressure, both positive and negative, and a change in proton aurora intensity is very short, possibly below the 2 minutes time resolution of the SI-12 images. (3) Some events (22 April and the end of 25 October) showed that the dynamic pressure induced aurora was predominantly located at the flanks, but this is not always seen (18 August, 19 January). (4) The dawn/dusk asymmetry shows that the aurora originates from magnetospheric protons that have undergone gradient/curvature drift, and precipitate due to some mechanism induced by the high dynamic pressure. This is consistent with *Liou et al.* [2007], who reported a similar pattern in the electron dominated aurora caused by high dynamic pressure, only with the highest intensity at dawn. (5) The intensification of the proton aurora during high dynamic pressure is seen regardless of the sign of the IMF B_z. However a much more dramatic effect of high solar wind dynamic pressure is seen when the IMF is southward (Figure 3) than when it is northward, consistent with many reports on the effect of dynamic pressure on the magnetosphere.

[31] The two theories used to explain the energy transfer from the solar wind to the magnetosphere are reconnection [*Dungey*, 1961], which is least efficient when IMF $B_z > 0$, and viscous interaction [*Axford and Hines*, 1961], which can be independent of the sign of B_z. We have shown five events where high dynamic pressure led to increased precipitation, even though the IMF was northward, which may be taken as a sign of viscous interaction. One possible viscous interaction mechanism is Kelvin–Helmholtz waves on the magnetopause flanks, which agrees with the flank location of the aurora seen in Figure 1a and at the end of the 25 October event (Figure 1d). This was the argument of *Zhou and Tsurutani* [2003], who observed dynamic pressure dependent flank auroras using the Polar UVI camera, which is sensitive mainly to electron produced auroral emissions.

[32] The onset condition for K–H waves is [e.g., *Baumjohann and Treumann*, 1997]

$$(\mathbf{k}\cdot\Delta\mathbf{v})^2 > \frac{n_1+n_2}{m_p\mu_0 n_1 n_2}\left((\mathbf{k}\cdot\mathbf{B}_1)^2+(\mathbf{k}\cdot\mathbf{B}_2)^2\right) \quad (1)$$

where subscripts 1 and 2 denote the regions outside and inside the magnetopause, respectively. **B** is the magnetic field, n is the plasma density, **k** is the wave vector, and $\Delta\mathbf{v}$ is the shear velocity across the magnetopause. The inequality 1 shows that the shear velocity is a more important parameter in controlling K–H wave growth than the plasma density. Hence K–H wave activity is much more sensitive to changes in solar wind velocity, than density [e.g., *Engebretson et al.*, 1998]. In the 21–22 April event (Figure 1a), the proton aurora (and solar wind dynamic pressure) was clearly governed by changes in solar wind density, with an increasing solar wind velocity at the time of the drop in dynamic pressure. The 19 January event (Figure 1b) also showed a clear response to an increase in density, with a nearly constant velocity. The first period of the 15–16 December event (Figure 1e) had high solar wind density, while the last period had high velocity, with no apparent difference in the dynamic pressure control of the overall proton aurora intensity. These examples illustrate that it is the combined effect of the solar wind velocity and density, namely the dynamic pressure, $P = m_p n v^2$, and not the velocity in particular, that governs the global proton aurora, as would be expected if Kelvin–Helmholtz waves was the driving mechanism.

[33] A more direct mechanism was proposed by *Liou et al.* [2007], who observed persistent increased intensity in the electron-dominated aurora seen by Polar UVI during high solar wind dynamic pressure. They suggest that the increased intensity is caused by a decrease in the mirror ratio, B_m/B_{eq}, since the equatorial magnetic field increases most during compression, leading to a larger loss cone.

[34] Many of the features in the observations presented in this paper support the idea that the dynamic pressure induced aurora follows directly from the compression of the magnetosphere. A quick compression/expansion of the magnetosphere following a change in the solar wind dynamic pressure quickly turns on or off the precipitation mechanism, in agreement with, e.g., the observations in Figure 1d (on) and Figure 1a (off). However it is not clear why a reduced mirror ratio should lead to maxima at the flanks, and in particular why there is a higher intensity at the dawn sector than at noon, where the magnetosphere gets much more compressed than other MLTs. Another problem is that an increase in the size of the loss cone intuitively would lead to a short-lived intensification, lasting long enough for the protons within the loss cone to be emptied; only a few minutes, considering the bounce periods of >1 keV protons. This means that a continuous supply of new protons, or an acceleration of "old" protons is needed to explain the persistent high intensity. An explanation for

both the flank location and the persistence of proton aurora may be increased magnetospheric convection triggered by the compression [*Lukianova*, 2003; *Boudouridis et al.*, 2004, 2005, 2008]. In a steady state, charged particles from the tail cannot enter closed trajectories (the symmetric ring current), and are led around the Earth, to the flanks, with a larger fraction of the protons at the dusk flank due to magnetic field gradient and curvature drifts. At the flanks, the particles encounter a more compressed magnetic field, and some precipitate. Particles that are not on closed trajectories are convected out in the magnetosheath, and are lost from the magnetosphere. *Shue et al.* [2002] studied global electron aurora, driven by convection, located primarily on the flanks (two-cell aurora) during substorm growth phase. The strikingly similar distribution observed in the compression aurora events presented by *Liou et al.* [2007], and the distribution (with opposite asymmetry) in the proton aurora events shown in Figures 1a and 1d, further support the suggestion that convection plays a role in the generation of the aurora during high pressure.

[35] *Anderson and Hamilton* [1993] reported onset (cessation) of electromagnetic ion cyclotron waves (EMIC) in the magnetosphere in response to a contracting (expanding) magnetosphere. These waves are believed to result from the increased energy anisotropy of magnetospheric protons during compression. Interaction with EMIC waves is a possible acceleration mechanism, scattering protons into the loss cone.

[36] While the aurora in Figure 1a was primarily located at the flanks, other events have a maximum located near midnight. This is most clearly seen in Figures 1b and 1c. The different locations may suggest different precipitation mechanisms. One possible mechanism at the nightside is pitch angle scattering in areas with small magnetic field line curvature radius. If a particle with gyro radius r_g encounters an area where $\frac{R_c}{r_g} \lesssim 8$, where R_c is the field line curvature, the particles can be scattered into the loss cone due to violation of the first adiabatic invariant [*Sergeev et al.*, 1983]. This is most likely to happen on the stretched magnetic field lines on the nightside magnetosphere. A study by *Boudouridis et al.* [2008] indicates that a sudden increase in dynamic pressure induces enhanced magnetotail reconnection. The resulting newly closed field lines would have a very small curvature, and so could contribute to pitch angle scattering.

[37] *Lee et al.* [2007] studied the response of ENAs to solar wind dynamic pressure variations, and found that the flux of ENAs exhibits a global increase (decrease) in response to magnetospheric compression (decompression). They claim that this behavior is due to adiabatic energization, with a reinforcing effect from the inward motion of the charged particles to areas where the density of neutral atoms is higher. One of the events studied by *Lee et al.* [2007] was the 18 August 2001 event. The flux of <10 keV energetic neutral hydrogen in this event showed a very similar development as the global average proton aurora intensity (Figure 1c). However the relative increase in <10 keV hydrogen was less than 30%, while the global proton aurora increased by ≈100% from the intensity prior to the dynamic pressure increase. The similarity in the time development of ENA emissions and proton aurora intensity may suggest a production mechanism of common origin during high solar wind dynamic pressure. However the large discrepancy in the relative changes in the 18 August event indicates that adiabatic energization is not sufficient to account for the entire increase in proton aurora without some additional mechanism.

4. Conclusions

[38] We have shown that (1) High solar wind dynamic pressure leads to a persistent higher intensity in global proton aurora, also when the IMF $B_z > 0$. The dynamic pressure induced aurora is most intense on the nightside and flanks. (2) The proton aurora intensity responds promptly to variations in the solar wind dynamic pressure. (3) We observe a dawn–dusk asymmetry, with the highest intensity at dusk, which shows that the protons have been subjected to gradient/curvature drift in the magnetosphere. (4) When IMF $B_z < 0$, the effect of high solar wind dynamic pressure is more dramatic, in terms of proton aurora intensity. Any mechanism, or mechanisms, that aim to explain the dynamic pressure proton aurora must account for the above observations.

[39] The relative changes in solar wind density and velocity during some of the events we have studied suggest that Kelvin–Helmholtz waves on the magnetopause are most likely not the cause for the dynamic pressure induced aurora. Our observations suggest that the precipitation is caused by a mechanism directly connected to the compression of the magnetosphere, possibly: (1) Adiabatic energization, (2) increase in the size of the loss cone due to a decrease in mirror ratio, (3) enhanced magnetospheric convection, (4) a reduced curvature radius on the nightside magnetic field lines, (5) interaction with EMIC waves due to increased energy anisotropy, or a combination of these mechanisms.

[40] Since all these mechanisms may be present simultaneously, further study is needed to differentiate between them. Adiabatic energization, increase in the size of the loss cone, and EMIC waves caused by energy anisotropy depend on the relative increase of the magnetic field strength when the magnetosphere is compressed. It is thus expected that these processes are more important at high latitude magnetic field lines, which experience a larger relative increase during compression. In situ measurements of EMIC wave activity during high dynamic pressure could provide further support for this theory. Since the efficiency of the pitch angle scattering on stretched field lines depends on the mass and energy of the particles, the precipitation of energetic protons should increase much more than the electron precipitation where this is the most important mechanism. This could be tested by in situ measurements of particle precipitation. The flank location (two-cell structure) which was particularly prominent in two of the above events (22 April and 25 October), may be an auroral signature of magnetospheric convection during high dynamic pressure, which has received increasing attention in the recent years. A study of the ionospheric convection in these events could provide a test for this hypothesis.

[41] **Acknowledgments.** This study was financed by the IPY-ICESTAR project, The Research Council of Norway, contract 176045/S30. We thank the IMAGE-FUV team and PI S. Mende for the SI-12 data. The geomagnetic data were provided by the WDC for Geomagnetism, Kyoto.

We thank the ACE SWEPAM and MAG instrument teams and the ACE Science Center for providing the ACE data.

[42] Wolfgang Baumjohann thanks Athanasios Boudouridis and another reviewer for their assistance in evaluating this paper.

References

Anderson, B. J., and D. C. Hamilton (1993), Electromagnetic ion cyclotron waves stimulated by modest magnetospheric compressions, *J. Geophys. Res.*, *98*, 11,369–11,382.

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

Baumjohann, W., and R. A. Treumann (1997), *Advanced Space Plasma Physics*, Imperical College Press, Covent Garden, London.

Boudouridis, A., E. Zesta, L. R. Lyons, P. C. Anderson, and D. Lummerzheim (2003), Effect of solar wind pressure pulses on the size and strength of the auroral oval, *J. Geophys. Res.*, *108*(A4), 8012, doi:10.1029/2002JA009373.

Boudouridis, A., E. Zesta, L. R. Lyons, P. C. Anderson, and D. Lummerzheim (2004), Magnetospheric reconnection driven by solar wind pressure fronts, *Ann. Geophys.*, *22*, 1367–1378.

Boudouridis, A., E. Zesta, L. R. Lyons, P. C. Anderson, and D. Lummerzheim (2005), Enhanced solar wind geoeffectiveness after a sudden increase in dynamic pressure during southward IMF orientation, *J. Geophys. Res.*, *110*, A05214, doi:10.1029/2004JA010704.

Boudouridis, A., E. Zesta, L. R. Lyons, P. C. Anderson, and A. J. Ridley (2008), Temporal evolution of the transpolar potential after a sharp enhancement in solar wind dynamic pressure, *Geophys. Res. Lett.*, *35*, L02101, doi:10.1029/2007GL031766.

Burton, R. K., R. L. McPherron, and C. T. Russell (1975), An empirical relationship between interplanetary conditions and *Dst*, *J. Geophys. Res.*, *81*, 4204–4214.

Coumans, V., J.-C. Gérard, B. Hubert, and M. Meurant (2006), Global auroral proton precipitation observed by IMAGE-FUV: Noon and midnight brightness dependence on solar wind characteristics and IMF orientation, *J. Geophys. Res.*, *111*, A05210, doi:10.1029/2005JA011317.

Craven, J. D., L. A. Frank, C. T. Russel, E. E. Smith, and R. P. Lepping (1986), Global auroral responses to magnetospheric compressions by shocks in solar wind: Two case studies, in *Solar Wind-Magnetospheric Coupling*, edited by Y. Kamide and J. A. Slavin, pp. 377–380, Terra Sci., Tokyo.

Dungey, J. W. (1961), Interplanetary magnetic field and the auroral zones, *Phys. Rev.*, 6, 47–48.

Engebretson, M., K.-H. Glassmeier, M. Stellmacher, W. J. Hughes, and H. Lühr (1998), The dependence of high-latitude Pc5 wave power on solar wind velocity and on the phase of high-speed solar wind streams, *J. Geophys. Res.*, *103*, 26,271–26,283.

Frey, H. U., S. B. Mende, V. Angelopoulos, and E. F. Donovon (2004), Substorm onset observations by IMAGE-FUV, *J. Geophys. Res.*, *109*, A10304, doi:10.1029/2004JA010607.

Frey, H. U., et al. (2003), Summary of quantitative interpretation of IMAGE far ultraviolet auroral data, *Space Sci. Rev.*, *109*, 255–283.

Gérard, J., B. Hubert, M. Meurant, V. I. Shematovich, D. V. Bisikalo, H. Frey, S. Mende, G. R. Gladstone, and C. W. Carlson (2001), Observation of the proton aurora with IMAGE FUV imager and simultaneous ion flux in situ measurements, *J. Geophys. Res.*, *106*, 28,939–28,948.

Hardy, D. A., M. S. Gussenhoven, and D. Brautigam (1989), A statistical model of auroral ion precipitation, *J. Geophys. Res.*, *94*, 370–392.

Lee, D.-Y., and L. R. Lyons (2004), Geosynchronous magnetic field response to solar wind dynamic pressure pulse, *J. Geophys. Res.*, *109*, A04201, doi:10.1029/2003JA010076.

Lee, D.-Y., S. Ohtani, P. C. Brandt, and L. R. Lyons (2007), Energetic neutral atom response to solar wind dynamic pressure enhancements, *J. Geophys. Res.*, *112*, A09210, doi:10.1029/2007JA012399.

Liou, K., P. T. Newell, T. Sotirelis, and C.-I. Meng (2006), Global auroral response to negative pressure impulses, *Geophys. Res. Lett.*, *33*, L11103, doi:10.1029/2006GL025933.

Liou, K., P. T. Newell, J.-H. Shue, C.-I. Meng, Y. Miyashita, H. Kojima, and H. Matsumoto (2007), "Compression aurora": Particle precipitation driven by long-duration high solar wind ram pressure, *J. Geophys. Res.*, *112*, A11216, doi:10.1029/2007JA012443.

Lukianova, R. (2003), Magnetospheric response to sudden changes in solar wind dynamic pressure inferred from polar cap index, *J. Geophys. Res.*, *108*(A12), 1428, doi:10.1029/2002JA009790.

Mende, S. B., H. U. Frey, T. J. Immel, D. G. Mitchell, P. C. Brandt, and J. C. Gérard (2002), Global comparison of magnetospheric ion fluxes and auroral precipitation during a substorm, *Geophys. Res. Lett.*, *29*(12), 1609, doi:10.1029/2001GL014143.

Mende, S. B., H. U. Frey, T. J. Immel, J.-C. Gérard, B. Hubert, and S. A. Fuselier (2003), Global imaging of proton and electron aurorae in the far ultraviolet, *Space Sci. Rev.*, *109*, 211–254.

Mende, S. B., et al. (2000), Far ultraviolet imaging from the IMAGE spacecraft. 3. Spectral imaging of Lyman-α and OI 135.6 nm, *Space Sci. Rev.*, *91*, 287–318.

Meurant, M., J.-C. Gérard, B. Hubert, V. Coumans, C. Blockx, N. Østgaard, and S. B. Mende (2003), Dynamics of global scale electron and proton precipitation induced by a solar wind pressure pulse, *Geophys. Res. Lett.*, *30*(20), 2032, doi:10.1029/2003GL018017.

Meurant, M., J.-C. Gérard, C. Blockx, B. Hubert, and V. Coumans (2004), Propagation of electron and proton shock-induced aurora and the role of the interplanetary magnetic field and solar wind, *J. Geophys. Res.*, *109*, A10210, doi:10.1029/2004JA010453.

Sergeev, V. A., E. M. Sazhina, N. A. Tsyganenko, and F. Søraas (1983), Pitch-angle scattering of energetic protons in the magnetotail current sheet as the dominant source of their isotropic precipitation into the nightside ionosphere, *Planet. Space Sci.*, *31*, 1147–1155.

Shue, J.-H., P. T. Newell, K. Liou, C.-I. Meng, Y. Kamide, and R. P. Lepping (2002), Two-component aurora, *Geophys. Res. Lett.*, *29*(10), 1379, doi:10.1029/2002GL014657.

Søraas, F., K. Aarsnes, K. Oksavik, and D. S. Evans (2002), Ring current intensity estimated from low-altitude proton observations, *J. Geophys. Res.*, *107*(A7), 1149, doi:10.1029/2001JA000123.

Spann, J. F., M. Brittnacher, R. Elsen, G. A. Germany, and G. K. Parks (1998), Initial response and complex polar cap structures of the aurora in response to the January 10, 1997 magnetic cloud, *Geophys. Res. Lett.*, *25*, 2577–2580.

Vorobyev, V. G. (1974), SC-associated effects in auroras, *Geomagn. Aeron.*, *14*, 72–74.

Wanliss, J. A., and K. M. Showalter (2006), High-resolution global storm index: *Dst* versus SYM-H, *J. Geophys. Res.*, *111*, A02202, doi:10.1029/2005JA011034.

Zhou, X.-Y., and B. T. Tsurutani (1999), Rapid intensification and propagation of the dayside aurora: Large scale interplanetary pressure pulses (fast shocks), *Geophys. Res. Lett.*, *26*, 1097–1100.

Zhou, X.-Y., and B. T. Tsurutani (2003), Dawn and dusk auroras caused by gradual, intense solar wind ram pressure events, *J. Atmos. Terr. Phys.*, *66*, 153–160, doi:10.1016/j.jastp.2003.09.008.

Zhou, X.-Y., R. J. Strangeway, P. C. Anderson, D. G. Sibeck, B. T. Tsurutani, G. Haerendel, H. U. Frey, and J. K. Arballo (2003), Shock aurora: FAST and DMSP observations, *J. Geophys. Res.*, *108*(A4), 8019, doi:10.1029/2002JA009701.

K. M. Laundal and N. Østgaard, Department of Physics and Technology, University of Bergen, Allégaten 55, N-5007 Bergen, Norway. (karl.laundal@ift.uib.no; nikolai.ostgaard@ift.uib.no)

Paper II

K. M. Laundal and N. Østgaard (2009)

Asymmetric auroral intensities in the Earth's Northern and Southern hemispheres

Nature Vol. 460, doi:10.1038/nature08154

LETTERS

Asymmetric auroral intensities in the Earth's Northern and Southern hemispheres

K. M. Laundal[1] & N. Østgaard[1]

It is commonly assumed that the aurora borealis (Northern Hemisphere) and aurora australis (Southern Hemisphere) are mirror images of each other because the charged particles causing the aurora follow the magnetic field lines connecting the two hemispheres. The particles are believed to be evenly distributed between the two hemispheres, from the source region in the equatorial plane of the magnetosphere. Although it has been shown that similar auroral features in the opposite hemispheres can be displaced tens of degree in longitude[1,2] and that seasonal effects can cause differences in global intensity[3,4], the overall auroral patterns were still similar. Here we report observations that clearly contradict the common assumption about symmetric aurora: intense spots are seen at dawn in the Northern summer Hemisphere, and at dusk in the Southern winter Hemisphere. The asymmetry is interpreted in terms of inter-hemispheric currents related to seasons, which have been predicted[5,6] but hitherto had not been seen.

On 12 May 2001, the Wideband Imaging Camera (WIC)[7] on the Imager for Magnetopause-to-Aurora Global Exploration (IMAGE) satellite and the Polar Visible Imaging System (VIS) Earth camera[8] recorded completely asymmetric auroral intensity distributions in the two hemispheres, with higher intensity in the northern dawn, and in the southern dusk (Fig. 1). A spatiotemporal analysis (Fig. 2) reveals that the dusk aurora in the Southern Hemisphere was persistent, lasting until 22:30 Universal Time (UT) and that the northern dawn spot was a transient feature lasting about 10 min and occurring twice, at 21:43 and at 22:10 UT.

The charged particles causing the spots at dawn and dusk originate from different regions in the magnetosphere, as sketched in the inset in Fig. 1b. Although we do not have the data to address this properly, the particles creating the two spots are probably accelerated by parallel electric fields and/or Alfven waves. The observed asymmetric intensity distribution may imply differences in acceleration in the two hemispheres.

It is well-known that the magnetic field strength in the two hemispheres is not the same and this may generate differences in auroral intensity between the hemispheres[9]. However, the differences in the magnetic field strength in the Earth's ionosphere where the spots are observed are very small ($<10\%$) and are not likely to explain the large asymmetries we observe.

The difference in ionospheric conductivity is expected to give rise to different auroral intensities in the two hemispheres. If a generator in the magnetosphere drives a current through the ionosphere, low ionospheric conductivity needs to be compensated for by a larger electric field, which accelerates the precipitating particles. An upward current is associated with precipitating electrons. This effect, which favours the dark hemisphere where the conductivity is low, has been observed statistically[3,4]. In this case (Fig. 1b), the Southern Hemisphere was in darkness, and the Northern Hemisphere was sunlit. This implies a lower conductivity in the Southern Hemisphere, and we would expect to see more intense aurora in the south than in the north, which is true for the dusk spot but not for the dawn spot. The asymmetry we observe therefore cannot be fully explained by this mechanism.

We know from previous studies that the orientation of the interplanetary magnetic field can have a strong impact on auroral asymmetries[1,2,10,11], in which the *By* component has been found to be the most important, just as it is in the large-scale convection patterns[12]. The interplanetary magnetic field conditions between 21:20–22:30 UT were dominated by a large positive *Bx* component, as sketched in Fig. 1b. Theoretical considerations[13] suggest that a dominant positive interplanetary magnetic field *Bx* component implies that the solar wind dynamo is more efficient and drives stronger high latitude (region 1) currents in the Southern Hemisphere. This idea is consistent with the persistent dusk spot but cannot explain the transient feature in the north.

Theoretical studies[5,6] predict that conductivity differences will lead to inter-hemispheric, field-aligned currents. However, these currents have never been observed. The theory predicts a pair of inter-hemispheric currents going up (down) at dusk and down (up) at dawn in the southern dark (northern sunlit) hemisphere. Furthermore, as these currents are related to conductivity gradients they should be observed in the vicinity of the terminator, which is consistent with our observations. The persistent character of the asymmetric southern dusk spot is enhanced by the more efficient solar wind dynamo in the Southern Hemisphere for large positive *Bx*. The spot at dawn coincides with a significant increase of tail reconnection in this region, which can explain its transient character. This also suggests that both Alfven waves and plasma flows and shears might be involved in creating the bright aurora. We therefore think that the asymmetry we observe is a confirmation of these predicted currents that to our knowledge has not been observed in both hemispheres before.

All our observational knowledge about interhemispheric differences in the aurora comes from a few ground-based observations with narrow field of view[14], statistical studies from one hemisphere[3,4], and some recent studies based on simultaneous imaging from space[1,2,10,11]. The latter has been possible only twice in history, first with Viking and Dynamic Explorer 1 in the mid-1980s, and then with Polar and IMAGE. Because neither of these satellites were operated to provide conjugate images, the coverage is sporadic, and thus we do not know how common direct observations of interhemispheric currents could be. However, the most intense interhemispherical current systems should be observable from the ground and not only statistically[15].

[1]Department of Physics and Technology, Allegt 55, University of Bergen, N-5007 Bergen, Norway.

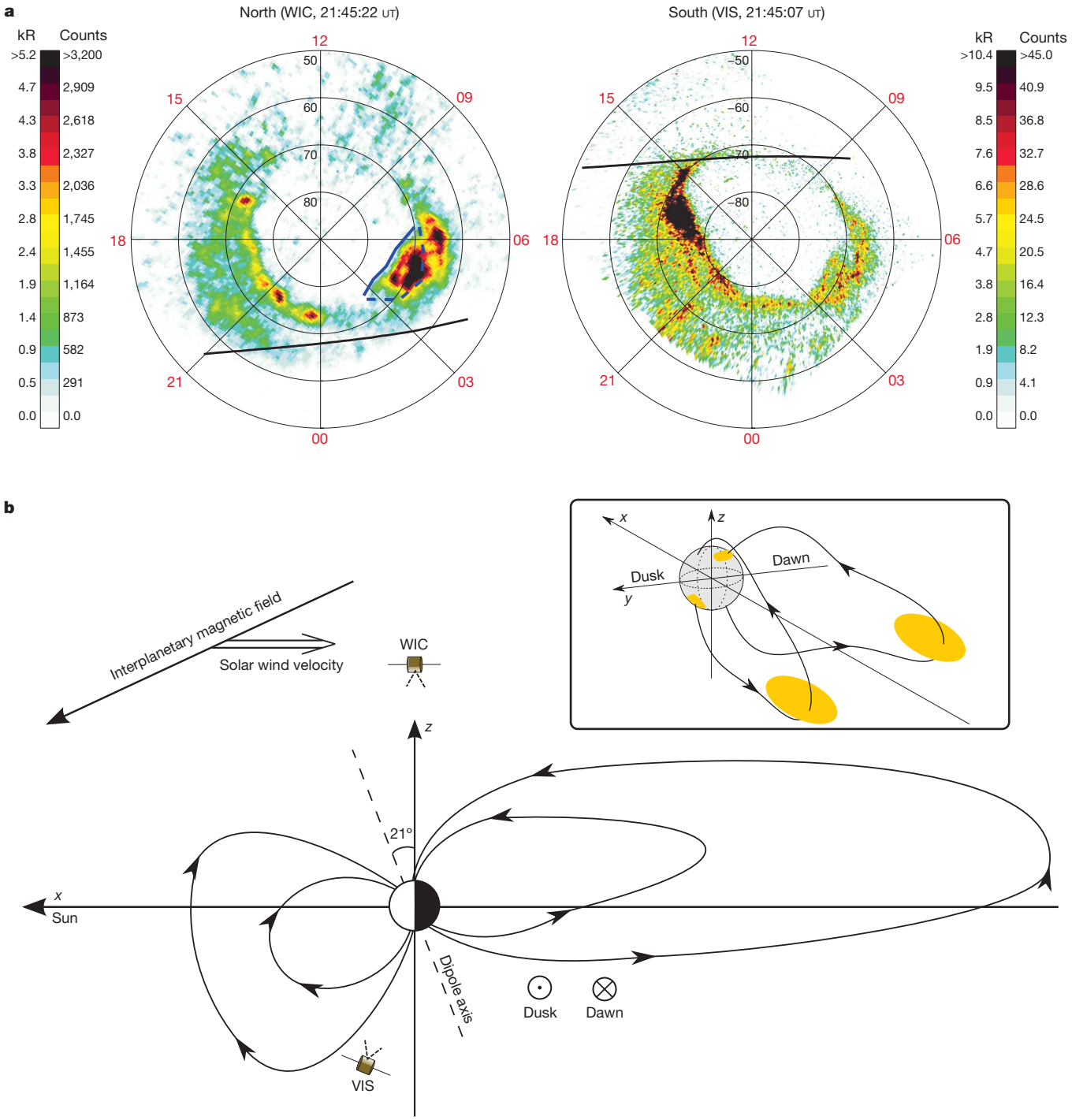

Figure 1 | Simultaneous ultraviolet images of the aurora in both hemispheres. a, Simultaneous images of the aurora at 21:45 UT 12 May 2001, showing completely different intensity distributions. WIC, image of the Northern Hemisphere from the IMAGE satellite's WIC camera; VIS, image of the Southern Hemisphere from the Polar VIS Earth Camera. The colour bars show intensity in both counts and kilo-Rayleigh (kR)[16,17]. The exposure times for WIC and VIS are 10 s and 32.5 s, and the UT at the middle of the exposure is shown above the images; the cadences for WIC and VIS are 2 min and 1 min. The images are shown in magnetic apex coordinates looking down at the magnetic North Pole (and through the Earth for the southern image), with the direction of the Sun up, dawn to right, and dusk to the left. Sunlight-induced dayglow emissions have been subtracted from the images. The black curves indicate the sunlight terminator. The blue lines show the location of the poleward boundary of the aurora at 21:37 (dashed) and 21:45 (solid) UT. This means that in less than 10 min the boundary, which is also the boundary between open and closed magnetic field lines, moved about 8° magnetic latitude towards the pole, signifying a local increase of tail reconnection. **b**, A conceptual presentation showing the seasonal conditions and geometry of the magnetosphere, and the orientation of the interplanetary magnetic field, measured by ACE ($|By|$ was less than 1 nT). The interplanetary magnetic field moves with the solar wind velocity. The Earth's magnetic dipole axis (shown with a dashed line) was tilted at an angle of 21° towards the Sun at 21:45 UT. The coordinate system in this sketch is Geocentric Solar Magnetospheric, with the x axis pointing towards the Sun, the y axis perpendicular to the Earth's magnetic dipole axis (and to the x axis) and the z axis fulfilling the system. The inset illustrates that the spots at northern dawn and southern dusk originate from completely different regions in the magnetosphere.

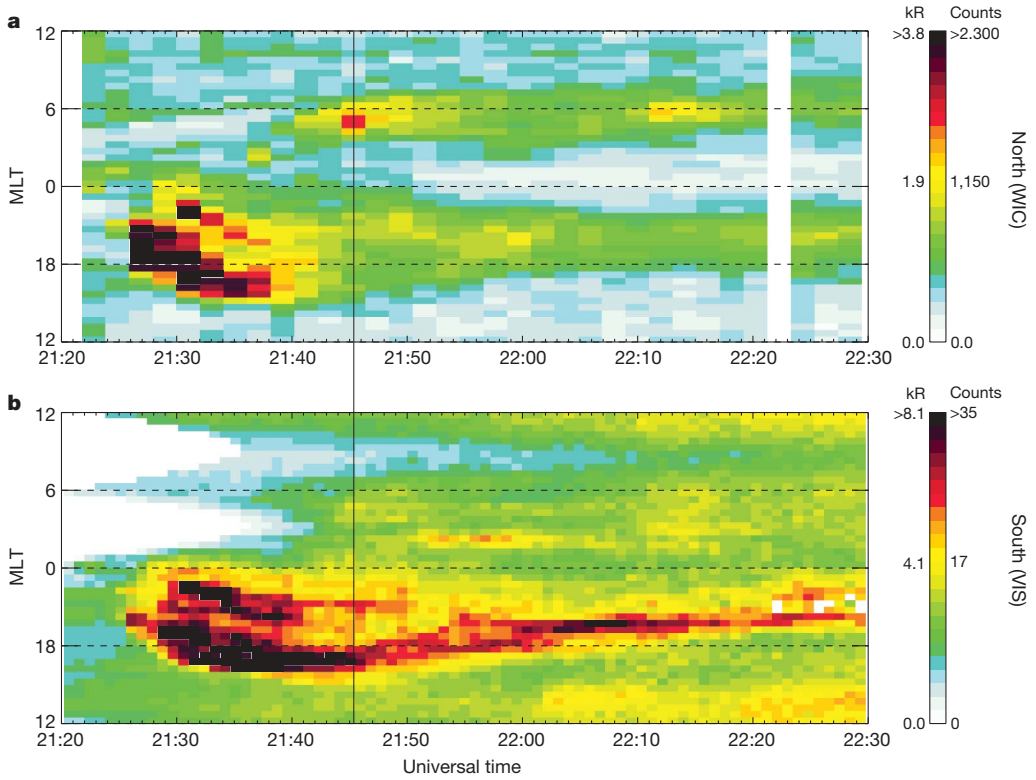

Figure 2 | Spatio-temporal distribution of auroral intensity in both hemispheres. The intensity (integrated between 63° and 85° magnetic latitude) is shown as a function of UT (*x* axis) and Magnetic Local Time (MLT, *y* axis), in the Northern (**a**) and Southern (**b**) hemispheres. Panel **b** reveals that the spot at southern dusk coincides with the western edge of an auroral substorm, seen clearly in both hemispheres between 21:25 and 21:40 UT. The count rate in both cameras depends on the energy flux of precipitating particles. However, the cameras have different sensitivity and pass-band, giving different counts and kR in the two hemispheres[16,17]. The VIS camera (124–149 nm) is most sensitive to the atomic oxygen (O I) line at 130.4 nm (ref. 16), while the WIC camera (140–190 nm) is sensitive to molecular nitrogen (N_2) emissions in the Lyman–Birge–Hopfield (LBH) band and a few emission lines of atomic nitrogen[7,17]. An important difference in energy flux dependence between the two cameras is that emissions in the LBH band are more reduced by molecular oxygen (O_2) absorption. Heating of the atmosphere by sunlight will affect the scale height of N_2 more than of O_2, possibly causing a relatively higher intensity in the sunlit Northern Hemisphere. This means that the observed differences in the dusk sector could be underestimated, and the difference at dawn may be exaggerated. Another difference is that the observed emissions decrease with increasing mean electron energy, but this effect is stronger for the O I line (VIS). This means that high electron energy could contribute to the asymmetry at dusk. However, the large relative difference observed at 6 MLT, combined with the high absolute intensity in the north, >5.2 kR in Fig. 1a (with a peak intensity of 10 kR), leads us to conclude that the observed asymmetry is not an instrumental effect.

Received 17 March; accepted 21 May 2009.

1. Østgaard, N. *et al.* Observations and model predictions of auroral substorm asymmetries in the conjugate hemispheres. *Geophys. Res. Lett.* **32,** L05111, doi:10.1029/2004GL022166 (2005).
2. Østgaard, N., Mende, S. B., Frey, H. U., Sigwarth, J. B., Aasnes, A. & Weygand, J. Auroral conjugacy studies based on global imaging. *J. Atmos. Solar-Terr. Phys.* (special issue) **69,** 249–255 (2007).
3. Newell, P. T., Meng, C.-I. & Lyons, K. M. Suppression of discrete aurorae by sunlight. *Nature* **381,** 766–767 (1996).
4. Liou, K., Newell, P. T. & Meng, C.-I. Brittnacher, M. & Parks, G. Seasonal effects on auroral particle acceleration and precipitation. *J. Geophys. Res.* **106,** 5531–5542 (2001).
5. Richmond, A. D. & Roble, R. G. Electrodynamic effects of therospheric winds from the NCAR thermospheric general circulation model. *J. Geophys. Res.* **92,** 12365–12376 (1987).
6. Benkevich, L., Lyatsky, W. & Cogger, L. L. Field-aligned currents between the conjugate hemispheres. *J. Geophys. Res.* **105,** 27727–27737 (2000).
7. Mende, S. B. *et al.* Far ultraviolet imaging from the IMAGE spacecraft. 2. Wideband FUV imaging. *Space Sci. Rev.* **91,** 271–285 (2000).
8. Frank, L. A. *et al.* The Visible Imaging System (VIS) for the Polar spacecraft. *Space Sci. Rev.* **71,** 297–328 (1995).
9. Stenbaek-Nielsen, H. C., Wescott, T. N., Davis, T. N. & Peterson, R. W. Auroral intensity differences at conjugate points. *J. Geophys. Res.* **78,** 659–671 (1973).
10. Fillingim, M. O., Parks, G. K., Frey, H. U., Immel, T. J. & Mende, S. B. Hemispheric asymmetry of the afternoon electron aurora. *Geophys. Res. Lett.* **32,** L03113, doi:10.1029/2004GL021635 (2005).
11. Stubbs, T. J. & Vondrak, R. R. Østgaard, N. Sigwarth, J. B. & Frank, L. A. Simultaneous observations of the auroral oval in both hemispheres under varying conditions. *Geophys. Res. Lett.* **32,** L03103, doi:10.1029/2004GL021199 (2005).
12. Heppner, J. P. & Meynard, N. C. Empirical high-latitude electric field models. *J. Geophys. Res.* **92,** 4467–4489 (1987).
13. Cowley, S. W. H. Asymmetry effects associated with the X-component of the IMF in a magnettically open magnetosphere. *Planet. Space Sci.* **29,** 809–818 (1981).
14. Sato, N., Nagaoka, T., Hashimoto, K. & Saemundsson, T. Conjugacy of isolated auroral arcs and non-conjugate auroral break-ups. *J. Geophys. Res.* **103,** 11641–11652 (1998).
15. Lyatskaya, S., Lyatsky, W. & Khazanov, G. V. Relationship between substorm activity and magnetic disturbances in the two polar caps. *Geophys. Res. Lett.* **35,** L20104, doi:10.1029/2008GL035187 (2008).
16. Frank, L. A. & Sigwarth, J. B. Simultaneous images of the northern and southern auroras from the Polar spacecraft: an auroral substorm. *J. Geophys. Res.* **108,** 8015, doi:10.1029/2002JA009356 (2003).
17. Frey, H. U. *et al.* Summary of quantitative interpretation of IMAGE Far Ultraviolet auroral data. *Space Sci. Rev.* **109,** 255–283 (2003).

Acknowledgements We are indebted to the IMAGE and Polar teams for the design and successful operations of the two missions. We especially thank S. B. Mende for the use of IMAGE FUV WIC data and J. B. Sigwarth for the use of Polar VIS Earth Camera data. This study was supported by the Norwegian Research Council, through the IPY-ICESTAR project 176045/S30

Author Contributions N.Ø is responsible for the project planning, K.M.L. is responsible for the data processing. The data were identified and analysed by K.M.L. and N.Ø. The manuscript was written by K.M.L. and N.Ø.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to N.Ø. (nikolai.ostgaard@uib.no).

Paper III

K. M. Laundal, N. Østgaard, K. Snekvik and H. U. Frey (2010)

Inter-hemispheric observations of emerging polar cap asymmetries

Journal of Geophysical Research Vol. 115, doi:10.1029/2009JA015160

Interhemispheric observations of emerging polar cap asymmetries

K. M. Laundal,[1] N. Østgaard,[1] K. Snekvik,[2] and H. U. Frey[3]

Received 3 December 2009; revised 2 March 2010; accepted 1 April 2010; published 30 July 2010.

[1] In this paper we use simultaneous global UV images of the aurora in the two hemispheres to study differences in the polar cap boundary location. We show that the northern and southern auroral ovals circumvent the same amount of magnetic flux, providing additional evidence that the poleward boundary of the aurora coincides with the open/closed field line boundary. During a period of significant flux closure, large asymmetries in the polar cap boundaries developed between the hemispheres. The asymmetry was strongest in the regions where the polar caps contracted the most, suggesting that emerging interhemispheric polar cap asymmetries is an intrinsic phenomenon during substorm expansions, when magnetic flux closes rapidly in the tail. Utilizing the prolonged surveillance of the open/closed boundary location, we show that the growing asymmetries can be accounted for by differences in the ionospheric convection in the two hemispheres. The observations suggest that the differences in convection were due to seasonal differences between the hemispheres, and that the summer hemisphere responded more promptly to changes in magnetospheric convection than the winter hemisphere.

Citation: Laundal, K. M., N. Østgaard, K. Snekvik, and H. U. Frey (2010), Interhemispheric observations of emerging polar cap asymmetries, *J. Geophys. Res.*, *115*, A07230, doi:10.1029/2009JA015160.

1. Introduction

[2] The Earth's magnetosphere is the cavity in space which is dominated by the magnetic field generated in the planet's interior. Its shape differs from a dipole-like field most notably by a long tail which extends to $>100R_E$ (Earth radii) on the nightside of the Earth. This tail is the product of interaction with the interplanetary magnetic field (IMF), which is transported from the sun by the solar wind plasma. When the IMF and the Earth's field have opposite orientations, the field lines can merge on the dayside in a process called reconnection, and form open field lines. Open field lines are then transported, by the influence of the solar wind momentum, anti-sunward, and add to the tail on the nightside. When open field lines with foot-points in opposite hemispheres meet in the tail, they can reconnect, forming a closed field line. Owing to the stretched character of the newly closed field lines, the plasma now convects back to the dayside where it can once again merge with the IMF. This so-called Dungey cycle [*Dungey*, 1961] thus describes the circulation of plasma within the magnetosphere.

[3] The regions at the surface of the Earth which are threaded by open field lines are called the polar cap. The polar caps are very often seen to be encircled by a belt of precipitating particles (~keV), which ionize and excite atmospheric constituents to form the aurora borealis and aurora australis. Global auroral images can therefore be used to observe the shape and size of the polar caps, since it can be identified as the dim region surrounded by the bright aurora (the main exception to this is the polar cusp aurora which is located on open field lines, and the theta aurora). This technique has been substantiated [*Carbary et al.*, 2003; *Boakes et al.*, 2008] and utilized extensively in recent years [*Brittnacher et al.*, 1999; *Mende et al.*, 2003a, 2003b; *Milan et al.*, 2003, 2007, 2008, 2009; *Lam et al.*, 2006; *Hubert et al.*, 2006, 2008; *Boakes et al.*, 2009].

[4] Among the main findings established by these studies are that the polar cap expansion (dayside reconnection) and contraction (nightside reconnection) are quasi-uncoupled. The expansion of the polar cap happens mainly when the IMF has a southward component ($B_z < 0$ nT), whereas the contraction of the oval can happen almost explosively during the expansion phase of a substorm [e.g., *Milan et al.*, 2007]. A substorm expansion can be recognized in global auroral images by a sudden local brightening in auroral luminosity, followed by a rapid azimuthal and latitudinal expansion, lasting typically for a few tens of minutes. The quasi-uncoupled dayside and nightside reconnection is called the expanding/contracting polar cap paradigm, and was first suggested by *Siscoe and Huang* [1985], and expanded upon by, e.g., *Lockwood et al.* [1990] and *Cowley and Lockwood* [1992].

[5] The observational studies of the polar cap boundary cited above all focus on one hemisphere. The implicit assumption is that, when an appropriate magnetic coordinate

[1]Department of Physics and Technology, University of Bergen, Bergen, Norway.
[2]Finnish Meteorological Institute, Helsinki, Finland.
[3]Space Sciences Laboratory, University of California, Berkeley, California, USA.

0148-0227/10/2009JA015160

system is used (the two systems which are most-often used are AACGM and, as in the present study, Apex [*Richmond*, 1995], which are similarly defined), the polar caps in the two hemispheres are approximately similar. This assumption is largely necessitated by the lack of simultaneous measurements from both hemispheres. The assumption is certainly true for the total magnetic flux content, but the open/closed field line boundaries are not necessarily identical. Conjugate observations of the local auroral brightening at substorm onset have shown that large spatial deviations (tens of degrees in longitude) may happen, evidencing a temporary asymmetry in the field line about the equatorial plane [*Østgaard et al.*, 2004, 2005b]. *Stubbs et al.* [2005] also employed simultaneous UV images to find that the entire polar caps can be displaced from each other. The displacement seen in these studies was consistent with distorted magnetic field lines, the distortion being in the same direction as the IMF (in particular the y component). Statistical studies of the substorm onset location in relation to the IMF have confirmed these findings [*Liou et al.*, 2001; *Østgaard et al.*, 2007]. The IMF dependent perturbation in the magnetic field has also been demonstrated directly by in-situ measurements [e.g., *Wing et al.*, 1995]. Conjugate studies have also been undertaken, using all-sky cameras on the ground [e.g., *Sato et al.*, 2005] or on air-planes [*Stenbaek-Nielsen and Otto*, 1997, and references therein], and a combination of observations from ground and space [*Frey et al.*, 1999; *Vorobyev et al.*, 2001]. *Vorobyev et al.* [2001] focused on the poleward boundary of the nightside aurora, which they found to be displaced poleward in the Northern (Southern) Hemisphere compared to the Southern (Northern) Hemisphere when the IMF orientation was in the Parker spiral sector $B_x < 0$, $B_y > 0$ ($B_x > 0$, $B_y < 0$).

[6] In the present paper, we take advantage of two fortuitous constellations of the IMAGE and Polar satellites, both equipped with UV imagers, enabling simultaneous observations of the aurora in the Northern and Southern Hemispheres. The main focus is on a strong substorm occurring on 12 May 2001. In this event, we observe highly variable interhemispheric asymmetries in the polar caps, arising in conjunction with significant flux closure.

[7] The method which we employ is described in the next section. In section 3 we present the observations. The findings are discussed in section 4. Section 5 concludes the paper.

2. Method

[8] The IMAGE satellite [*Burch*, 2000] was launched on 25 March 2000 into an elliptical orbit, which after one year in operation had its apogee (~$7R_E$) over the north pole. The FUV instrument on this satellite included three imagers: The Wideband Imaging Camera (WIC), and two Spectropgrahic Imagers, SI-12 and SI-13. Due to their higher sensitivity and spatial resolution, we use WIC images in this study. WIC [*Mende et al.*, 2000] provided images in the LBH wavelength band (140–190 nm). Being mounted on the spinning satellite, and viewing radially outward, it produced images every 123 seconds (the satellite spin period), and had 10 seconds integration time.

[9] The Polar satellite was launched on 24 February 1996. In 2001, precession of its elliptical orbit had enabled prolonged observation of the Southern Hemisphere. In the present study we use the VIS Earth camera [*Frank et al.*, 1995], which observed UV emissions in the wavelength range of 124–149 nm. VIS Earth was mounted on a despun platform, and in the images used here the integration time was 32.5 s, with a nominal cadence of 54 s.

[10] The intensity observed by VIS Earth is dominated by the OI line at 130.4 nm (83% during quiet conditions according to *Frank and Sigwarth* [2003]). Most of the remaining signal stems from the OI emission at 135.6 nm and from the LBH band. Since WIC is only marginally sensitive to the OI lines, some differences are expected in the response of the two cameras to similar auroral features. However, as long as the two cameras observe a prominent poleward boundary in the aurora, these differences will not affect the result in the present study, even if the absolute intensities may differ.

[11] To calculate the polar cap flux content, an accurate determination of the poleward boundary of the aurora is needed at all local times, since this boundary is a proxy for the open/closed field line boundary (OCB). Earlier studies have used an automated routine for this purpose. Possible methods include a functional fit (a gaussian is often used), or a threshold intensity, often accompanied by automatic tests of how well these methods work. In regions where no valid boundary is obtained, interpolation is necessary, if the total magnetic flux content in the polar cap is to be calculated.

[12] Having experimented with various automated routines for determining the boundary, we did not find any single method which worked well in both hemispheres. This is likely because WIC and VIS Earth have significantly different count rates and signal to noise ratios. In the May 2001 event, additional complications arose from dayglow contamination in the northern images. The dayglow was removed using a functional fit to the background counts, but residual Poisson noise was still evident on the dayside part of the image. In many cases, automatic methods also tend to fail in regions where manual inspection clearly shows that a boundary between the background and auroral luminosity is well defined. Since the comparison between the cameras should be based on a common method, we therefore determined the boundaries by eye. This was done by separating the image into 1 hour wide MLT sectors, and plotting the intensity as a function of latitude. The boundary was placed where the intensity profile transition from background to aurora.

[13] Figure 1a shows this method applied on a best case scenario (solid) and a worst case scenario (dotted-dashed). These intensity profiles are from 20 to 21 MLT and 11 to 12 MLT in the WIC image taken 21:26:58 (Figure 3a). On the night side, the boundary can be determined accurately, because of the sharp transition from aurora to background seen at 71°. For the dayside intensity profile, an accurate determination is impossible, neither by eye, nor by any automatic method that we are aware of. These kinds of boundaries were therefore determined by looking at neighboring regions, assuming the OCB to be fairly uniform, or by looking at later images in which the dayside aurora could be recognized, assuming the boundary not to vary much in time. The latter assumption is justified at the dayside by the concurrent stable solar wind and IMF conditions. The two hemispheres were considered independently. All manually

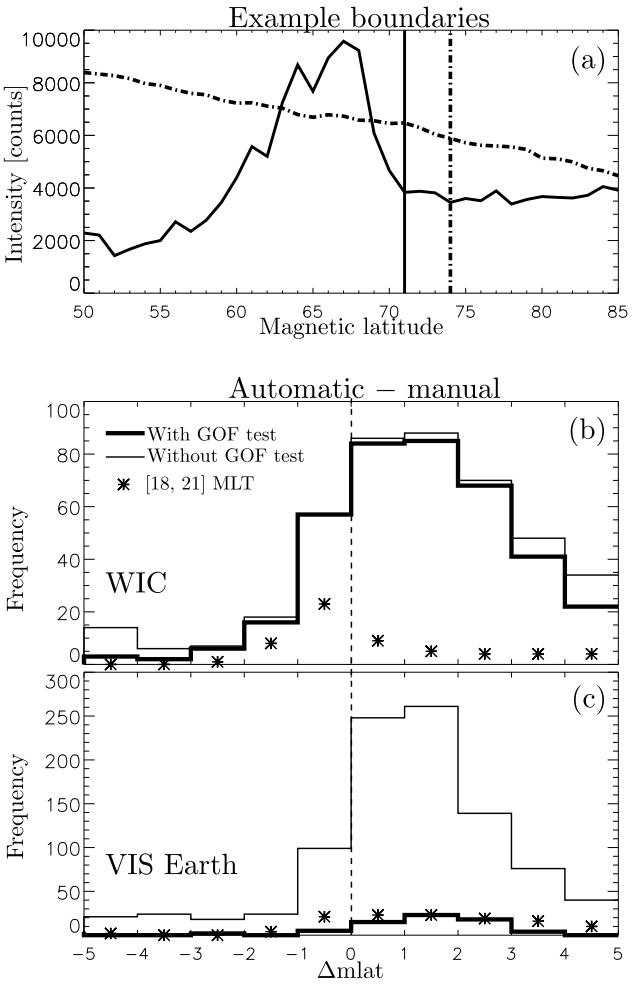

Figure 1. (a) Example of manual boundary determination applied on two intensity profiles, from the WIC image at 21:26:58 UT, 12 May 2001 (Figure 3a). The intensity profiles are from 20–21 MLT (solid) and 11–12 MLT (dotted-dashed). (b and c) The difference (Δ mlat in degrees) between boundaries determined by gaussian fit [*Carbary et al.*, 2003], and our manually determined boundaries, in WIC (Figure 1b) and VIS Earth (Figure 1c) in the 12 May 2001 event. The frequency denotes the number of cases in which the difference was within the bins defined at the *x* axis. The thick lines compares boundaries for which the gaussian fit passed the *Carbary et al.* [2003] goodness of fit (GOF) tests. The thin lines are without GOF tests. Asterisks denote the frequency for boundaries in the [18–21] MLT sector (without GOF tests), which will be studied in more detail in section 4.2.

determined boundaries were rounded to the closest integer. The prominent boundaries were assigned an accuracy of ±1°, and the less clear boundaries were assigned an accuracy of ±2°. In the first images in the Southern Hemisphere, the oval was outside the field of view in a small region post-midnight. In these cases, the accuracy was set to ±3°. We used low altitude in-situ particle precipitation measurements from DMSP F12, F13, F14 and F15, as well as NOAA-15 (Table 1 shows these boundaries for the May 2001 event) to determine the boundary accurately when and where these satellites crossed the OCB.

[14] In Figure 1b and 1c, we compare our manually determined boundaries to boundaries determined by fitting the intensity profile to a gaussian, plus a quadratic polynomial which most often aligns with the background [*Carbary et al.*, 2003]. The thick lines show the frequency of differences between using the *Carbary et al.* [2003] method (including a goodness of fit (GOF) test), and our manual method. We see a fairly symmetrical distribution in the case of WIC, centered at 1–2°, indicating that we place our boundaries slightly equatorward compared to the *Carbary et al.* [2003] method. For VIS Earth, very few of the boundaries passed the *Carbary et al.* [2003] GOF tests. The thin lines show the difference when no test was applied on the fitted function. Many of these boundaries were still placed quite close to our manually determined boundaries, although this distribution is also centered at 1–2°. The seemingly common bias between manually determined boundaries and boundaries determined by gaussian fits, indicates that we have treated the two hemispheres equally, and that the resulting boundaries are comparable. The asterisks show the difference of the two methods in the [18–21] MLT region (without GOF tests). This region had particularly sharp poleward boundaries, and will be discussed in more detail in section 4.2. The automatic and manual methods seem to give similar results in this region, in the WIC images. The still present deviation seen in VIS Earth, is likely caused by a slant viewing angle in the equatorward edge of the auroral oval, distorting the fitted function. The slant viewing angle did however not affect the higher latitudes, and the boundary could be easily identified.

[15] We have also compared the manually detected OCBs to boundaries which were determined as the first point, from the pole, in which the intensity exceeds some threshold value. In 68% of the cases, our boundaries in the WIC images are within 2° of the boundaries determined by a threshold value of 800 counts. In VIS Earth, 83% of the boundaries were within 2° of the boundaries using a 25 counts threshold. The mentioned thresholds were the values giving the highest number of matches between the two methods. In the [18–21] MLT sector, these threshold boundaries matched the manually determined boundaries (within 2°) in 90% of the cases in WIC, and 93% of the cases in VIS Earth.

Table 1. Poleward Boundaries of Precipitation, as Seen by the DMSP and NOAA Satellites During the Time When WIC and VIS Observed the Entire Ovals in Both Hemispheres, During the 12 May 2001 Event

Satellite	UT	Mlat	MLT
DMSP-F15	21:32:30	−74.8°	20:50
NOAA-15	21:32:52	−70.6°	07:10
DMSP-F13	21:37:20	75.2°	18:00
NOAA-15	21:41:15	−79.5°	20:00
DMSP-F13	21:47:20	71.5°	07:50
DMSP-F14	21:56:49	78.9°	20:50
DMSP-F12	22:07:47	78.6°	20:00
DMSP-F15	22:16:45	77.0°	21:50

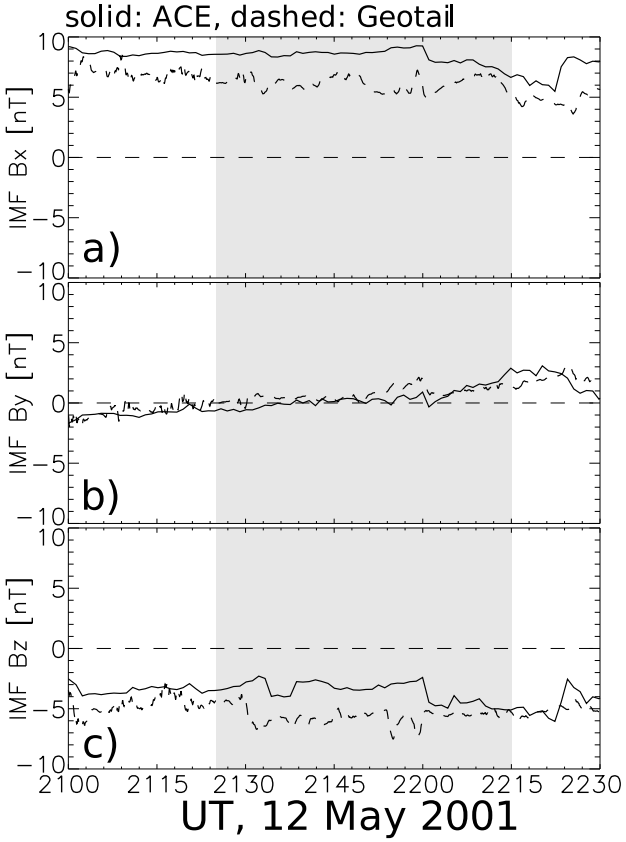

Figure 2. (a) IMF B_x, (b) B_y, (c) B_z, measured by ACE (solid) and Geotail (dashed) on 12 May 2001. The ACE measurements have been time shifted from its position in orbit around the L1 point to the Geotail position at $x_{GSE} = 12R_E$. The shaded region shows when IMAGE and Polar provided global coverage of the polar caps.

[16] Since some degree of subjectivity is impossible to avoid when the boundaries are determined manually, we include as auxiliary material print-outs of all images and boundaries from the 12 May 2001 event, for the sake of verifiability.[1]

[17] Figures 2a–2c show the IMF B_x, B_y, and B_z component during the May 2001 event. The dashed curves are measurements by Geotail, which was located at $(x, y, z)_{GSM} = (12, 8, 2)\ R_E$, and the solid curves are ACE measurements, time shifted to the Geotail position. The time shift, 52 minutes, was determined by a minimum variance analysis, and confirmed by matching a subsequent pressure increase (not shown) with its response in the geosynchronous magnetic field, measured by GOES-8. This is 15 minutes longer than $\Delta t = \Delta x/v_x$, which means that the phase fronts were oblique to the propagation [e.g., see *Haaland et al.*, 2007]. The solar wind speed and density (not shown) were constant at ≈620 km/s and ≈5 cm^{-3}, respectively, corresponding to a dynamic pressure slightly in excess of 3 nPa. The main point of Figure 2 is to emphasize the stable IMF orientation, with negative B_z, strongly positive B_x and almost zero B_y. The shaded region corresponds to the period with global coverage of the aurora in both hemispheres.

3. Observations

[18] Figure 3 shows five pairs of simultaneous images in the Northern (upper row) and Southern (middle row) Hemispheres, starting at 21:27 UT, and spanning almost 40 minutes. The OCB is indicated by a red (black) curve in the Northern (Southern) Hemisphere. The ionospheric OCBs in the two hemispheres can be compared in the lower row. The UT at middle of exposure is shown above the images. All five pairs have some overlapping integration time. All images were taken during substorm expansion and recovery. The onset was seen in the Northern Hemisphere at 21 MLT at 21:24:55 UT (not shown).

[19] The first pair of images (Figure 3a) were taken a few minutes into the substorm expansion phase. At this time, the boundaries were quite circular and symmetric in the two hemispheres. As revealed in the images taken ten minutes later (Figure 3b), a large asymmetry between the hemispheres emerged as the boundaries propagated poleward. In the pre-midnight sector, the southern boundary was ≈5° closer to the pole than its northern counterpart, while neighboring regions (15 MLT and post-midnight) had an opposite asymmetry. Figure 3c shows that the asymmetry six minutes later, though still present, was more uniformly distributed in local time. At this time, the OCBs underwent rapid poleward propagation around 3 MLT. Eight minutes later (Figure 3d), the post-midnight region exhibited large asymmetries, the southern boundary being closer to the pole than the northern boundary duskward of 4 MLT, and further from the pole dawnward of 4 MLT. 12 minutes later (Figure 3e), the sense of the asymmetry was largely similar, demonstrating that the emerging asymmetries were relatively stable. The five pairs of images also show that the initial circular shape of the polar cap was better retained in the Northern Hemisphere when compared to the Southern Hemisphere.

[20] The images in Figure 3c reveal large differences in intensity distributions, as reported by *Laundal and Østgaard* [2009]. In section 4.5 we discuss how the non-conjugate spots may be interrelated with the spatial asymmetries in the magnetic field signified by the OCBs.

3.1. Interpretation in Terms of Open Magnetic Flux

[21] Previous studies have used auroral images to calculate the amount of open magnetospheric flux [e.g., *Milan et al.*, 2003; *Hubert et al.*, 2008]. Since $\nabla \cdot \mathbf{B} = 0$, the magnetic flux entering through the surface of the Earth is equal to the flux which is leaving the surface. Since the closed magnetic field lines by definition extend the same amount of magnetic flux in both hemispheres, the open field lines must also contain exactly the same amount of magnetic flux in the two hemispheres. Hence, conjugate global images can be used as a test of the hypothesis that the auroral oval circumvents all open magnetospheric flux: If the amount of flux in the two measured polar caps is different, the hypothesis or the method is wrong. The open flux is given as the surface integral $\int B_\perp dA$, where $B_\perp$ is the Earth's magnetic field perpendicular to the surface at ionospheric altitudes. We use the International Geomagnetic Reference

[1] Auxiliary materials are available in the HTML. doi:10.1029/2009JA015160.

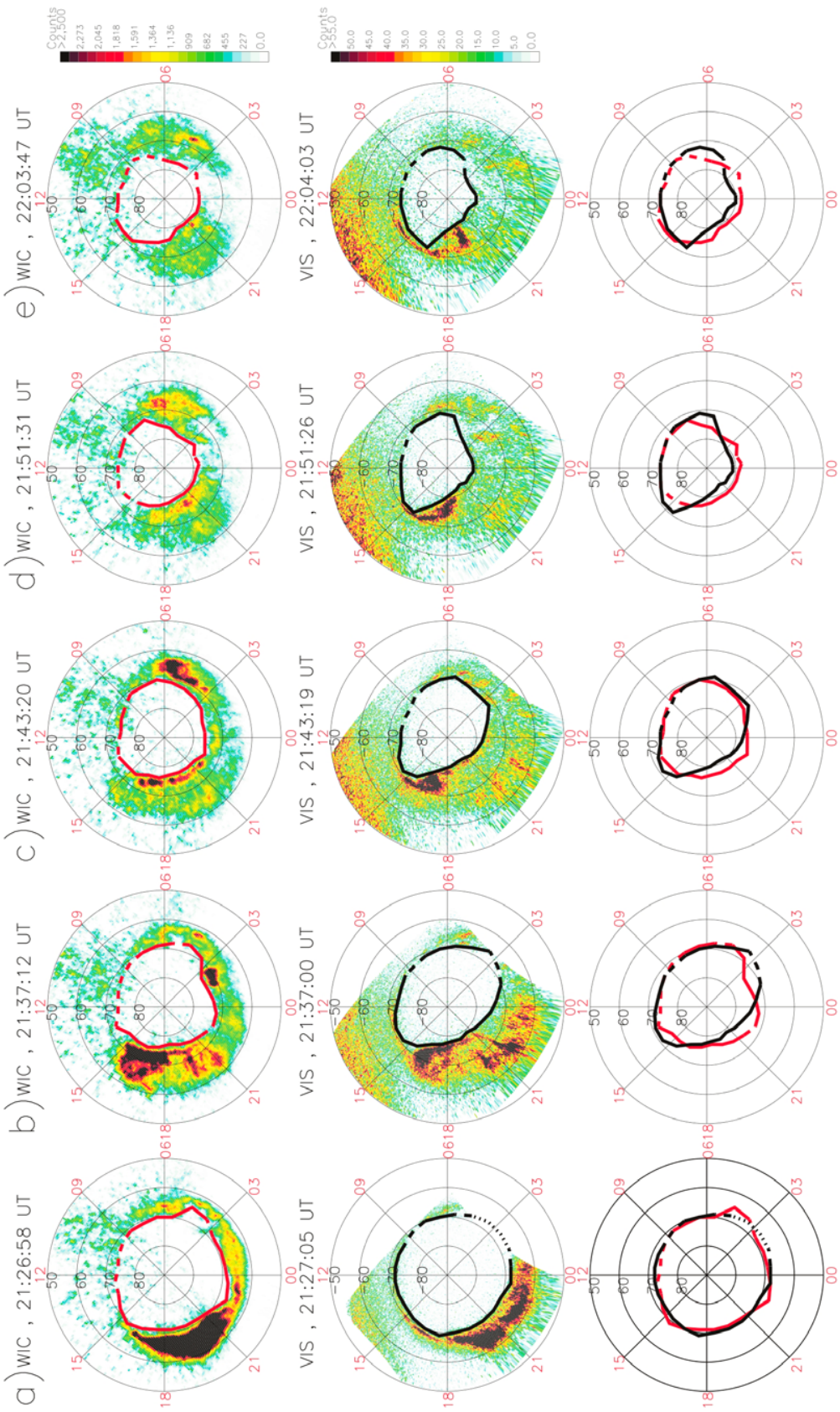

Figure 3. (a–e) Simultaneous images from the (top) Northern and (middle) Southern Hemisphere, with (bottom) the poleward boundary of the aurora in red (north) and black (south) shown in the images, and together for comparison. The UT of the middle of exposure is shown above the images. OCBs indicated by a solid line were clear and possible to determine independently. Dashed boundaries indicate less clear boundaries, and dotted lines are used where we have a lack of coverage (this was only the case for the post-midnight sector in the first images from the Southern Hemisphere).

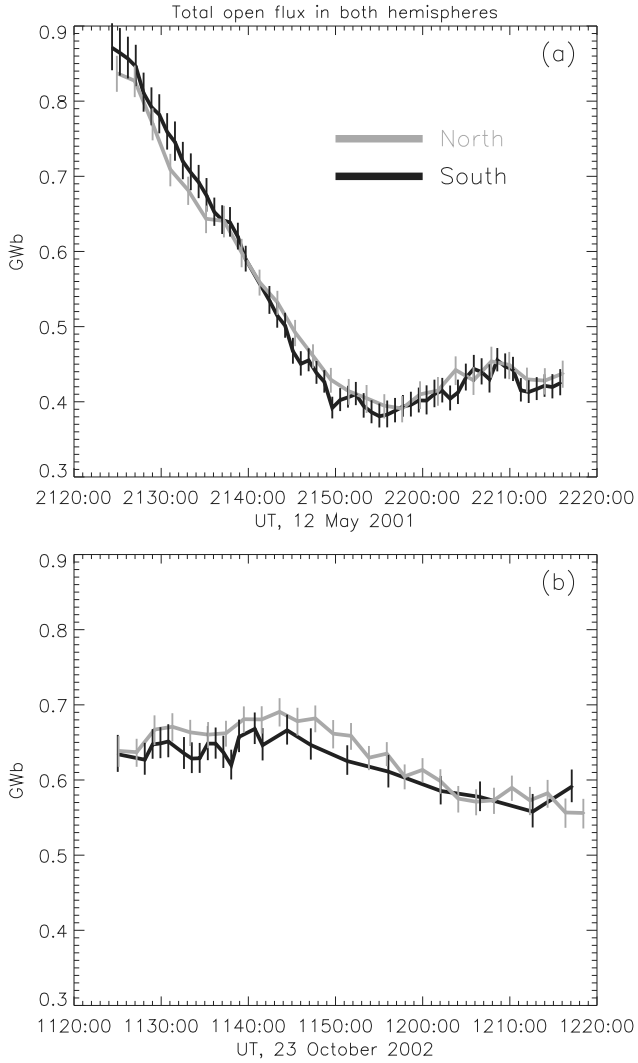

Figure 4. Total open flux in both hemispheres (a) 12 May 2001 and (b) 23 October 2002. The flux in the Northern (Southern) Hemisphere is shown in grey (black).

Field (IGRF) at 130 km altitude (the assumed emission height in the auroral images) for this purpose. The integration was performed in geographic coordinates.

[22] Figure 4 shows the total open flux in two hemispheres for the 12 May 2001 (Figure 4a) event, and from 23 October 2002 (Figure 4b). The latter event was studied in more detail by *Stubbs et al.* [2005]. The total open flux content can be written as a sum of the flux content in the 24 sectors, $\Phi = \sum_1^{24} \Phi_i$. With the assigned errors of 1, 2, or 3 degrees, each sector is associated with a corresponding error in the flux, $e^{\pm}_{\Phi,i}$, where superscripts + and − correspond to the error equatorward and poleward of the boundary, respectively. The total error of the flux was then calculated as $e^{\pm}_{\Phi} = \sqrt{\sum_1^{24}\left(e^{\pm}_{\Phi,i}\right)^2}$. Figure 4 shows that the measured amount of polar cap flux in the two hemispheres are similar for both events, within the applied error bars (i.e. they deviate by less than ≈0.05 GWb). This provides new evidence that the dim region encircled by the auroral oval, as seen by global FUV imaging, is indeed the ionospheric footprint of the open flux in the magnetosphere.

[23] We also notice that in the 12 May 2001 event, the initial amount of open flux in the magnetosphere was 0.85 GWb. *Milan et al.* [2008] showed that this number is high (the most probable value for open magnetic flux was ≈0.4 GWb in their study). According to *Milan et al.* [2009], the flux closure rate during substorms is high when the initial open flux content is high. This is also observed in the 12 May 2001 event, when ≈0.45 GWb closed in 25 minutes. This change in magnetic flux corresponds to a net mean closure rate of 300 kWb/s, or 300 kV.

[24] In Figure 5, we look in more detail at where, in terms of magnetic local time, the closing of flux took place in the 12 May 2001 event. Figure 5a–5d show the open flux content in six hours wide MLT sectors. Figure 5 clearly shows that in the first 15 minutes of the substorm, the closing of flux happened at the pre-midnight sector (d), followed by a steep drop in open flux content in the post-midnight sector (Figure 5a) during the following 10 minutes, in agreement with the observations of the OCB in Figure 3. This is consistent with tail reconnection happening in two steps, first at dusk and then at dawn, or that the X-line propagated from dusk to dawn. Figures 5a and 5d also show that the interhemispheric asymmetry appeared in conjunction with flux closure. More specifically, in the pre-midnight region (Figure 5d) an asymmetry appeared during the first 15 minutes, because the OCB moved poleward more rapidly in the Southern Hemisphere than in the Northern Hemisphere, resulting in more open flux in the Northern Hemisphere in this sector. Since the total open flux content is equal, an opposite asymmetry is expected elsewhere. This is observed post-midnight (Figure 5a) and (less clearly) post-noon (Figure 5c). In the following 10 minutes, when the closing of flux was most dominant post-midnight (Figure 5a), this region experienced a change in sign of the asymmetry. For the remaining period, the asymmetry was slowly reduced in the pre-midnight sector (Figure 5d). An opposite asymmetry, which balanced the total open magnetic flux content, was most clear pre-noon (Figure 5b). Because of the Poisson noise associated with the dayglow in the Northern Hemisphere, and the relatively low intensity of the dayside aurora in both hemispheres, an accurate determination of the boundary was difficult close to noon. As a consequence the asymmetries are generally within the error intervals in this region. However, the consistent behavior of the asymmetry seen between 21:30 and 21:40 in Figure 5c, and between 21:45 and 22:10 in Figure 5b, suggests that the asymmetries are real.

[25] In Figures 5e–5l we have divided the polar cap further into three hours wide MLT sectors. These plots show a more detailed picture of where the flux closed and how the interhemispheric asymmetries developed. Most prominently, the [0, 3] (Figure 5e) and [3, 6] (Figure 5f) MLT-sectors reveal that the asymmetry was much larger closer to midnight as the OCBs in this sector propagated poleward. After 21:50 UT, these sectors exhibited opposite asymmetries, which explains why the sum of these regions ([0, 6]), Figure 5a, was symmetric.

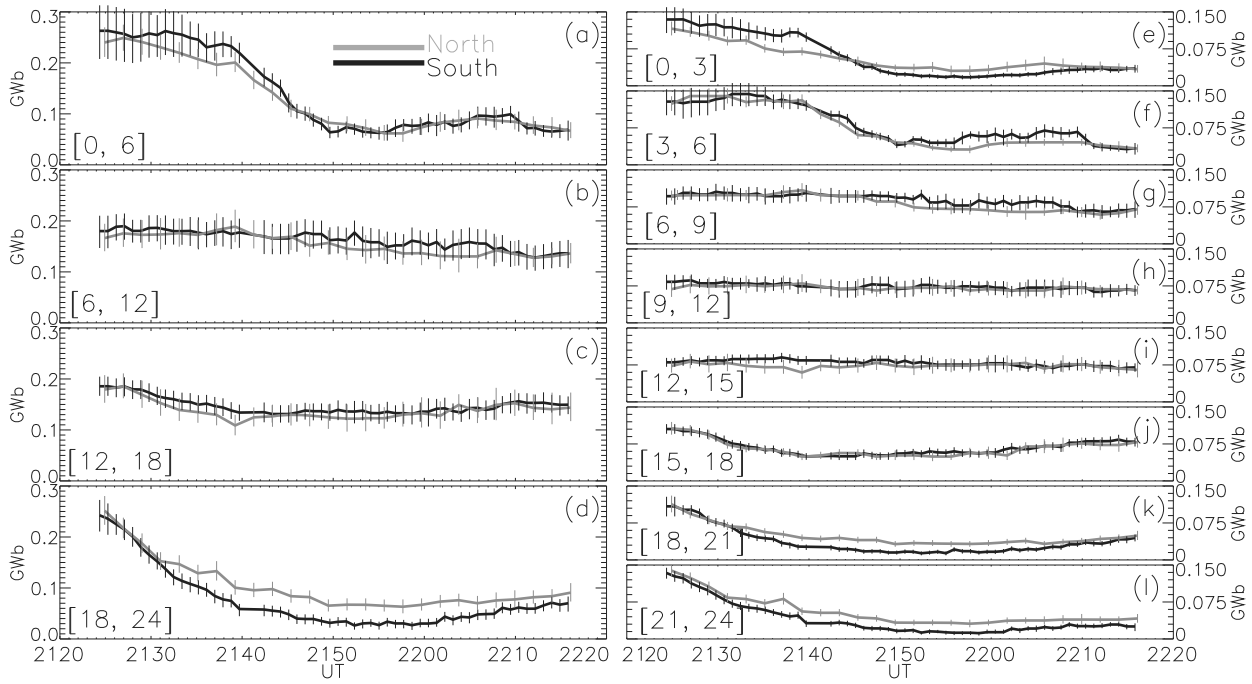

Figure 5. Open flux in different MLT sectors in both hemispheres (grey for the Northern Hemisphere, and black for the Southern Hemisphere). (a–d) Flux in 6 hours wide sectors, and (e–l) flux in 3 hours wide sectors. The error bars in Figure 5 are sums of the absolute errors, and hence relatively larger than the error bars in Figure 4.

[26] These observations show that auroral images from only one hemisphere is insufficient to determine local reconnection rate. In the [18, 24] MLT sector (Figure 5d), the reduction of magnetic flux differed by ≈50 MWb in the two hemispheres (20% of the initial flux content) during the first 20 minutes. Since the reconnection rate necessarily is equal in both hemispheres, this difference must be due to the emerging interhemispheric asymmetry.

4. Discussion

[27] The 50 minutes of global conjugate coverage of an auroral substorm presented in this paper reveals several interesting features: (1) The polar cap boundaries propagated poleward at different rates in the two hemispheres during the substorm expansion phase, leading to interhemispheric asymmetries in the polar cap boundaries. The interhemispheric asymmetry was as high as 5° in some regions. (2) The asymmetry differed from the classical view [e.g., *Cowley*, 1981a, 1981b] that polar cap asymmetries in essence is a global, uniform displacement. Instead, we observe that the asymmetry was largest in regions where the flux closure (tail reconnection) was strongest. Since the total open flux content must be the same in both hemispheres, such local asymmetries must be compensated for by an opposite asymmetry at other regions, which is what we observe.

[28] Comparing these findings to previous studies of interhemispheric asymmetries is problematic for two reasons. First, while many of the previous studies of interhemispheric asymmetries have focused only on the substorm onset location [*Østgaard et al.*, 2004, 2005b, 2007; *Liou et al.*, 2001], our findings show that interhemispheric asymmetries may change rapidly during the substorm expansion phase. This means that the formulas derived from onset observations [e.g., *Østgaard et al.*, 2004] may not be valid during the substorm expansion phase. A study of longitudinal asymmetries during the course of a substorm could resolve this issue. Second, studies which are restricted to local measurements of polar cap asymmetries [*Vorobyev et al.*, 2001] do not recognize the non-uniform nature in the asymmetry. For example, Figure 3b shows that if we were restricted to local measurements in one of the hemispheres, we would arrive at completely opposite conclusions, depending on which side of the magnetic midnight meridian we are.

[29] The only earlier study that has utilized conjugate images from IMAGE and Polar to study the entire auroral ovals, *Stubbs et al.* [2005], showed that the ovals (as represented by a best fit circle) were displaced in accordance with the expected B_y and B_x effects [e.g., *Cowley et al.*, 1991] and dipole tilt effect [e.g., *Oznovich et al.*, 1993]. Changes in asymmetry were seen to be directly related to changes in IMF orientation, consistent with newly opened field lines being subject to magnetic stresses [*Jørgensen et al.*, 1972]. The IMF control observed by *Stubbs et al.* [2005] is in contrast to the observations in the present paper, where large variations in asymmetry are seen, while the IMF remained fairly steady (Figure 2). The most obvious difference between the 12 May 2001 event, and the event studied by *Stubbs et al.* [2005], 23 October 2002 (Figure 4b), is the level of flux closure. The constant IMF and the rapid oval

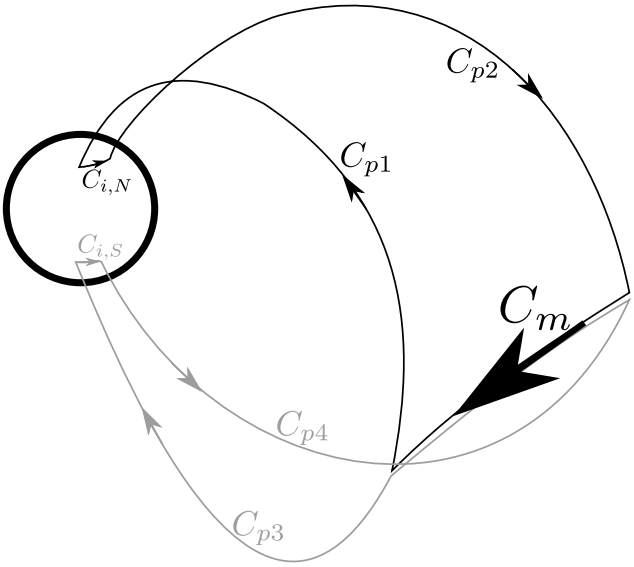

Figure 6. Paths of integration. C_m is along a segment of the X-line, and chosen so that it is similar for the two loops. The paths labeled C_p are along the magnetic field lines mapping from the X-line to the ionospheric footprints of the OCB, $C_{i,N}$ and $C_{i,S}$.

contraction suggests that the 12 May 2001 asymmetries were not directly driven by the IMF influence on newly open field lines, but rather by processes related to, or excited by tail reconnection.

[30] The progression of the OCBs from a symmetrical to an asymmetrical configuration implies that the boundaries moved at different velocities in the two hemispheres. In the following we show that the ionospheric OCB velocity depends on local reconnection rate, ionospheric convection velocity, and parallel electric fields. We also show how these quantities are related in a two-hemisphere system. Then we apply this relation on a segment of the OCBs in the 12 May 2001 event.

4.1. Open/Closed Boundary Motion

[31] The relation between polar cap motion and plasma convection can be found using a similar approach as proposed by *Vasyliunas* [1984] and applied in several reconnection studies [e.g., *Østgaard et al.*, 2005a]. We start with Faraday's law:

$$\oint (\mathbf{E} + \mathbf{u} \times \mathbf{B}) \cdot d\mathbf{l} = -\frac{\partial \Phi}{\partial t} \tag{1}$$

where Φ is the magnetic flux through a surface enclosed by the integration path on the left hand side. $\mathbf{u}$ is the velocity of the integration path with respect to the chosen coordinate system.

[32] Now we specify two integration paths, shown in Figure 6: $C_m + C_{p1} + C_{i,N} + C_{p2}$ and $C_m + C_{p3} + C_{i,S} + C_{p4}$, where $C_{i,N}$ and $C_{i,S}$ are along the ionospheric OCB in the Northern and Southern Hemispheres, respectively. C_m is along the segment of the X-line to which the two ionospheric paths map. The paths labeled C_p are chosen to be along magnetic field lines. For the integration paths C_p, $C_{i,N}$ and $C_{i,S}$ we assume the ideal MHD approximation to be true,

$$\mathbf{E} + \mathbf{v} \times \mathbf{B} \approx \mathbf{0}, \tag{2}$$

where $\mathbf{v}$ is the convection velocity of the plasma. A condition for this to hold is that the ion-neutral collision frequency is low. *Østgaard et al.* [2005a] showed that this requirement can be accommodated by placing the C_i integration paths at 300 km, which does not significantly change the latitudes inferred from auroral images (in which an emission height of 130 km were assumed). Equation (2) gives an expression for $\mathbf{E}$, which can be substituted in equation (1) everywhere, except along C_m. The integrand then, $([\mathbf{v} - \mathbf{u}] \times \mathbf{B}) \cdot d\mathbf{l}$, vanishes along C_p, since $\mathbf{B}$ is parallel to $d\mathbf{l}$. If $E_{\parallel} \neq 0$, equation (2) does not hold along C_p, but the integrals along the C_p's would cancel if the potential drop along these paths are the same. In section 4.3 we discuss the effect of non-canceling parallel electric fields. The integration path is chosen so that it is tangent to the magnetic field, and $d\Phi = \mathbf{B} \cdot \mathbf{n} dS$ is identically zero everywhere. Therefore the magnetic flux threading the surface is zero, and thus constant, and the right hand side of equation (1) vanishes. Since $\mathbf{B}$ is either parallel to the X-line or zero, $\mathbf{u} \times \mathbf{B} = \mathbf{0}$ along C_m. Hence, equation (1) reduces to

$$\int_{C_m} \mathbf{E} \cdot d\mathbf{l} = \int_{C_i} ([\mathbf{v} - \mathbf{u}] \times \mathbf{B}) \cdot d\mathbf{l}. \tag{3}$$

For the C_i integration path, $\mathbf{u}$ corresponds to the velocity of the ionospheric open/closed boundary. $\mathbf{v}$ is the ionospheric convection velocity.

[33] This equation demonstrates that, if there is no reconnection in the magnetosphere ($\mathbf{E} = \mathbf{0}$ on the left hand side), $\mathbf{u} = \mathbf{v}$, and the open/closed boundary moves exactly with the plasma convection (adiaroic convection). The equation also shows the equivalence between magnetic reconnection rate, quantified by $\mathbf{E}$, and plasma flow across the OCB, $\mathbf{v} - \mathbf{u}$ [*Vasyliunas*, 1984]. This identity has previously been used to assess tail reconnection rate from combined measurements of OCB location and plasma flow [e.g., *de la Beujardiere et al.*, 1991; *Blanchard et al.*, 1996; *Østgaard et al.*, 2005a; *Hubert et al.*, 2008].

[34] The left hand side of equation (3) is common for both paths of integration depicted in Figure 6, since conjugate segments of the OCB map to the same X-line. Setting the right hand sides in the two equations (for the two integration paths) equal to each other, we get the relation between ionospheric convection, open/closed boundary motion, and magnetic field in the two hemispheres:

$$\int_{C_{i,N}} ([\mathbf{v} - \mathbf{u}] \times \mathbf{B}) \cdot d\mathbf{l} = \int_{C_{i,S}} ([\mathbf{v} - \mathbf{u}] \times \mathbf{B}) \cdot d\mathbf{l}. \tag{4}$$

[35] From this equation it is clear that if the ionospheric convection $\mathbf{v}$ is different in the two hemispheres, $\mathbf{u}$ may also be different, and the OCBs may become asymmetrical. However, this statement is not yet exact, since interhemispheric differences in $\mathbf{v}$ may also be balanced by differences in $\mathbf{B}$, and by differences in the integration paths. We thus

need to make further considerations in order to use this equation quantitatively.

[36] First, we make one approximation regarding the integration paths: Let C_i coincide with circles of magnetic latitude (the OCB is circular along C_i seen in magnetic coordinates). We then introduce average values of the quantities in the integrand, in order to solve the integral:

$$\int_{C_i} ([\mathbf{v}-\mathbf{u}]\times\mathbf{B})\cdot d\mathbf{l} = LB_{\perp}(v-u) \tag{5}$$

$$\Rightarrow L_N B_{\perp,N}(v_N-u_N) = L_S B_{\perp,S}(v_S-u_S) \tag{6}$$

where L is the length of C_i, $B_\perp$ is the absolute magnetic field perpendicular to $d\mathbf{l}$, and v and u are velocities in the equatorward direction along magnetic meridians (the velocities are assumed to be horizontal).

[37] In general, $B_{\perp,N} \neq B_{\perp,S}$ and $L_N \neq L_S$, even when the OCB segments are on the same magnetic latitude. This is because the Apex coordinate system (and also the similarly defined AACGM system) is irregular: In the region above 60° magnetic latitude, the geographic distance corresponding to one degree of magnetic latitude ranges from ≈93 km (≈97 km) to ≈131 km (≈165 km) in the Northern (Southern) Hemisphere. The length of one degree of magnetic longitude will also vary along a circle of magnetic latitude.

[38] Conjugate images give us an estimate for the OCB locations, which we can differentiate to get the OCB velocity. The resulting unit is magnetic latitude per second, which, according to the previous paragraph is not proportional to m/s. In order to take this effect into account, we transform equation (6) to magnetic coordinates. To do this, we recognize that the OCB velocity, u', measured in magnetic coordinates relates to the OCB velocity in equation (6), u, by $u' = u/d$, where d is the dimensionless ratio between a unit length along a meridian in the magnetic system (not constant), and a (constant) unit length in a regular system (e.g., the length of one degree magnetic latitude, divided by the length of one degree latitude in a regular system, $2\pi R_E/360$). Similarly, $v' = v/d$. As discussed above, d is a function of space (on the surface of the Earth). From now on, primed quantities refer to quantities measured in magnetic coordinates. Multiplying equation (6) by $\frac{d_N d_S}{d_N d_S}$, we get,

$$d_N L_N B_{\perp,N}(v'_N-u'_N) = d_S L_S B_{\perp,S}(v'_S-u'_S) \tag{7}$$

$$\Rightarrow \Phi_N(v'_N-u'_N) = \Phi_S(v'_S-u'_S) \tag{8}$$

where we have interpreted $dLB_\perp$ as the magnetic flux Φ in a thin strip along the OCB of length L and width d (since d is dimensionless, we will have to multiply the equation by a unit length to justify this). An advantage of the Apex (and AACGM) coordinate system is that the magnetic flux in areas spanned by the same magnetic coordinates in the two hemispheres, is equal. This means that, if the OCBs start out being symmetrical, so that $\Phi_N = \Phi_S$, equation (8) implies that they can become asymmetrical ($u'_N \neq u'_S$) only if $v'_N \neq v'_S$. Emerging OCB asymmetries thus presupposes differences in the ionospheric equatorward convection in the two hemispheres.

[40] If, on the other hand, the OCBs are on different latitudes, $\Phi_N \neq \Phi_S$ (assuming they span the same longitudinal angle). In the Apex/AACGM systems, the amount of magnetic flux in equally wide concentric circles diminish towards the pole (in the longitudinal direction, the flux content is constant in these circles). A consequence of this can be seen if we put the convection equal to 0 in both hemispheres, so that the OCB moves only in response to magnetic reconnection. Then, $u'_S = \frac{\Phi_N}{\Phi_S}u'_N$. If, say the southern OCB is poleward of the northern OCB, $\Phi_N > \Phi_S$, and the southern OCB will move faster than the northern OCB. Hence, if the ionospheric footprints of an X-line in the tail (flux closure) are on different latitudes, the asymmetry will be reinforced. We emphasize that this is more of a geometric effect, introduced by the definition of the Apex coordinates, rather than an increase in the field line distortion. Nevertheless, this property must be taken into account here, and it is also useful to be aware of, since it implies that latitudinal asymmetries may be self-reinforcing. If the asymmetry is measured in magnetic flux, as in Figure 5, the self-reinforcement effect is compensated for.

4.2. OCB Motion: The 12 May 2001 Event

[41] We now solve equation (8) for the [18, 21] MLT sector in the 12 May 2001 event. The conjugate images have given us the average OCB location, from which we can find u'_S and u'_N, and calculate Φ_N and Φ_S. All quantities are considered as constant along the OCB segment. We also need to assume that the two segments of the OCB are indeed conjugate, and not displaced in longitude. To get the difference in ionospheric convection from equation (8), we will have to make assumptions for the convection velocity in one of the hemispheres. These assumptions necessarily introduce significant errors. However, the purpose of this exercise is to give a rough assessment of what the ionospheric convection must be, in order to account for the observed asymmetries. The choice of the [18, 21] sector is due to its prominent OCB, large asymmetry, and the fact that two DMSP spacecraft, carrying instruments which could measure the plasma convection, crossed the OCB in the two hemispheres. We will use these measurements, and physical considerations, to substantiate our estimate of the convection.

[42] The quantities in equation (8) for the [18, 21] MLT sector are shown schematically in Figure 7. Figure 8a shows the average location of the OCB in this sector (thin, solid), with error (dashed). We have interpolated the values to a common time resolution, and then calculated four minutes averages (thick, solid). The Northern Hemisphere boundaries are shown in grey, and the Southern Hemisphere in black. The change in latitude over time gives u' for both hemispheres, shown in Figure 8b. u' is shown in units of m′/s = (m/d)/s, where d is the scaling factor described in section 4.1. At the OCBs considered in Figure 8, d is close to 1 ($d \in [0.93, 1.12]$), so that m′/s is always close to m/s. However, we keep the primes for consistency with the preceding section, and to emphasize the principal difference between the magnetic coordinate system and a regular coordinate system. Figure 8b shows that the poleward velocity of the southern OCB is higher than the poleward velocity of the northern OCB, which leads to the observed increasing asymmetry. Figure 8c shows the evolution of Φ_N/Φ_S (in our calculation, we used the magnetic flux within a strip with a width of 1° magnetic latitude, centered at the OCB latitude). The ratio is 1 where the OCBs are collocated,

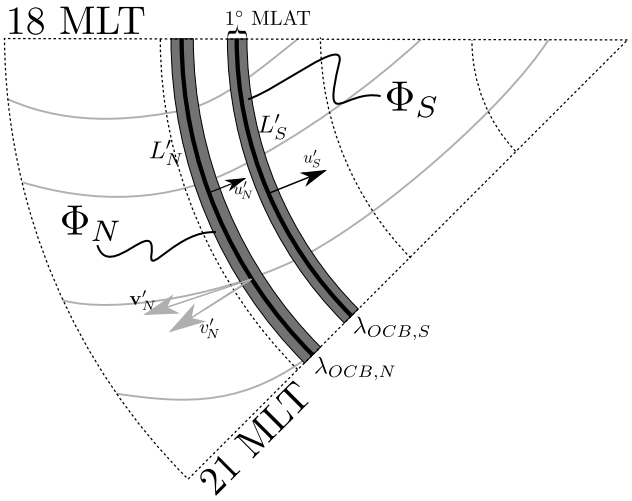

Figure 7. Schematic illustration of the quantities in equation (8). The OCB motion, u' is derived from measurements of the OCB location (λ_{OCB}). This motion depends on reconnection rate, and ionospheric convection. An illustration of the convection pattern is shown as grey contours (for clarity, only one hemisphere is shown). v' is the projection of the ionospheric convection $\mathbf{v}'$ onto a magnetic meridian. The grey strips along the OCBs mark the regions containing the flux Φ_N and Φ_S in equation (8) (for numerical reasons, we used a width of 1° magnetic latitude for the strips). Figure 7 is shown in magnetic coordinates, and for consistency with the text, we use primes to denote quantities which can take different values when transformed to a geographic coordinate system (except for λ_{OCB}).

and otherwise in agreement with decreasing magnetic flux towards the pole.

[43] The unknown quantities in equation (8) are v'_S and v'_N. First we consider the case that $v'_S = v'_N$. The resulting convection velocity is shown in Figure 8d. During the first ten minutes, the convection is unrealistically strong. This is a strong indication that the equatorward ionospheric convection must have been different in the two hemispheres during this period. After 21:35 UT, the convection is reduced, but still relatively strong (fluctuating between ≈−500 m′/s and ≈2500 m′/s). This behavior is also unrealistic, but considering the inherent uncertainties in our method, we can not rule out the possibility of symmetrical (or almost symmetrical) convection during the last 10 minutes of the period.

[44] In order to obtain some reasonable value, we will set the convection in one hemisphere to a fixed positive (equatorward) value, and then use equation (8) to calculate the other. In Figure 8e, the grey lines show v'_N = 600 m′/s (dashed) and v'_N = 900 m′/s (solid). The corresponding convection in the Southern Hemisphere is shown as black curves. In both cases, the asymmetry between hemispheres is strong ($\Delta v' \sim$ 500 m′/s) in the first ten minutes, and smaller towards the end. This shows that a difference in v of ≈500 m′/s, lasting for ≈10 minutes can account for the observed asymmetries. The decrease in equatorward convection seen in the Southern Hemisphere does however seem unreasonable, considering the increase observed in previous studies of ionospheric convection during substorm expansion [e.g., *Provan et al.*, 2004].

[45] We now use convection measurements from the two OCB crossings by DMSP F15 and F13, as well as the results from Figures 8d and 8e to suggest a more realistic scenario. The convection measurements are shown in Figure 9. Figures 9c and 9d show the value of the convection vector composed from the RPA (parallel to the satellite track) and IDM (perpendicular to the track) measurements projected onto a magnetic meridian in the Northern and Southern Hemispheres, respectively (positive in the equatorward direction). The vertical dotted bars show the time when the spacecraft crossed the OCB, as determined by particle precipitation measurements (Table 1). The values are averages within 40 seconds bins (the nominal resolution is 4 seconds), and the vertical bars denote the standard error. Only measurements labeled "good" were used for this purpose. From Figure 9, we get that $v_S \approx$ 300 m/s at 21:32:30 UT, and $v_N \approx$ 600 m/s at 21:37:20 UT (diamonds in Figure 8f). Figures 9a and 9b show the velocity vectors along the satellite tracks, plotted on top of the auroral images closest to the OCB crossing. In Figures 9a and 9b, the blue vectors denote measurements of "good" quality, while the red vectors include measurements labeled "caution."

[46] In Figure 8f, we assume the convection in the Northern Hemisphere to increase from 0 m′/s to 900 m′/s in the first few minutes of the period, and decrease to 600 m′/s at 21:35 UT. These velocities were chosen for two reasons: First, it does not imply a reduction of the convection in the Southern Hemisphere during the first ten minutes (as opposed to the v_N = const. case). Second, the velocities at the times of the OCB crossings by the DMSP spacecraft were chosen so that they are in accordance with the measurements ($d \approx 1$ so that m′/s ≈ m/s). This scenario confirms that an interhemispheric difference in equatorward convection of ≈500 m/s (using $d \approx 1$) during the first 10 minutes of the period can account for the observed asymmetry. During the remaining period, much of the still growing asymmetry is handled by the self-reinforcement effect discussed in the previous section. However, some rather fluctuating differences in convection seem to remain.

4.3. Effect of $E_\parallel$

[47] So far we have assumed $E_\parallel = 0$ along the C_p integration paths, so that equation (2) is valid. If we allow for parallel electric fields to be present, the integrals along C_p become $\int_{C_p} E_\parallel dl = \Delta V$, since the remaining terms in the integrand in equation (1) are perpendicular to $d\mathbf{l}$. If the potential drop along the two C_p's are equal, the effect of parallel electric fields are canceled by the opposite directions of the integration paths. If, on the other hand, there is a spatial gradient perpendicular to the magnetic field in the electric potentials, so that the two field aligned integrals do not cancel, we are left with a net potential drop in addition to the terms in equation (3). Assuming this net potential drop is only present in one hemisphere, we can follow the same procedure as derived above, and arrive at equation (8), only with the potential drop ΔV at one side of the equality,

$$d_N L_N B_{\perp,N}(v'_N - u'_N) = d_S L_S B_{\perp,S}(v'_S - u'_S) + \Delta V \qquad (9)$$

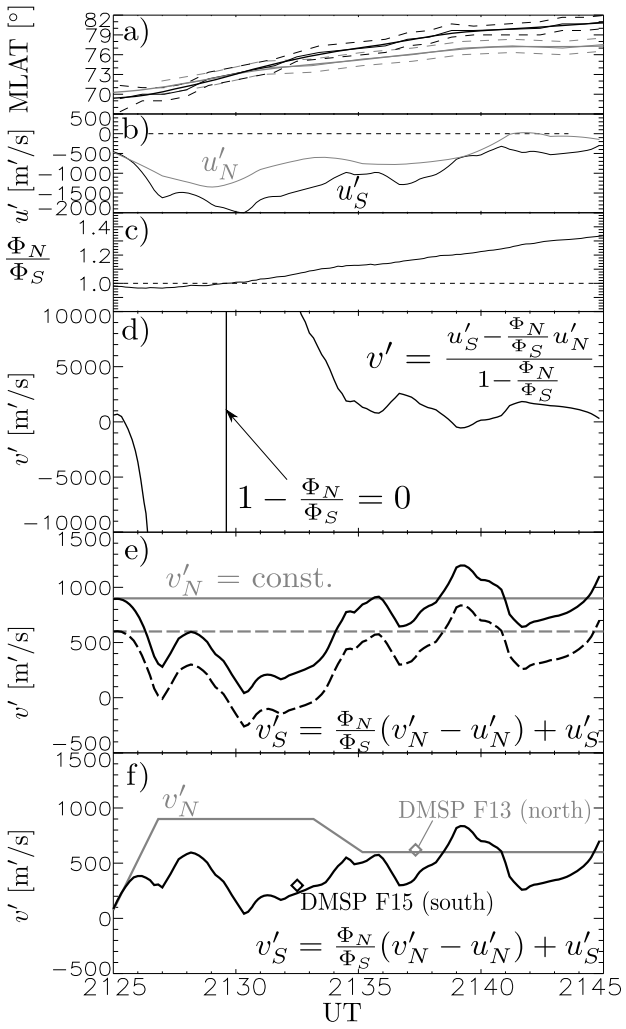

Figure 8. Equation (8) solved for the [18, 21] MLT sector on 12 May 2001. (a) Magnetic latitude of the OCB in the Northern (grey) and Southern (black) Hemispheres. The thick curves were averaged over four minutes. The dashed curves show the uncertainty. (b) OCB equatorward motion. The unit is m′/s = (m/*d*)/s. (c) Ratio of the magnetic flux in thin strips along the OCBs, Φ_N/Φ_S. (d) Equatorward convection velocity, assuming $v'_N = v'_S$. (e) Equatorward convection velocities in the south, assuming v'_N = 600 m′/s (dashed) and v'_N = 900 m′/s (solid). (f) Equatorward convection velocity in the south, assuming the convection in the Northern Hemisphere follows the grey curve. The diamonds mark convection measurements by two DMSP spacecraft.

For the dimensions to be consistent, we write the Φ's as $dLB_\perp$. This equation implies that, to accommodate the potential drop, plasma must be allowed to cross magnetic field lines ($v' - u' \neq 0$, at least in one of the hemispheres), in such a way that differences in u' may be introduced between hemispheres. If, for simplicity, we assume $d_N L_N B_{\perp,N} = d_S L_S B_{\perp,S} = LB_\perp$ and $v_N = v_S$ (we skip the primes now, since d is assumed to be 1), and let ΔV be positive (this could correspond to an electric field anti-parallel to the C_{p3} integration path (parallel to the magnetic field) in Figure 6), equation (9) gives the difference between u_N and u_S introduced by the potential drop:

$$u_N - u_S = \frac{\Delta V}{LB_\perp}. \tag{10}$$

For a magnetic field of 50,000 nT and $L \approx 2000$ km (which corresponds approximately to a longitude segment of 45° at 75° latitude), we get $LB_\perp \approx 100$ Tm. For a parallel potential drop of 5 kV, this corresponds to a difference between u_N and u_S of 50 m/s. If this was allowed to go on for 20 minutes, it would result in an interhemispheric asymmetry of ≈0.5°. A difference in u of 50 m/s is thus much less than the typical differences observed in Figure 8b. However, it should be noted that this is an average value for the entire 45° wide sector. If we reduced L, thus sharpening the gradient in ΔV, the difference in u would grow, and could become a significant factor in producing local interhemispheric field line asymmetries.

4.4. Asymmetric Convection

[48] The scenario suggested by Figure 8f is that the observed asymmetry in the 18–21 MLT sector in the 12 May 2001 event can be accounted for by a brief (~10 minutes) period of asymmetrical meridional convection ($\Delta v \approx 500$ m/s) in the two hemispheres. The next question is why the convection becomes asymmetric. In very general terms, the answer to this question is that either the high-altitude (magnetospheric) region, which is the ultimate source of the ionospheric convection, enforces the asymmetry, or that the ionosphere responds differently to symmetrical forcing, or a combination of these effects. In the 12 May 2001 event, two parameters which are believed to contribute in separate regions stand out: 1) The IMF B_x component was strong (B_y was not), which, according to *Cowley* [1981b] can lead to asymmetrical lobes, forcing the equatorial plane in the negative z direction. Owing to the same forces, the northern half of the field lines might be expected to be pushed inward more efficiently, leading to the convection asymmetry. However, this can not be the sole cause, since the OCBs became asymmetric while the IMF was unchanging. 2) The Northern Hemisphere was highly sunlit, and the Southern Hemisphere was in darkness. The situation depicted in Figure 8f resembles a northern, sunlit hemisphere which responds promptly to a burst of earthward magnetospheric convection, while the southern winter ionosphere lags behind. The difference in ionospheric response could be due to an excess of parallel electric fields in the southern winter hemisphere compared to the summer hemisphere [e.g., *Newell et al.*, 1996], which could decouple the southern ionosphere from the magnetosphere. Another possibility is that the pre-existing convection in the northern summer hemisphere, which according to statistical studies [*Ruohoniemi and Greenwald*, 2005] of ionospheric convection should be stronger than in the winter hemisphere, more easily accommodates the change in magnetospheric convection. The increased meridional convection imposed by the magnetosphere, which these explanations presuppose, has been shown to be characteristic of the substorm expansion phase [*Provan et al.*, 2004; *Bristow and Jensen*, 2007]. The Northern Hemisphere flow in Figure 8f is stronger than typical observations reported by *Provan et al.* [2004] and

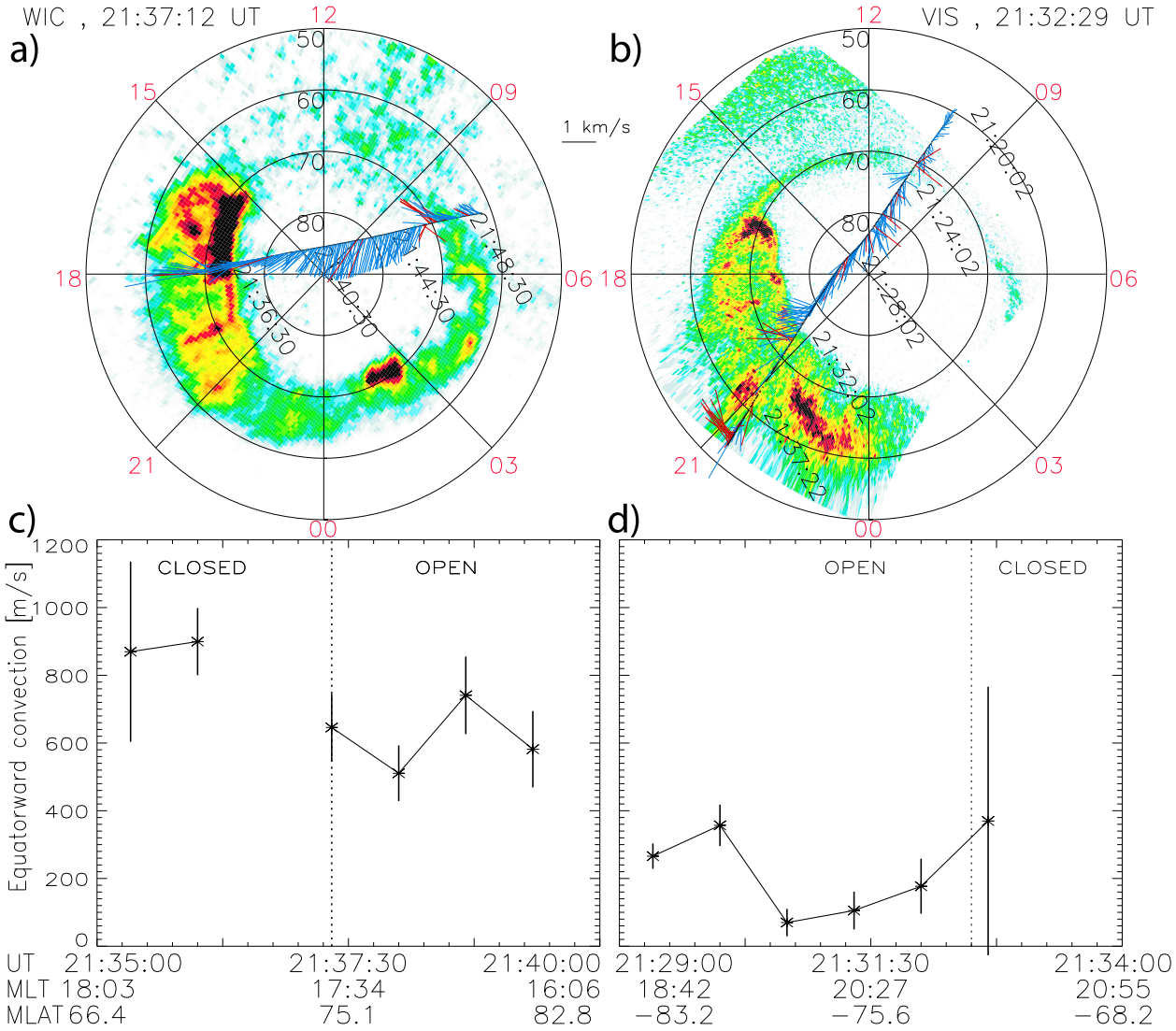

Figure 9. Ionospheric convection measurements from DMSP F13 and F15. (a) WIC image from the Northern Hemisphere, taken approximately at the time when DMSP F13 crossed the open/closed boundary at the dusk side. Horizontal convection measurements along the satellite track is also shown. Blue color indicates data of good quality, while the red vectors include data labeled "caution". (b) Same format as Figure 9a, but with satellite track of DMSP F15, which crossed the southern OCB at dusk approximately when the image was taken. (c) Average convection velocity in the Northern Hemisphere projected onto a magnetic meridian (positive in the equatorward direction). Each data point is an average of the ("good") measurements obtained at the UT on the x-axis ± 20 seconds. The error bars show $s/\sqrt{n}$, where s is the standard deviation, and n (≥2) is the number of measurements. The vertical dotted bar marks the time of the OCB crossing. (d) Same format as in Figure 9c, but for the Southern Hemisphere (DMSP F15).

Bristow and Jensen [2007], but still at a reasonable level, considering the above average strength of the substorm.

[49] The increased meridional flow is also in accordance with the theory by *Cowley and Lockwood* [1992], who proposed that the destruction of open flux implied by tail reconnection leaves the magnetosphere-ionosphere system in an excited state, which is then brought back towards an equilibrium by convection, thus making reconnection a key prerequisite for convection. This concept is illustrated in Figure 10, which is a copy of *Cowley and Lockwood* [1992, Figure 7]. In Figure 10a, there is no generation or destruction of open magnetic flux, and consequently, there is no flow across the OCB. In Figure 10b, impulsive reconnection in the tail has closed an amount of flux, dF, causing the ionospheric footprint of the X-line to move to higher latitudes. The convection, which has not yet commenced, will now start to transport the OCB to the new equilibrium, shown as a dotted-dashed circle. The excited convection is shown as arrowed loops in Figure 10c. In this panel, the reconnection has ceased, and the OCB is transported

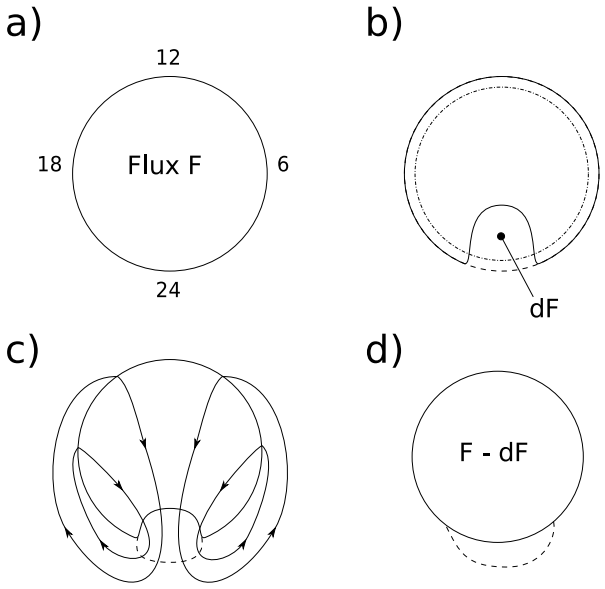

Figure 10. Cartoon showing how ionospheric convection is excited after magnetic flux closes. The solid curves indicate the open/closed field line boundary. (a) Static situation, in which no flux is opened on the dayside or closed on the nightside, and consequently there is no convection. The total open flux content is *F*. (b) Impulsive tail reconnection has closed flux (*dF*), causing a perturbation to the boundary. The dashed-dotted line indicates the new equilibrium OCB. (c) The tension associated with the perturbed boundary in the magnetosphere excites convection (arrowed solid loops), transporting the OCB towards a new equilibrium. The dashed curve indicates the boundary of the closed flux. (d) A new equilibrium has been achieved. The dashed curve indicates the location of the redistributed, newly closed flux. Figure 10 is a copy of *Cowley and Lockwood* [1992, Figure 7]. With kind permission of Springer Science + Business Media.

towards the new equilibrium with the flow (**u** = **v** in equation (3)). The new equilibrium is reached in Figure 10d, and the convection has stopped. The closed flux, *dF* has been redistributed to the area indicated by the dashed curve. In reality, this process happens more continuously, and the presence of reconnection on the dayside and on the nightside is what leads to the familiar two-cell convection pattern.

[50] According to this picture, the magnetospheric convection should be strongest where the reconnection is strongest. Therefore, due to differences in the ionospheric response, interhemispheric asymmetries in the OCB might also be expected to be strongest in these regions. This is in excellent agreement with the observations presented in Figure 5, which shows that the asymmetry is strongest in the regions where more flux closes. The conspicuously more circular OCBs in the northern than in the Southern Hemisphere (Figure 3) might also be explained in terms of the Figure 10 cartoon; if the Northern Hemisphere responded more promptly to the magnetospheric convection excited after flux closure in the tail, the ionospheric OCB would reach equilibrium (circle) earlier than in the Southern Hemisphere. In this process, the elevated convection in the Northern Hemisphere, will also transport the OCB at the regions flanking the X-line to relatively lower latitudes, which may explain the opposite asymmetries at dawn and dusk.

4.5. OCB Asymmetries and Interhemispheric Currents

[51] The 12 May 2001 event also exhibited non-conjugate aurora which has been interpreted as ionospheric signatures of interhemispheric currents [*Laundal and Østgaard*, 2009]. One of the non-conjugate spots occurred at the dawn side in the Northern Hemisphere, between 21:40 and 21:50 UT. This coincides remarkably well with significant flux closure in this region (Figure 5a), and a concurrent change in asymmetry (most evident in Figure 5e). This strongly suggests that the current and the field line perturbations signified by the OCB asymmetries are interrelated. The simplest realization of the field line perturbation, is to consider the field lines to be deformed only in the radial direction. In that case, the OCB asymmetry on the dawn-side is consistent with an additional perturbation magnetic field in the outward direction, while the nightside asymmetry is consistent with an inward perturbation field. By Ampere's law, the shear in the perturbation field between these regions implies a current anti-parallel to the field lines (electrons down in the Northern Hemisphere), approximately collocated with the non-conjugate dawn spot in the north. Since in large-scale space plasmas, the curl in magnetic field in general produces the current, and not vice versa [*Parker*, 1996; *Vasyliunas*, 2005], the cause for the non-conjugate spot (which signifies a current) may therefore be found if we can explain the OCB asymmetry (which signifies $\nabla \times \mathbf{B}$).

[52] To this end, we need to know the parameters which govern the convection during the substorm expansion phase, and hence the emerging asymmetries in the magnetic field. We have pointed out two possible candidates in the 12 May 2001 event: The IMF B_x, and the dipole tilt angle (through interhemispheric differences in ionospheric conductivity). To properly address this question, more statistics is needed. Because of the lack of global conjugate images, the most practicable way of doing this is by MHD modeling, or by a superposed epoch analysis of substorm observations in one hemisphere. This is reserved for a future work.

5. Conclusions

[53] We have used two serendipitous constellations of the IMAGE and Polar spacecraft, both equipped with UV imagers, to study the conjugate auroral ovals. We have focused on the size and shape of the dim region encircled by the bright aurora. The magnetic flux content in these regions were found to be equal. This indicates that the polar caps, as observed in UV images, are cross sections of the open magnetic flux, as is often assumed.

[54] In one of the events, 12 May 2001, a significant amount of flux closed during the expansion phase of a substorm. Concurrently, the open/closed boundary in the two hemispheres became highly asymmetrical. The emerging asymmetries were strongest in the regions where magnetic flux closed most rapidly. In neighboring regions, an opposite asymmetry was observed, preserving equal amounts of open flux.

[55] These observations strongly suggest that the interhemispheric asymmetry at the footprints of the magnetic field lines increases, and hence the magnetic field is deformed during the course of a substorm. This has not been recognized in previous studies of interhemispheric symmetry. The local nature of the asymmetries seen in this study implies that local measurements of the asymmetry may give an incomplete view of magnetospheric geometry. A complete view of interhemispheric asymmetries, and hence its causes, can only be given if temporal and spatial variations are considered.

[56] We have shown that open/closed boundary asymmetries can only arise if the ionospheric convection velocities are different in the two hemispheres, or in the presence of spatial gradients in field aligned electric potential drops along the OCB. A quantitative analysis of the [18, 21] MLT sector revealed that most of the emerging asymmetry in this sector could be accounted for if the ionospheric convection in the meridional direction differed by ≈500 m/s in the two hemispheres for the first ~10 minutes of the expansion phase. We suggest that the cause for the different convection velocities are differences in the ionospheric response to an increase in magnetospheric convection during the substorm expansion phase. The difference in ionospheric response can be due to the large seasonal differences. The strong B_x component in the IMF could also contribute to an OCB asymmetry [*Cowley*, 1981a], although this effect seems less likely, since the IMF remained almost constant, while the asymmetries changed rapidly.

[57] The clear dependence of the OCB asymmetry on reconnection rate (Figure 5) suggests that the observations presented in this paper represent an intrinsic phenomenon during substorm expansions, when flux closes rapidly in the tail [e.g., *Milan et al.*, 2009]. We have identified the IMF and ionospheric conductivity differences in the two hemispheres as two potential sources for the asymmetry. In other events, when $B_y \neq 0$, we might expect polar cap asymmetries to appear under the influence of the different flow strengths in the two convection cells [*Provan et al.*, 2004]. However, to firmly determine the parameters responsible for the asymmetries, more statistics is needed.

[58] **Acknowledgments.** The authors thank S.B. Mende and the IMAGE FUV team for the use of IMAGE FUV data. We thank J. Sigwarth for the use of Polar VIS Earth data. We thank C. Smith for the ACE magnetic field data and D. McComas for the ACE solar wind data. We gratefully acknowledge the Center for Space Sciences at the University of Texas at Dallas and the U.S. Air Force for providing the DMSP thermal plasma data. The DMSP boundaries were determined from plots at The Johns Hopkins University Applied Physics Laboratory web page. Marit Sandanger and Finn Søraas assisted in the processing of NOAA data. Geotail magnetic field data were provided by T. Nagai through DARTS at Institute of Space and Astronautical Science, JAXA in Japan. This study was financed by the IPY-ICESTAR project, The Research Council of Norway, contract 176045/S30.

[59] Masaki Fujimoto thanks the reviewers for their assistance in evaluating this paper.

References

Blanchard, G. T., L. R. Lyons, O. de la Beaujardiére, R. A. Doe, and M. Mendillo (1996), Measurement of the magnetotail reconnection rate, *J. Geophys. Res.*, *101*, 15,265–15,276.

Boakes, P. D., S. E. Milan, G. A. Abel, M. P. Freeman, G. Chisham, B. Hubert, and T. Sotirelis (2008), On the use of IMAGE FUV for estimating the latitude of the open/closed magnetic field line boundary in the ionosphere, *Ann. Geophys.*, *26*, 2759–2769.

Boakes, P. D., S. E. Milan, G. A. Abel, M. P. Freeman, G. Chisham, and B. Hubert (2009), A statistical study of the open magnetic flux content of the magnetosphere at the time of substorm onset, *Geophys. Res. Lett.*, *36*, L04105, doi:10.1029/2008GL037059.

Bristow, W. A., and P. Jensen (2007), A superposed epoch study of SuperDARN convection observations during substorms, *J. Geophys. Res.*, *112*, A06232, doi:10.1029/2006JA012049.

Brittnacher, M., M. Fillingim, G. Parks, G. Germany, and J. Spann (1999), Polar cap area and boundary motion during substorms, *J. Geophys. Res.*, *104*, 12,251–12,262, doi:10.1029/1998JA900097.

Burch, J. L. (2000), IMAGE Mission Overview, *Space Sci. Rev.*, *91*, 1–14.

Carbary, J. F., T. Sotirelis, P. T. Newell, and C.-I. Meng (2003), Auroral boundary correlations between UVI and DMSP, *J. Geophys. Res.*, *108* (A1), 1018, doi:10.1029/2002JA009378.

Cowley, S. W. H. (1981a), Asymmetry effects associated with the x-component of the IMF in a magnetically open magnetosphere, *Planet. Space Sci.*, *29*, 809–818.

Cowley, S. W. H. (1981b), Magnetospheric asymmetries associated with the y-component of the IMF, *Planet. Space Sci.*, *29*, 79–96.

Cowley, S. W. H., and M. Lockwood (1992), Excitation and decay of solar wind-driven flows in the magnetosphere-ionosphere system, *Ann. Geophys.*, 103–115.

Cowley, S. W. H., J. P. Morelli, and M. Lockwood (1991), Dependence of convective flows and particle precipitation in the high-latitude dayside ionosphere on the x and y components of the interplanetary magnetic field, *J. Geophys. Res.*, *96*, 5557–5564, doi:10.1029/90JA02063.

de la Beujardiere, O., L. R. Lyons, and E. Friis-Christensen (1991), Sondrestrom radar measurement of the reconnection electric field, *J. Geophys. Res.*, *96*, 13,907–13,912, doi:10.1029/91JA01174.

Dungey, J. W. (1961), Interplanetary magnetic field and the auroral zones, *Phys. Rev. Lett.*, *6*, 47–48.

Frank, L. A., and J. B. Sigwarth (2003), Simultaneous images of the northern and southern auroras from the Polar spacecraft: An auroral substorm, *J. Geophys. Res.*, *108*(A4), 8015, doi:10.1029/2002JA009356.

Frank, L. A., J. B. Sigwarth, J. D. Craven, J. P. Cravens, J. S. Dolan, M. R. Dvorsky, P. K. Hardebeck, J. D. Harvey, and D. W. Muller (1995), The visible imaging system VIS for the Polar spacecraft, *Space Sci. Rev.*, *71*, 297–328.

Frey, H. U., S. B. Mende, H. B. Vo, M. Brittnacher, and G. K. Parks (1999), Conjugate observation of optical aurora with polar satellite and ground-based cameras, *Adv. Space Res.*, *23*, 1647–1652.

Haaland, S. E., G. Paschmann, M. Forster, J. M. Quinn, R. B. Torbert, C. E. McIlwain, H. Vaith, P. A. Puhl-Quinn, and C. A. Kletzing (2007), High-latitude plasma convection from cluster EDI measurements: method and IMF-dependence, *Ann. Geophys.*, *25*, 239–253.

Hubert, B., S. E. Milan, A. Grocott, C. Blockx, S. W. H. Cowley, and J.-C. Gérard (2006), Dayside and nightside reconnection rates inferred from IMAGE-FUV and Super Dual Auroral Radar Network data, *J. Geophys. Res.*, *111*, A03217, doi:10.1029/2005JA011140.

Hubert, B., S. E. Milan, A. Grocott, S. W. H. Cowley, and J. C. Gerard (2008), Open magnetic flux and magnetic flux closure during sawtooth events, *Geophys. Res. Lett.*, *35*, L23301, doi:10.1029/2008GL036374.

Jørgensen, T. S., E. Friis-Christensen, and J. Wilhjelm (1972), Interplanetary magnetic-field directions and high-latitude ionospheric currents, *J. Geophys. Res.*, *77*, 1976–1977.

Lam, M. M., M. Pinnoch, and E. F. Donovan (2006), Observations of nightside magnetic reconnection during substorm growth and expansion phases, *J. Geophys. Res.*, *111*, A05209, doi:10.1029/2005JA011356.

Laundal, K. M., and N. Østgaard (2009), Asymmetric auroral intensities in the Earth's Northern and Southern Hemispheres, *Nature*, *460*, 491–493.

Liou, K., P. T. Newell, D. G. Sibeck, C.-I. Meng, M. Brittnacher, and G. Parks (2001), Observation of imf and seasonal effects in the location of auroral substorm onset, *J. Geophys. Res.*, *106*, 5799–5810, doi:10.1029/2000JA003001.

Lockwood, M., S. W. H. Cowley, and M. P. Freeman (1990), The excitation of plasma convection in the high-latitude ionosphere, *J. Geophys. Res.*, *95*, 7961–7972, doi:10.1029/JA095iA06p07961.

Mende, S. B., H. U. Frey, T. J. Immel, J.-C. Gérard, B. Hubert, and S. A. Fuselier (2003a), Global imaging of proton and electron aurorae in the far ultraviolet, *Space Sci. Rev.*, *109*, 211–254.

Mende, S. B., H. U. Frey, B. J. Morsony, and T. J. Immel (2003b), Statistical behavior of proton and electron auroras during substorms, *J. Geophys. Res.*, *108*(A9), 1339, doi:10.1029/2002JA009751.

Mende, S. B., et al. (2000), Far ultraviolet imaging from the IMAGE spacecraft. 2. wideband FUV imaging, *Space Sci. Rev.*, *91*, 271–285.

Milan, S. E., M. Lester, S. W. H. Cowley, K. Oksavik, M. Brittnacher, R. A. Greenwald, G. Sofko, and J.-P. Villain (2003), Variations in the polar cap area during two substorm cycles, *Ann. Geophys.*, *21*, 1121–1140.

Milan, S. E., G. Provan, and B. Hubert (2007), Magnetic flux transport in the dungey cycle: A survey of dayside and nightside reconnection rates, *J. Geophys. Res.*, *112*, A01209, doi:10.1029/2006JA011642.

Milan, S. E., P. D. Boakes, and B. Hubert (2008), Response of the expanding/contracting polar cap to weak and strong solar wind driving: Implications for substorm onset, *J. Geophys. Res.*, *113*, A09215, doi:10.1029/2008JA013340.

Milan, S. E., A. Grocott, C. Forsyth, S. M. Imber, P. D. Boakes, and B. Hubert (2009), A superposed epoch analysis of auroral evolution during substorm growth, onset and recovery: open magnetic flux control of substorm intensity, *Ann. Geophys.*, *27*, 659–668.

Newell, P. T., C.-I. Meng, and K. M. Lyons (1996), Suppression of discrete aurorae by sunlight, *Nature*, *381*, 766–767.

Østgaard, N., S. B. Mende, H. U. Frey, T. J. Immel, L. A. Frank, J. B. Sigwarth, and T. J. Stubbs (2004), Interplanetary magnetic field control of the location substorm onset and auroral features in the conjugate hemisphere, *J. Geophys. Res.*, *109*, A07204, doi:10.1029/2003JA010370.

Østgaard, N., J. Moen, S. B. Mende, H. U. Frey, T. J. Immel, P. Gallop, K. Oksavik, and M. Fujimoto (2005a), Estimates of magnetotail reconnecton rate based on IMAGE FUV and EISCAT measurements, *Ann. Geophys.*, *23*, 123–134.

Østgaard, N., N. A. Tsyganenko, S. B. Mende, H. U. Frey, T. J. Immel, M. Fillingim, L. A. Frank, and J. B. Sigwarth (2005b), Observations and model predictions of substorm auroral asymmetries in the conjugate hemispheres, *Geophys. Res. Lett.*, *32*, L05111, doi:10.1029/2004GL022166.

Østgaard, N., S. B. M. abd, H. U. Frey, J. B. Sigwarth, A. Aasnes, and J. Weygand (2007), Auroral conjugacy studies based on global imaging, *J. Atmos. Terr. Phys.*, *69*, 249–255.

Oznovich, I., R. W. Eastes, R. E. Hoffman, M. Tur, and I. Glaser (1993), The aurora at quiet magnetospheric conditions: Repeatability and dipole tilt angle dependence, *J. Geophys. Res.*, *98*, 3789–3797, doi:10.1029/92JA01950.

Parker, E. N. (1996), The alternative paradigm for magnetospheric physics, *J. Geophys. Res.*, *105*, 10,587–10,625, doi:10.1029/95JA02866.

Provan, G., M. Lester, S. B. Mende, and S. E. Milan (2004), Statistical study of high-latitude plasma flow during magnetospheric substorms, *Ann. Geophys.*, *22*, 3607–3624.

Richmond, A. D. (1995), Ionospheric electrodynamics using magnetic apex coordinates, *J. Geomagn. Geoelectr.*, *47*, 191–212.

Ruohoniemi, J. M., and R. A. Greenwald (2005), Dependencies of high-latitude plasma convection: Consideration of interplanetary magnetic field, seasonal, and universal time factors in statistical patterns, *J. Geophys. Res.*, *110*, A09204, doi:10.1029/2004JA010815.

Sato, N., A. Kadokura, Y. Ebihara, H. Deguchi, and T. Saemundsson (2005), Tracing geomagnetic conjugate points using exceptionally similar synchronous auroras, *Geophys. Res. Lett.*, *32*, L17109, doi:10.1029/2005GL023710.

Siscoe, G. L., and T. S. Huang (1985), Polar cap inflation and deflation, *J. Geophys. Res.*, *90*, 543–547.

Stenbaek-Nielsen, H. C., and A. Otto (1997), Conjugate auroras and the interplanetary magnetic field, *J. Geophys. Res.*, *102*, 2223–2232, doi:10.1029/96JA03563.

Stubbs, T. J., R. R. Vondrak, N. Østgaard, J. B. Sigwarth, and L. A. Frank (2005), Simultaneous observations of the auroral ovals in both hemispheres under varying conditions, *Geophys. Res. Lett.*, *32*, L03103, doi:10.1029/2004GL021199.

Vasyliunas, V. M. (1984), Steady state aspects of magnetic field line merging, in *Magnetic Reconnection in Space and Laboratory Plasmas*, *Geophys. Mongr. Ser.*, vol. 30, edited by E. Hones Jr., pp. 25–31, AGU, Washington D. C.

Vasyliunas, V. M. (2005), Relation between magnetic fields and electric currents in plasmas, *Ann. Geophys.*, *23*, 2589–2597.

Vorobyev, V. G., O. I. Yagodkina, D. Sibeck, K. Liou, and C.-I. Meng (2001), Aurora conjugacy during substorms: Coordinated antarctic ground and polar ultraviolet observations, *J. Geophys. Res.*, *106*, 24,579–24,591, doi:10.1029/2001JA900025.

Wing, S., P. T. Newell, D. G. Sibeck, and K. B. Baker (1995), A large statistical study of the entry of interplanetary magnetic field y-component into the magnetosphere, *Geophys. Res. Lett.*, *22*, 2083–2086, doi:10.1029/95GL02261.

H. Frey, Space Sciences Laboratory, University of California, 7 Gauss Way, Berkeley, CA 94720-7450, USA. (hfrey@ssl.berkeley.edu)

K. M. Laundal and N. Østgaard, Department of Physics and Technology, University of Bergen, Allegt. 55, N-5007 Bergen, Norway. (karl.laundal@ift.uib.no; nikolai.ostgaard@ift.uib.no)

K. Snekvik, Finnish Meteorological Institute, PO Box 503, FIN-00101 Helsinki, Finland. (kristian.snekvik@fmi.fi)

Paper IV

K. M. Laundal, N. Østgaard, H. U. Frey and J. M. Weygand (2010)

Seasonal and IMF dependent polar cap contraction during substorm expansion phase

Journal of Geophysical Research in press, doi:10.1029/2010JA015910

Seasonal and IMF dependent polar cap contraction during substorm expansion phase

K. M. Laundal,[1] N. Østgaard,[1] H. U. Frey,[2] and J. M. Weygand[3]

Abstract. Recent observations from simultaneous imaging in two hemispheres have shown that the polar caps can attain considerably different shapes as the auroral ovals contract during substorm expansion phase. In this paper we use images from 2770 substorms to study the evolution of the polar cap boundary location statistically. We show that during the first 26 minutes after substorm expansion phase onset, the polar cap boundary location depends on seasons, IMF B_y, and IMF B_x. For different signs of B_y, with $|B_y| > 3$ nT, the asymmetry in polar cap boundary observed at onset increases during expansion phase, consistent with an increase in tail reconnection of field lines with asymmetrical footprints. When $B_x > 2$ nT and $|B_y| < 2$ nT, the polar cap boundary dawnward of the onset propagates slightly further poleward compared to negative B_x conditions. In the sunlit hemisphere, the polar cap boundary evolves from a pronounced equatorward displacement at onset, to an almost reversed displacement during the expansion phase, compared to substorms observed in darkness. Substorms in the dark hemisphere also have a much more pronounced bulge than substorms in the sunlit hemisphere. If the interpretation of the poleward auroral boundary as being coincident with the open/closed field line boundary (OCB) is correct, the seasonal differences in OCB locations imply seasonal differences in the ionospheric convection during substorm expansion phase.

1. Introduction

It is now well established that the Earth's magnetic field can be perturbed, such as to imply temporal inter-hemispheric asymmetries in magnetic field line footprints. This can be clearly seen by observing the aurora simultaneously in the two hemispheres, since similar auroral forms are ionospheric footprints of the same magnetic field lines (e.g., *Østgaard et al.* [2004]). The distortion of the field lines, which is implied by such observations, is ultimately an effect of the interaction between the Earth's magnetic field, and the interplanetary magnetic field (IMF), which is transported away from the sun by the solar wind plasma.

Most of our quantitative knowledge about how the IMF orientation affects inter-hemispheric asymmetries stems from studies of the onset of auroral substorms (*Liou et al.* [2001]; *Østgaard et al.* [2005, 2007]; *Wang et al.* [2007]). The onset of a substorm can be seen as a moment of transition between a relatively calm and steady state, to a disturbed state. After the onset, the magnetotail becomes more dipolar (e.g. *Baumjohann et al.* [1999]), and excessive reconnection of open magnetic field lines takes place (as observed by e.g. *Blanchard et al.* [1997]). The significant and global changes that occur in the magnetospheric geometry during this period are also likely to change the prevailing inter-hemispheric differences.

In two recent studies of conjugate global auroral images during the course of a substorm, large and varying asymmetries in intensity (*Laundal and Østgaard* [2009]) and magnetic field line footprint (*Laundal et al.* [2010]) were reported. The latter study showed that, as the polar caps contracted during a substorm expansion phase, their shapes became different. The asymmetries were largest in the regions where the polar caps contracted the most. It was shown that the asymmetries could be accounted for, if excitation of ionospheric convection lagged by ≈ 10 minutes in one hemisphere, compared to the other. The inter-hemispheric differences in this event could either have been due to the large x component of the IMF (the y component was negligible), or the large seasonal differences.

In the present paper, we investigate statistically the effect of these parameters on auroral poleward propagation. The basis for the study is $\sim 30,000$ images from 3943 substorms observed by the Wideband Imaging Camera (WIC) during the lifetime of the Imager for Magnetopause-to-Aurora Global Exploration (IMAGE) satellite, and identified by *Frey et al.* [2004]; *Frey and Mende* [2006]. In their study, substorms were defined as clear local auroral brightenings, expanding in latitude and longitude for at least 20 minutes. In addition, they eliminated events which occurred less than 30 minutes after the previous onset. This definition is a quantification of the qualitative auroral substorm description presented by *Akasofu* [1964]. Several subsequent substorm studies have substantiated and elaborated on the *Akasofu* [1964] substorm picture. Of particular importance to the present paper are the studies of the poleward auroral boundary motion during substorms, inferred from global images, both in event studies [*Craven and Frank*, 1987; *Brittnacher et al.*, 1999; *Milan et al.*, 2003], and statistically [*Mende et al.*, 2003; *Gjerloev et al.*, 2008]. One of the findings of *Mende et al.* [2003] and *Gjerloev et al.* [2008] was that the poleward boundary on average propagates furthest poleward close to the magnetic local time of the substorm onset. It is the aim of the present paper to investigate the dependence of the poleward propagation on seasons and IMF orientation. We focus on the B_x and B_y components of the IMF, since they are believed to affect the two hemispheres differently, producing inter-hemispheric asymmetries.

We use an automated routine to identify the poleward boundary of the aurora, assumed to be co-located with the open/closed magnetic field line boundary (OCB), at substorm onset and at four time steps up to 26 minutes into

[1]University of Bergen, Dept. of Physics and Technology, Norway
[2]Space Sciences Laboratory, University of California, USA
[3]Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California, USA

0148-0227/10/$9.00

the substorm expansion phase. Looking at average values of the location of this boundary for substorms selected according to the concurrent IMF orientation and seasonal conditions, we find that the morphology of the polar cap during substorm expansion phase is significantly affected by these parameters. Two mechanisms are proposed to dominate in the production of these asymmetries: Tail reconnection of magnetic field lines with footprints at asymmetrical points in the ionosphere, and an asymmetrical response in the ionosphere to excitation of magnetospheric convection.

In the next section, we describe the method that was used to compile the set of OCBs which constitutes the basis for this study, as well as the parameters according to which the substorms are binned. In Section 3, average boundaries are presented, and differences due to season and IMF are reported. In Section 4, we discuss various mechanisms which might explain these observations, and Section 5 concludes the paper.

2. Method

The basis for this study are images from substorm onset and expansion phase during the 3943 substorms that *Frey et al.* [2004]; *Frey and Mende* [2006] identified in images from the WIC camera on the IMAGE satellite. The IMAGE satellite was launched on 25 March 2000, and it provided data until December 2005. During its first years, the apogee at $\approx 7R_E$ was close to the North pole, and in the later years of the mission, apsidal precession had moved it to the southern hemisphere. Most of the substorms in this study are from the first years, because of longer continuous monitoring of the auroral zone, and because of an increasingly inaccurate pointing in the last years of the mission. We made a rough evaluation of the pointing accuracy in each observed substorm, based on the image from substorm onset: First, we removed a few events after manual inspection showing the aurora to be positioned away from the auroral zone. Second, we made sure that the location of the substorm onset, as reported by *Frey et al.* [2004]; *Frey and Mende* [2006], was associated with a local maximum in auroral intensity. Having experimented with different subsets of the resulting data set, we conclude that the inclusion of events late in the IMAGE mission did not change the results significantly, other than contributing to the statistical basis. The images used in this study are from the WIC [*Mende et al.*, 2000]. WIC provided images in the Lyman-Birge-Hopfield wavelength band (140 - 190 nm). Being mounted on the spinning satellite, it produced images every 123 seconds (the satellite spin period), and had 10 seconds integration time.

The coordinate system used in this study is Apex coordinates (*Richmond* [1995]). This coordinate system is based on the International Geomagnetic Reference Field (IGRF), and it produces symmetrical coordinates in the two hemispheres at points which are on the same field lines in the IGRF (the assumed emission height is 130 km). Interhemispheric asymmetries are therefore readily understood as deviations from the IGRF. The difference between Apex, and the other commonly used system for auroral zone studies, AACGM, is negligible, since these systems are defined almost similarly.

Measurements of the IMF were obtained from the Advanced Composition Explorer (ACE) spacecraft, which is located in orbit around the L1 point, $\sim 250R_E$ sunward of the Earth. The measurements were time shifted to the dayside magnetopause using the *Weimer et al.* [2003]; *Weimer* [2004] method, which uses a minimal variance analysis to include the geometry of the IMF in the time-shift estimate. For each substorm, one value for the components of the IMF was assigned (if data was available): The average in the period 30 minutes prior to onset, until 20 minutes after onset. Changing this definition to other time windows close to onset did not produce noticeable differences in our results. Neither did imposing constraints on the variability of the IMF.

The substorms were also binned according to seasons, parameterized by the tilt angle of the dipole axis. This is the axis of a best-fit dipole to the IGRF, and it is tilted at an angle of $\approx 12^\circ$ with respect to the Earth's rotational axis. The values of the tilt angle therefore range between -35° (minimum at northern winter solstice) and 35° (maximum at northern summer solstice). We use the terms positive tilt/summer/sunlit interchangeably. However, since the best-fit dipole axis (which also gives the magnetic poles in Apex coordinates) does not pass through the center of the Earth, the tilt angle does not correspond to a unique location of the sunlight terminator seen in Apex coordinates. Having experimented with bins according to UT and various thresholds for the tilt, we can safely say that this discrepancy does not significantly affect our results.

2.1. OCB identification

Several studies have used auroral images to determine the open/closed field line boundary (e.g. *Milan et al.* [2007]; *Hubert et al.* [2008]; *Boakes et al.* [2009]), and the validity of the method has been substantiated by *Carbary et al.* [2003]; *Boakes et al.* [2008], by comparing with boundaries inferred from low-altitude in-situ particle precipitation measurements. These authors also found a systematic bias, with the UV determined boundary being equatorward of the particle boundary at dawn, and slightly poleward at dusk. Despite deviations, we shall use the terms OCB and poleward auroral boundary interchangeably. We do this, believing that while the actual OCB may be slightly displaced from our observed boundaries, the two boundaries are highly correlated, and that any bias is independent of substorm selection criteria (dipole tilt angle and IMF orientation).

To determine the OCB we use a scheme first suggested by *Carbary et al.* [2003]: The intensity as a function of latitude is found in 1 h wide magnetic local time (MLT) sectors, with a latitudinal resolution of 1° magnetic latitude. Using a least-squares method, the following function is fitted to the resulting intensity profile:

$$f(\lambda) = a_1 e^{-\frac{1}{2}\left(\frac{\lambda - a_2}{a_3}\right)^2} + a_4 + a_5\lambda + a_6\lambda^2 \qquad (1)$$

where λ is magnetic latitude and a_i are constants determined by the least-squares method. The Gaussian defined by $a_{1,2,3}$ usually coincides with the aurora, and the quadratic function defined by $a_{4,5,6}$ handles the background, including the the sunlight induced dayglow. In the case of a successful fit, the boundary is assumed to be located one full width at half maximum (FWHM) poleward of the peak of the Gaussian:

$$\lambda_{OCB} = a_2 + FWHM = a_2 + 2a_3\sqrt{2ln(2)}. \qquad (2)$$

In applying this method to a large number of intensity profiles (more than 500,000 intensity profiles were fitted for this study), a precise test of the goodness of fit is essential. We use a set of criteria which are slightly different from the criteria used by *Carbary et al.* [2003] and *Boakes et al.* [2008]. Since we rely on these validation studies, we have made a detailed comparison between the two methods of testing the fit. We find that our more liberal criteria yield between 2 and 4 times more valid boundaries, without introducing significant errors. An extensive description of our method, and a detailed comparison with earlier methods is given in the appendix.

We note that the use of Equation 1 introduces a bias towards single oval events. According to *Gjerloev et al.*

[2008], the latitudinal intensity profile during substorm expansion phase is in general approximated more effectively by a double Gaussian, than a single Gaussian. Double auroral ovals are believed to be most pronounced during the recovery phase of substorms [*Elphinstone et al.*, 1995], and possibly more frequent in substorms that occur during sawtooth events [*Henderson et al.*, 2006]. The results in the present study may therefore be more representative of the expansion phase of isolated substorms. According to *Huang et al.* [2009], isolated substorms are associated with less open flux (OCB at higher latitudes) at substorm onset, and less flux closure during the expansion phase, compared to substorms during sawtooth events.

We are not concerned with the total flux content of the magnetosphere, and therefore do not need to make an indirect estimate of the boundaries which are not successfully fitted, in order to make the OCB a closed loop. Only boundaries satisfying the chosen requirements are used. We only use boundaries from 12 hours wide sectors on the nightside, and 12 hours wide sectors centered at substorm onset. In order to promote images with good coverage and clear ovals, we discard an image if less than 5 out of these 12 boundaries are successfully identified.

3. Observations

With the above selection criteria, and with some events late in the IMAGE mission removed because of errors in satellite pointing, we are left with 2770 substorms. Substorms from both hemispheres were used, although the vast majority were observed in the northern hemisphere. For the substorms observed in the southern hemisphere, we changed the sign of IMF B_y and the tilt angle, assuming that these parameters affect the two hemispheres symmetrically.

3.1. Format of Figures 1, 2 and 3

We present three figures with common format, one showing the average behavior of all substorms (Figure 1), one where the substorms are binned according to tilt angle (Figure 2), and one where they are binned according to the value of B_y (Figure 3).

For each substorm, we pick nine images: The onset image, and images number 3, 4, 6, 7, 9, 10, 12, and 13 after onset. The average OCBs in the onset images are labeled $t = 0$, while boundaries from consecutive images (image number 3 and 4, 6 and 7 etc.) from the expansion phase are labeled by the average time, in minutes, after the onset images: 7.2, 13.3, 19.5, and 25.6 (the satellite spin period is 123 s). The grouping of consecutive images in common bins is done to increase the statistical significance. Different times are marked by different colors in the plots. The upper limit of 25.6 minutes was chosen for the present study for the sake of clarity of the figures, and because of reduced statistics in the later phase of the substorm. In Figures 2 and 3, where the substorms are chosen according to tilt and B_y, different signs of these parameters are distinguished by dashed and solid lines.

In Figures 1, 2 and 3, the column labeled 'a' shows the average OCBs in 12 one hour wide magnetic local time sectors, centered at midnight. In columns b, c, and d, the boundaries contributing to the average were binned by their location relative to the MLT of the substorm onset. Negative numbers indicate boundaries that are duskward of the onset, while positive numbers are dawnward of the onset. In columns c and d, the substorms were grouped depending on the location of the substorm onset relative to the median of onset locations for the whole *Frey et al.* [2004]; *Frey and Mende* [2006] data set, which is 22.9 MLT. Columns c show substorms with onset duskward of the median, while columns d show onsets dawnward of the median. The substorm onsets furthest towards noon (at either side of midnight) occurred at 16.9 and 4.2 MLT, constituting the most westward and eastward substorms included in columns c and d respectively.

For the boundaries at onset, and at $\bar{t} = 25.6$, which are the times with the least statistical basis, the thickness of the curves constitute the average plus/minus half the standard error, $s/\sqrt{n}$, where s is the standard deviation in each bin, and n is the number of points upon which the average is based. For clarity, we only show the error for the selection criteria giving the least number of valid boundaries (e.g., the sunlit hemisphere, where the Gaussian fit is less likely to succeed). The thickness of these curves can therefore be regarded as the maximum error in each figure (with the single exception of Figure 3c). The lower panels show the number of points used to determine each of the average boundaries.

3.2. Average OCBs for all substorms

In Figure 1 we show the average evolution of the OCB for all substorms with acceptable data quality.

On average, substorm onset can be seen as a time when the magnetosphere/ionosphere changes from an undisturbed state to a disturbed state. The black curve in Figure 1a therefore tells us that on average, during quiet conditions, the OCB on the nightside (and likely for the whole oval) is furthest equatorward at midnight. This is perhaps not surprising, since we know that the oval generally is displaced towards midnight [*Meng et al.*, 1977], presumably under the influence of the anti-sunward momentum of the solar wind, and intuitively, midnight is where this effect converges to a maximum. Another interesting point is that the latitude increases faster with distance from midnight towards dawn, than towards dusk. This effect is probably even underestimated in the data, since UV image determination of the OCB has been shown to have a bias towards equator at the dawnside, compared to the boundaries inferred from low-altitude in-situ precipitation measurements (*Carbary et al.* [2003]; *Boakes et al.* [2008]). We show in Section 3.6 that the region dawnward of the onset is associated with a larger increase in the width of the oval, than on the dusk side, which could account for some of the observed difference between these regions.

For the $t > 0$ curves, we note the following: 1) The average boundary propagates poleward monotonically for at least 25.6 minutes after onset. 2) Local maxima are observed in all $t > 0$ curves, but the maximum is much clearer in the plots where we look at the relative MLT (b, c, d).

These two observations are in very good agreement with what we expect from *Akasofu* [1964]; *Mende et al.* [2003]; *Gjerloev et al.* [2008]. The local maximum is the poleward boundary of the auroral bulge, and we shall refer to this maximum as the bulge from now on. If the interpretation of the poleward boundary as coincident with the OCB is correct, this pattern can only mean that the expansion phase of substorms, on average, are associated with tail reconnection, closing magnetic flux to allow precipitation on closed field lines to expand poleward. Further, the tendency for the poleward propagation to be centered at onset shows that the X-line on average is located close to the same magnetic local time as the substorm onset.

Figures 1c and 1d show quite different shapes of the OCB, but they both reveal a tendency that the poleward expansion is stronger duskward of the onset. This could be an effect of stronger anti-sunward convection closer to midnight, transporting the OCB (in the sense described by *Cowley and Lockwood* [1992]) equatorward faster in this region, rectifying the effect of tail reconnection.

No smoothing was applied to the curves in Figure 1 (nor in the following figures), but still no clear signs of statistical

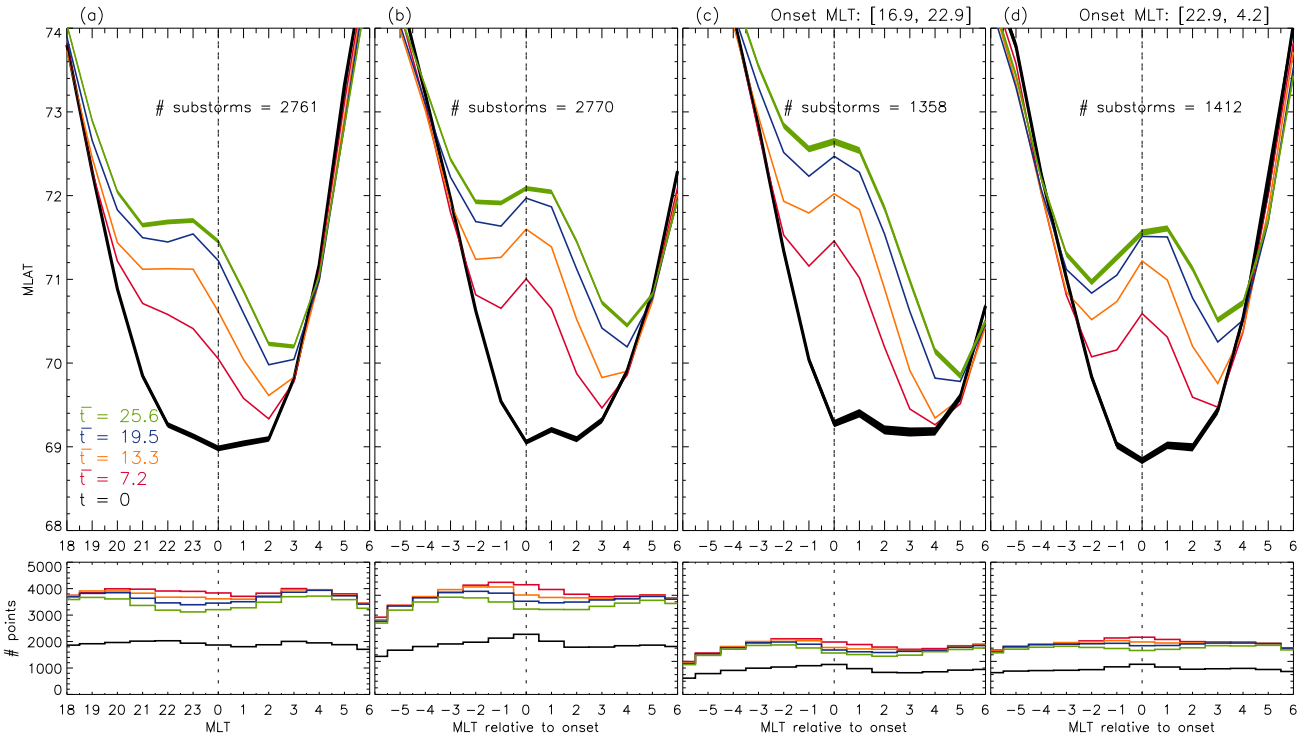

Figure 1. The figure shows average OCBs at fixed magnetic local times (a), and at magnetic local times relative to the MLT of the substorm onset (b, c, and d). Different colors indicate the time relative to substorm onset. Figures c and d are based on substorms with onset westward and eastward of 22.9 MLT, respectively. For the onset curves and the last expansion phase curves, the thickness constitutes the error intervals. The histograms in the lower row shows the number of points upon which each average is based. See Section 3.1 for details.

Table 1. The average magnetic latitude of the poleward boundary of the aurora at the onset magnetic local time, as a function of time [minutes] after substorm onset, from Figure 8 in *Mende et al.* [2003], *Gjerloev et al.* [2008], and this study. *Gjerloev et al.* [2008] used a normalized substorm time, and the numbers marked by * correspond to their $T = 0.5$ and $T = 1.0$, where $T = 1.0$ denotes the end of the expansion phase

Substorm time	*Mende et al.* [2003]	*Gjerloev et al.* [2008]	This study
0	68.0	68.7	69.0
7.2	70.3	-	71.0
13.3	70.9	71.7*	71.6
19.5	71.3	-	72.0
25.6	71.7	74*	72.1

noise is observed. The $t > 0$ curves all seem to follow the same pattern. This is a good indication that they represent real phenomena, and not something which is artificially produced by the fitting and selection methods. Although the curves become less smooth due to the reduced statistics, all the following figures show a consistent pattern which is hard to discount as artificial.

We can compare the boundary at the substorm onset location to two earlier studies using global UV imagers. *Mende et al.* [2003] looked at the average of 91 substorms seen by IMAGE WIC, and *Gjerloev et al.* [2008] studied 116 substorms seen by the Visible Imaging System (VIS) Earth camera on the Polar satellite. Our statistical data set is more than a factor of 20 larger than in these previous studies. In Table 1, we compare the average poleward boundary at the substorm onset MLT in this study (Fig. 1b) and in the studies by *Gjerloev et al.* [2008] and *Mende et al.* [2003]. Compared to the *Mende et al.* [2003] study, we consistently observe higher latitudes, but the difference is only 1° at onset, and decreasing after that. Comparison with *Gjerloev et al.* [2008] is more difficult to make, because they applied a normalization technique where $T = 0$ was defined as onset, and $T = 1$ as the end of the expansion phase. Their $T = 0$ boundary is only 0.3° equatorward of our onset boundary. In Table 1, we also present the boundary for $T = 0.5$ from *Gjerloev et al.* [2008], which compares well to the $\bar{t} = 13.3$ boundary in our study, and $T = 1.0$ which is further poleward than our last boundary. It is likely, from the monotonic poleward propagation in Figure 1 and from *Mende et al.* [2003], that the end of the expansion phase on average comes later than 25.6 minutes after onset, which might explain this discrepancy.

Mende et al. [2003] and *Gjerloev et al.* [2008] used a double Gaussian instead of a single Gaussian to fit the auroral oval. As stated earlier, the exclusion of many pronounced double ovals in the present study may represent a bias and contribute to the differences seen in Table 1. *Gjerloev et al.* [2008] also specifically required that the auroral bulge was in darkness. As we will show, this selection criterium has a profound influence on the average evolution of the substorm aurora, and also contributes to the differences.

3.3. Average OCBs for substorms selected according to dipole tilt angle

In Figure 2 we present two separate groups of substorms: Solid lines show substorms occurring when the dipole tilt angle was greater than 15°, and the dashed lines show substorms occurring when the dipole tilt angle was less than

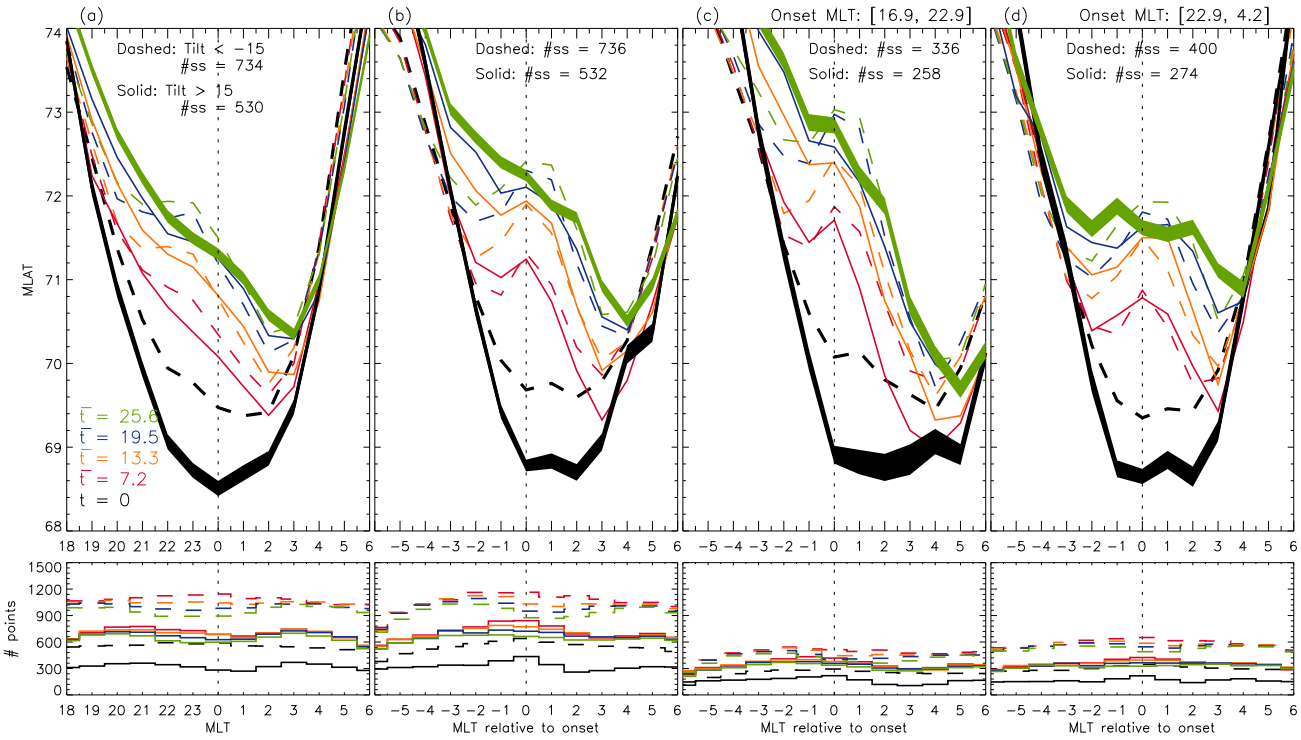

Figure 2. The figure shows average OCBs for substorms with onsets during periods when the dipole tilt angle was greater than 15° (solid) and less than −15° (dashed). The format is the same as in Figure 1. The format is also explained in Section 3.1. #ss denotes number of substorms in each bin.

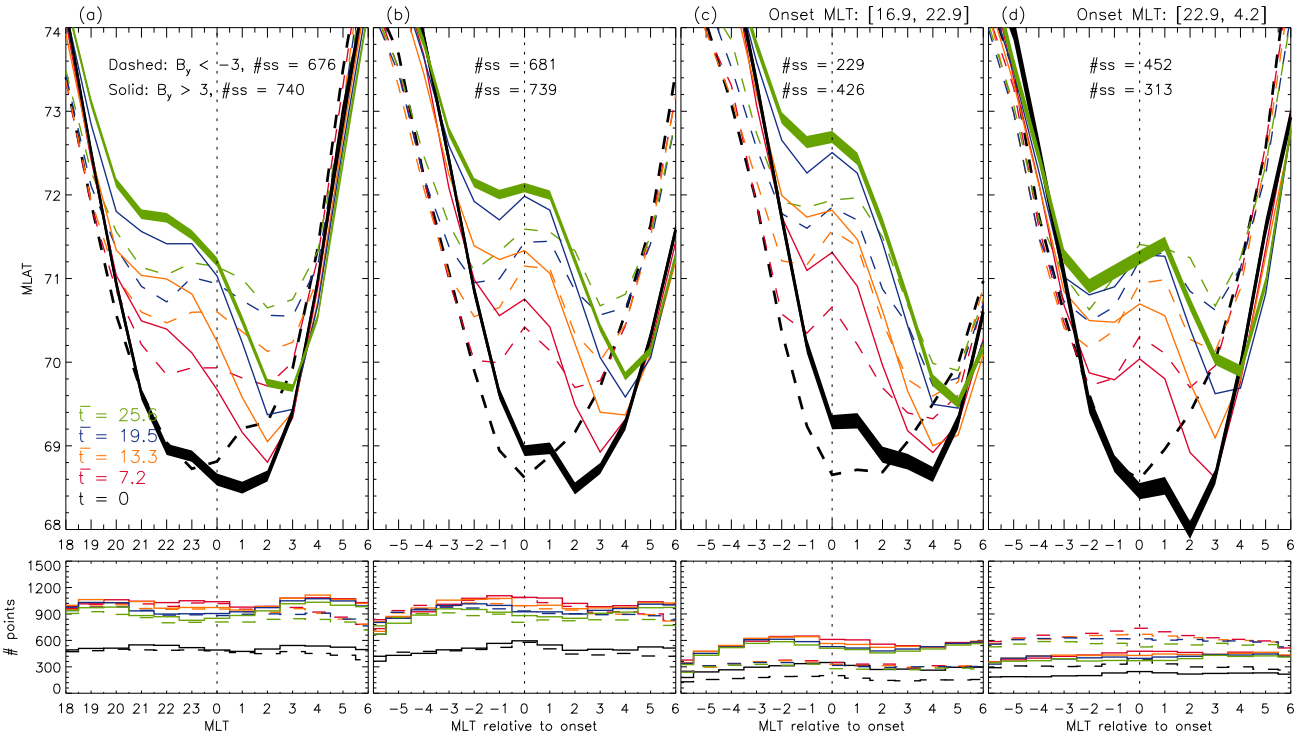

Figure 3. The figure shows average OCBs for substorms with onsets when the IMF $B_y < -3$ nT (dashed) and IMF $B_y > 3$ nT (solid). The format is the same as in Figure 1. The format is also explained in Section 3.1.

−15°. For observations in the southern hemisphere, we have changed the sign of the tilt angle, so that positive tilt always corresponds to sunlit conditions. As is clear from the lower panels in this figure, the Gaussian fit is less likely to succeed in sunlit conditions, and the error intervals (filled by black and green colors) are therefore only shown for positive tilt.

We first note that there is a clear tendency that the summer hemisphere OCB is on lower latitudes at substorm onset, than in the winter hemisphere. This agrees well with earlier studies [*Oznovich et al.*, 1993], and the presumption that the solar wind momentum, which affects the summer hemisphere more, displaces the oval towards midnight.

Two features stand out in the $t > 0$ curves: 1) The bulge is much more pronounced in the winter hemisphere. This is seen in all panels and for all $t > 0$. This tendency becomes increasingly clear as t increases. The difference is most pronounced when the OCBs are sorted according to their location relative to the onset MLT. 2) The asymmetry between seasons seen at onset is severely reduced during the expansion phase, and at some MLTs, even reversed. We show in Section 3.6 that the width of the oval increases faster in the summer, contributing to the apparent decrease in OCB asymmetry. The only MLTs where the winter OCB is leading the summer OCB is at the bulge. The asymmetries in the OCB shown in Figure 2a resemble the seasonal differences in the magnetic latitude of the substorm auroral electrojet, reported by *Wu et al.* [1991]. We discuss possible explanations for the asymmetries in Section 4.

Increasing the threshold of 15° gave the expected result of increased differences, but decreased statistics.

3.4. Average OCBs for substorms selected according to the value of IMF B_y

In Figure 3, we group substorms according to the average value of (time shifted) IMF B_y in the period 30 minutes prior to, and 20 minutes after onset. Dashed curves had $B_y < -3$ nT, and solid curves had $B_y > 3$ nT. Because of the Parker spiral configuration of the IMF, this figure is very similar to what we get if we select substorms according to $B_x > 3$ nT and $B_x < -3$ nT instead. In Section 3.5 we try to isolate the effect of the two components.

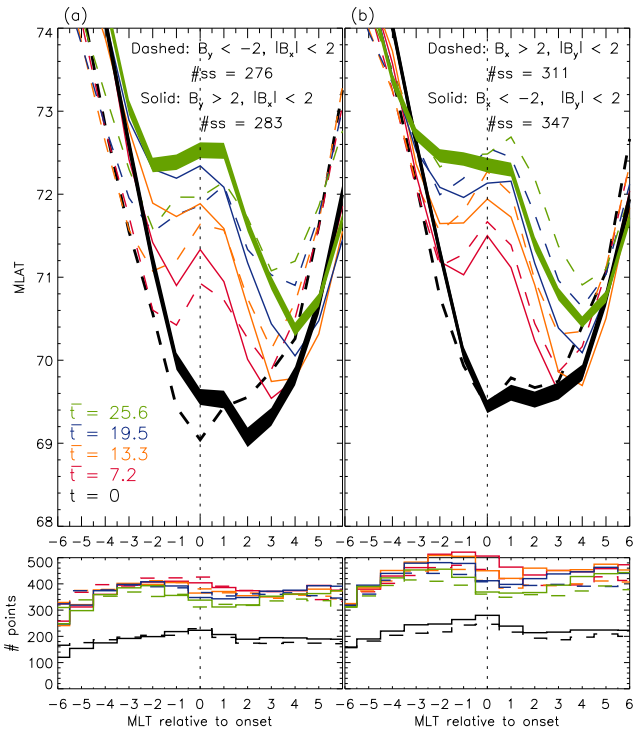

Figure 4. a) Average OCBs during substorms selected according to the value of B_y (a) and B_x (b), with the absolute value of the other component kept less than 2. Dashed lines denote substorms when the IMF orientation presumably was in the Parker spiral sector $B_y < 0$, $B_x > 0$, while the solid lines show $B_y > 0$, $B_x < 0$. Both figures show the OCBs at a given distance from the onset MLTs, similar to the format in Figures 1b, 2b, and 3b

At substorm onset, the OCBs are quite similarly shaped for different signs of B_y, but the boundaries for $B_y > 3$ nT are displaced dawnward, compared to $B_y < -3$ nT. This is most clear in panels b and c. During the expansion phase, the shift remains stable for the boundaries furthest out on the flanks, which are not affected by the emerging bulge. This shift in the boundary location is consistent with a global B_y dependent displacement of the auroral ovals, in the direction consistent with a net "penetration" of the IMF B_y into the magnetosphere (e.g., *Wing et al.* [1995]) and the "dipole plus uniform field" model, described by *Cowley et al.* [1991].

During the expansion phase ($t > 0$), there is a very clear tendency in Figure 3 that the poleward propagation of the OCB is skewed duskward when B_y is positive. A duskward skew was also apparent in Figure 1, where averages of all substorms were considered. However, this tendency seems to increase when B_y is positive. When B_y is negative, a much weaker duskward skew is seen in Figures 3a, b, and c. In Figure 3d, the $B_y < -3$ nT substorms are even skewed slightly dawnward.

The most striking effect of increasing the $|B_y|$ threshold beyond 3 nT was to move the average boundaries equatorward, while decreasing the statistics. Decreasing the threshold led to smaller asymmetries.

3.5. B_x effect on substorm expansion

Figure 4 shows OCBs grouped according to B_y (a) and B_x (b), with the other component kept less than 2, in absolute value. The format is the same as in panels b in the previous figures. Dashed curves correspond to the Parker spiral sector $B_y < -2$ nT and $B_x > 2$ nT, while solid curves correspond to $B_y > 2$ nT and $B_x < -2$ nT. We notice that Figure 4a is very similar to Figure 3b, as we might expect.

Figure 4b shows that the OCB is consistently on slightly higher latitudes when B_x is positive. This is the exact opposite of what we expect from a "partial penetration" of the IMF into the nightside magnetosphere. Since the boundaries sorted by B_y in the same Parker spiral sector shows the opposite asymmetry, we conclude that this is not an effect of the correlation between B_x and B_y.

The reduced threshold, 2 nT rather than the 3 nT threshold in Figure 3, was chosen to keep the number of substorms in each group high, despite the introduction of an additional constraint on the B_y (B_x) component in Figure 4a (4b).

3.6. Seasonal and IMF influence on the FWHM of the auroral oval

The definition of the OCB as being one FWHM poleward of the peak of the Gaussian introduces a certain ambiguity: Poleward motion of the OCB could signify a shift in the peak location, or a widening of the oval. To distinguish between these effects, we show in Figure 5 the distributions of the FWHM with a similar format and selection criteria as in panels b in Figures 1, 2, 3, and 4. This figure shows that, as expected, the oval becomes increasingly wide at the location of the bulge during the expansion phase. The widening of the oval is more severe dawnward of the onset, than duskward of the onset, accounting for at least some of the difference in OCB expansion between these regions.

Figure 5 also shows that the IMF orientation has only a small effect on the FWHM (c and d). However, during summer, the oval is generally wider than during winter (b). Further, the oval widens faster during the expansion phase in the substorms observed in the summer hemisphere. This

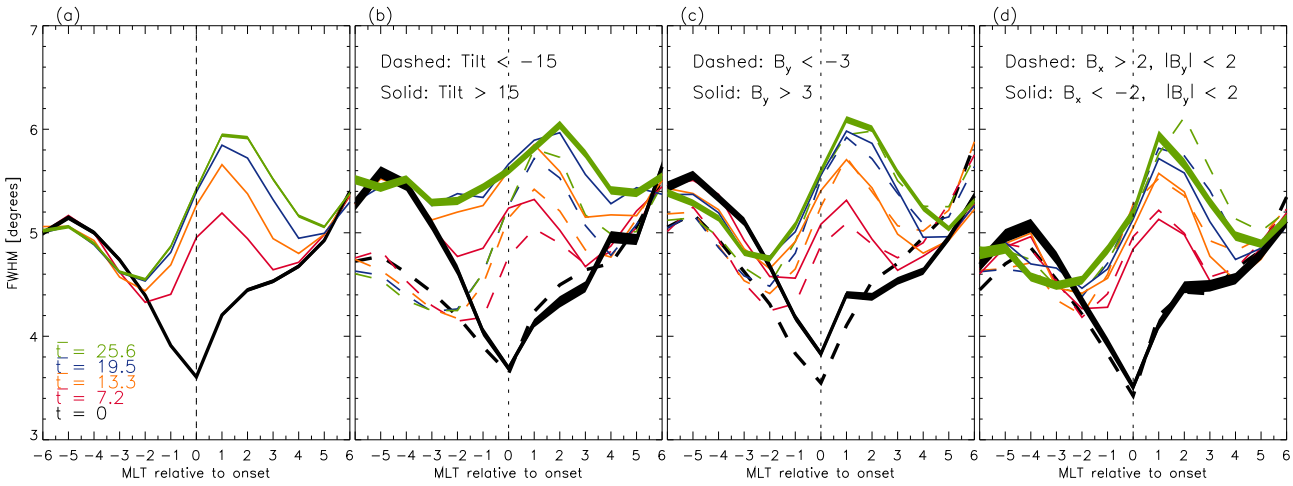

Figure 5. The full width at half maximum (FWHM) of the same intensity profiles used in a) Figure 1b, b) Figure 2b, c) Figure 3b, and d) Figure 4c. The number of points (n) which each average FWHM is based upon is shown below the mentioned figures. The width of the solid $t = 0$ and $\bar{t} = 25.6$ curves shows the standard error of the mean.

means that part of the stronger poleward expansion seen in the summer hemisphere, compared to the winter hemisphere, is due to a widening of the oval.

From the validation studies by *Carbary et al.* [2003] and *Boakes et al.* [2008], we can not be sure that Equation 2 is valid for all possible ranges of the FWHM. If this equation for the OCB introduces a FWHM dependent error, this shortcoming will however not affect the qualitative comparison between OCBs for different seasons and IMF orientations, unless the FWHM also changes with these parameters. Therefore, the more rapid widening of the auroral oval in the summer hemisphere could indicate that we overestimate the poleward propagation, compared to the winter hemisphere.

This is further supported by considering the reduction of open magnetic flux from substorm onset to $\bar{t} = 7.2$ indicated by Figure 2b: In the summer hemisphere, the boundary propagates further to the pole, traversing a larger area and hence more magnetic flux, than the OCB in the winter hemisphere. Since the amount of traversed magnetic flux should on average be equal during the two seasons, this difference constitutes a paradox unless it is balanced somewhere else along the OCB, which does not seem to be the case. This paradox could be explained if the widening of the oval causes an overestimation of the poleward propagation in the summer hemisphere, compared to the winter hemisphere. In that case, the asymmetry in the true OCB location could be less reduced than what is indicated by Figure 2, and the bulge could be even more prominent in the winter hemisphere, compared to the summer hemisphere.

4. Discussion

For the average OCBs at substorm onset, we have shown the following: 1) The average OCB in the summer hemisphere is displaced equatorward compared to the OCB in the winter hemisphere. 2) When $B_y > 3$ nT, the entire nightside polar cap is displaced towards dawn, compared to $B_y < -3$ nT conditions.

The seasonal asymmetry at onset can be understood as a the solar wind momentum affecting the summer hemisphere more than the winter hemisphere. This would expectedly lead to the oval being displaced further towards the nightside in the summer hemisphere, which is what we observe. The displacement between the average boundaries for different signs of B_y are in agreement with the "dipole plus uniform field" model, described by *Cowley et al.* [1991].

During the expansion phase of the substorm, the following features are seen in the statistics: 1) The summer hemisphere boundaries propagate poleward faster than the winter hemisphere boundaries, reducing the asymmetry seen at onset. 2) The bulge is much more pronounced in the substorms observed during winter, compared to the summer hemisphere substorms. 3) When $B_y > 3$ nT, the poleward propagation is stronger on the dusk side of the onset MLT, compared to $B_y < -3$ nT conditions. 4) When $B_x > 2$ nT, the OCB is on slightly higher latitudes during the substorm expansion phase, compared to $B_x < -2$ nT ($|B_y| < 2$ nT), implying a faster poleward propagation following the substorm onset when $B_x > 2$ nT.

These effects are more surprising, in terms of the direct effects of the solar wind and IMF on the magnetosphere. However, substorms are associated with significant and rapid changes in the tail, which to some extent can be considered to be internally driven. In the following, we discuss how inter-hemispheric differences in the shape of the polar cap can arise, as a consequence of these changes.

4.1. Asymmetric ionospheric response to magnetospheric convection

The poleward propagating OCB which is seen in the expansion phase of substorms signifies increased tail reconnection. Tail reconnection allows a reconfiguration of the magnetosphere, via convecting flux and plasma (e.g. *Cowley and Lockwood* [1992]). The enhancement in earthward magnetospheric convection subsequently excites equatorward convection in the ionosphere. However, the time it takes to get the ionosphere in a balanced motion with the magnetosphere likely depends on at least two factors: 1) The ionospheric conductance, which governs the collision frequency, and hence the effective friction with the neutral wind (e.g., *Song et al.* [2009]), as well as the susceptibility of the energy carried by Alfven waves (*Scholer* [1970]), and 2) the time history of the system, because of the inertia of the prevailing convection. Depending on seasons and IMF, these factors may very well be different in the two hemispheres, causing differences in ionospheric response to the expansion phase magnetospheric convection. As discussed by *Laundal et al.* [2010], inter-hemispheric differences in the ionospheric convection lead to inter-hemispheric differences in magnetic field line footprints, including field lines on the separatrix.

Ionospheric convection is observable directly from satellites and radars on the ground, and indirectly by magnetometers. Quite a few studies of its response to the onset

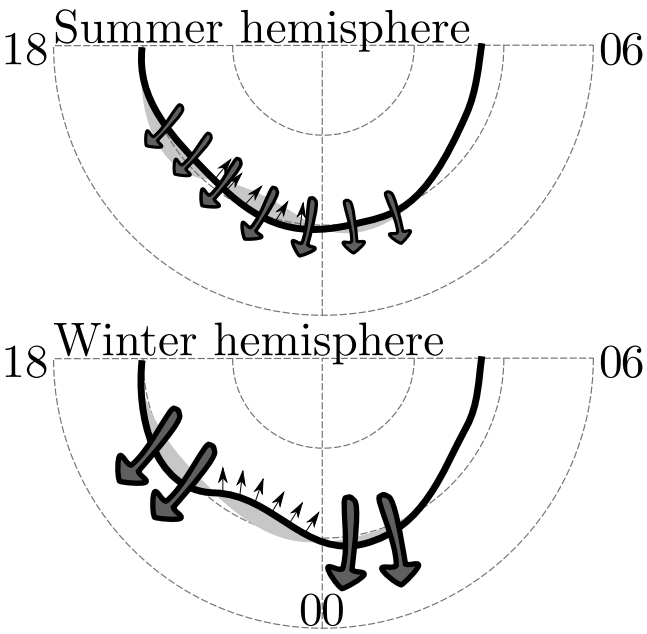

Figure 6. The black curves show the OCBs in the summer and winter hemispheres, and the shaded regions show the inter-hemispheric differences. The small poleward arrows denote propagation of the bulge caused by tail reconnection. The thick arrows show ionospheric convection, transporting the boundary equatorward. The observed OCB asymmetries seem to imply a more severe suppression of the convection at the bulge in the winter hemisphere, compared to the summer hemisphere. In neighboring regions, the convection is stronger in the winter hemisphere.

of substorm expansion phases exist. *Lyons et al.* [2001] reported a reduction in convection immediately following the onset. *Bristow and Jensen* [2007] reported an overall reduction, but also a rotation to a more meridional flow. *Grocott et al.* [2002] and *Provan et al.* [2004] observed an enhancement of the convection. *Blanchard et al.* [1997] found that there is an increase in ionospheric convection, but it may lag by ~ 20 minutes compared to the onset, while *Grocott et al.* [2009] found the convection response to depend on the latitude of the substorm onset. It is evident that the ionospheric convection response to substorm onset and expansion can be highly variable. Neither of these studies considered seasonal differences.

Figure 6 shows a conceptual illustration of two main features in the ionospheric convection, which can explain the observed differences in the poleward propagation of the OCB: 1) The equatorward ionospheric convection is generally stronger in the winter hemisphere, compared to the summer hemisphere, effectively reducing the asymmetry seen at onset. 2) The presence of a bulge in the winter hemisphere, and not in the summer hemisphere, indicates that the convection is weaker in the bulge region in the winter hemisphere. A local suppression of the ionospheric flow in the substorm bulge was observed by e.g., *Provan et al.* [2004]. These authors also observed fast flow in the surrounding regions, in agreement with our proposed pattern for the winter hemisphere convection.

If this explanation is true, our observations imply that the flow suppression in the bulge is more prominent in the winter hemisphere. It also implies that apart from the bulge, the ionospheric convection is stronger in the winter hemisphere. These differences could be caused by the differences in conductance, changing the ionospheric friction and coupling efficiency, as mentioned above. This is in agreement with *Milan et al.* [2009], who pointed out that the brighter the substorm aurora is, the more sluggish the ionospheric convection response will be, because of the increased conductance. During substorms, the conductance is expected to be high and relatively smooth when the ionosphere is sunlit, and low in the dark hemisphere, except for at the bulge, where it is increased by energetic particle precipitation.

In the event study by *Laundal et al.* [2010], in which simultaneous measurements from IMAGE WIC and Polar VIS Earth were used, large asymmetries were observed in the polar caps, which developed during the expansion phase of a substorm. The asymmetries were largest in the region with the strongest poleward propagation, in good agreement with what we now document statistically. During that event $|B_y|$ was less than 2 nT, and B_x was strongly positive. The observed asymmetries were in general opposite to what we would expect from the statistical average OCB for different B_x, shown in Figure 4b. However, there was also a large seasonal difference, with the winter hemisphere boundaries going further poleward than the summer hemisphere boundaries. This is in agreement with the latitudinal asymmetries in the bulge seen in Figure 2b. This suggests that the asymmetries in the event study was generated by seasonal differences, rather than by B_x. We note that the asymmetries that were observed by *Laundal et al.* [2010] were much larger, up to 5° different in the two hemispheres, than what is seen in the average case.

4.2. Closing of field lines with asymmetric footprints

The orientation of the IMF is also known to cause inter-hemispheric differences in convection (e.g. *Heppner and Maynard* [1987]; *Ruohoniemi and Greenwald* [2005]; *Haaland et al.* [2007]). These asymmetries are often attributed to curvature forces acting on newly opened magnetic field lines on the dayside (e.g. *Jørgensen et al.* [1972]; *Cowley et al.* [1991]), and subsequent pressure gradients in the lobes (*Khurana et al.* [1996]). *Provan et al.* [2004] showed that B_y also affects the convection during substorms. Differences in ionospheric convection due to different signs of B_y, and possibly B_x, may therefore contribute to the observed asymmetries in Figure 3 and 4. However, another effect also exists, which could be important in generating the asymmetries for different signs of B_y. As magnetic field lines approach the tail X-line, they will not be able to reconnect with field lines with symmetrical footprints in the opposite hemisphere: For positive (negative) B_y, the northern hemisphere field lines will have footprints duskward (dawnward) of the field lines in the southern hemisphere with which they merge [*Østgaard et al.*, 2004]. We would expect that this leads to an increased poleward propagation on the dusk side when $B_y > 0$, and an increased poleward propagation on the dawn side when $B_y < 0$. This is exactly what we observe.

There is no clear analogy to the asymmetric tail reconnection mechanism for different signs of B_x. It does seem reasonable that B_x could have an effect on ionospheric convection, however this has not yet been firmly established. Figure 4b shows that the OCB close to the MLT of the onset on average ends up on higher latitudes when B_x is positive, indicating that the equatorward convection in this region is stronger when B_x is negative.

4.3. Self-reinforcing asymmetries

Another effect, which might explain part of the changing asymmetries seen during substorm expansion phase, arises from the definition of the Apex coordinate system (and any other similar coordinate systems). In this coordinate system, the amount of magnetic flux in equally wide concentric

circles centered at the origin is reduced towards the pole (the amount of flux is constant along these circles, i.e. in the longitudinal direction). Assume that flux closure in the tail is limited to a fixed longitudinal region. If a given amount of magnetic flux closes in the tail, this necessitates a poleward propagation, which encompasses an area corresponding to the magnetic flux which has been closed. The latitudinal dimension of this area (assuming the longitudinal dimension to be fixed), will be different, measured in degrees, depending on where the boundary was before the reconnection event. If the boundaries were asymmetrical in the two hemispheres, the asymmetry is expected to be reinforced as flux closes. This may explain some of the growing asymmetries that we observe for different signs of B_y, but likely not all.

For different signs of tilt angle (Figure 2), the onset asymmetry was observed to be reduced, and even reversed. This can only be explained by different convection in the two hemispheres, and the self-reinforcement effect implies that the convection must be stronger than what is implied by a strictly linear relation between OCB motion in the two hemispheres. For a more detailed explanation of this effect, see *Laundal et al.* [2010].

5. Conclusions

A statistical analysis of auroral images from 2770 substorms have revealed the following:

1) The OCB propagates poleward monotonically during the first 26 minutes after substorm onset. The rate of poleward propagation decreased with each 7 minutes step. At the onset MLT, a local maximum (bulge) in the poleward boundary develops. The poleward propagation is stronger to the west of the onset than to the east. 2) The poleward expansion of the OCB is faster during summer than in winter, largely reversing the asymmetry seen at onset. The shape of the polar cap is also more circular for summer conditions, with almost no sign of a bulge. 3) At substorm onset, an almost uniform dawnward shift is seen in the polar cap boundary for $B_y > 3$ nT, compared to $B_y < -3$ nT. During expansion phase, the polar caps have different shapes for the two signs of B_y, attaining a more pronounced east/west asymmetry with respect to the onset location when $B_y > 3$ nT. 4) Positive B_x produces a faster poleward propagation, and a more pronounced bulge than negative B_x. The effect is different than what we would expect if the asymmetry was an effect of the high correlation with B_y.

Rising asymmetries in the OCB for different signs of IMF B_x, B_y and dipole tilt imply instantaneous differences between the hemispheres, as previously reported by *Laundal et al.* [2010]. The observations presented here further imply that the convection in the first few minutes after substorm onset is different if there are seasonal differences between the hemispheres. Differences in ionospheric convection may also play a role in the observed asymmetries for different signs of B_x and B_y. In the case of B_y, however, much of the growing asymmetry may arise from increased tail reconnection of field lines with asymmetrical footprints.

A logical next step in exploring the proposed differences in ionospheric convection is to study the seasonal dependence of the convection response to magnetospheric substorms. This may be done in event studies, using conjugate measurements, or statistically. Global MHD models which are able to reproduce the observed asymmetrical polar cap boundary could also be used to test this hypothesis.

Appendix: Identifying the open/closed boundary

Each image was divided into 1 h wide MLT sectors, producing 24 profiles of intensity as a function of latitude (1° resolution was used). To check that the function $f(\lambda)$ (see Eq. 1) is successfully fitted to an observed intensity profile, we first check that the following criteria are fulfilled:

- c_1: The minimum height of the Gaussian peak, a_1, must be > 50 WIC counts.
- c_2: The position of the peak, $a_1 \in [50°, 85°]$ magnetic latitude.
- c_3: The peak is located at least one full width at half maximum (FWHM) from the last latitude bin in the field of view.
- c_4: The FWHM must be greater than 1.
- c_5: The FWHM must be less than 30% of the latitudes spanned by the field of view.

These requirements primarily ensure that the Gaussian is fitted to a positive, relatively sharp, and fully observed bump in the intensity profile (presumably the oval), and not some artifact of the background. Having passed the above requirements, we make a more subtle test that the fitted profile resembles the observations:

$$\frac{1}{\overline{f_{obs}}}\sqrt{\sum_{i=1}^{N}\frac{(f(\lambda_i)-f_{obs}(\lambda_i))^2}{N}} < 0.2 \qquad \text{(A1)}$$

i.e., the root mean square deviation (RMSD), divided by the mean of the observed intensity, must be less than 0.2. The RMSD requirement is a test of how well the function $f(\lambda)$ represents the actual intensity. In the case of double ovals, a single Gaussian is a poor representation, and the profile will in most cases be rejected by this requirement (the conditions c_1 to c_5 would in general not reject a double oval).

Carbary et al. [2003] used the fractional standard deviation (FSD) as a measure of how well the intensity was fitted, instead of the RMSD. The FSD is defined as

$$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{f(\lambda_i)-f_{obs}(\lambda_i)}{f_{obs}(\lambda_i)}\right)^2}, \qquad \text{(A2)}$$

which they required should be less than 0.2. The principal difference between these two methods is that the FSD becomes very sensitive to deviations at a few single points if the observed intensity is low (i.e. at the background), whereas the RMSD condition is normalized to the mean. This means that our method is more liberal.

To compare the *Carbary et al.* [2003] method to the method used in this study, we investigate the effects of the FSD and RMSD acceptance criteria. Equation 2 is intended to give a representation of the poleward boundary of the aurora. Whether or not this is accomplished can be fairly easily determined by eye. Therefore, we have looked at 4000 intensity profiles, 2000 randomly picked from the entire data set, and 2000 which include the substorm onset aurora. Each profile was categorized as a successful or a failed boundary determination, constituting a set answer to which the automatic methods can be compared. In the case of random profiles (substorm onset profiles), 533 (133) out of 2000 fitted profiles gave a poor representation of the boundary.

The effects of the acceptance criteria are summarized in Table 2. The criteria are given at the first row and columns: 1) $\bigcap_{i=1}^{5} c_i$: Requirements c_1 to c_5, listed above, are all fulfilled, 2) The fractional standard deviation is less than 0.2, and 3) The root mean square deviation divided by the mean of the observed intensity (RMSD') is less than 0.2. The upper half of the table is based on the random profiles, while the lower half is based on the substorm onset profiles. The upper numbers in each cell show the percentage that was accepted by the given combination of criteria. That is, the

FSD requirement accepted 32.7% (22.5%) of the functions fitted to random (onset) intensity profiles, while the RMSD requirement accepted 77.8% (88.0%) of the fits (diagonal cells). The condition used in this study is that both the RMSD and the c_1 to c_5 criteria are fulfilled. This combination accepted 63.2% (86.9%) of the fitted functions in the case of random (onset) profiles (upper right cell). Compared to the *Carbary et al.* [2003] method, we get approximately 2 (random) and 4 (onset) times more accepted boundaries.

An important question with respect to the statistical analysis is whether our method includes a large amount of erroneously accepted boundaries. The lower numbers (italic) show the percentage of the boundaries that were accepted by the given combination of criteria, which should have been rejected. Out of the boundaries accepted by both the RMSD criterium and the c_i criteria, 4.3% (3.5%) should have been rejected for random (onset) profiles. Only 1.6% (0.6%) of the boundaries accepted by the *Carbary et al.* [2003] conditions should have been rejected. Although our method accepts a larger fraction of wrong boundaries, the contribution to the average (which we are interested in) is small: If there are, say 5% wrong boundaries, which represent a systematic error of 5° (we saw no tendency towards a systematic error while manually inspecting the boundaries, and many of the poor fits produced errors smaller than 5°), the overall contribution to the average is $0.05 \cdot 5^\circ = 0.25^\circ$. This is less than the error with which UV images can determine the OCB, according to *Carbary et al.* [2003]; *Boakes et al.* [2008]. Thus, we conclude that our method introduces negligible new errors when compared to previous methods, while significantly increasing the number of boundaries that can be used in the statistical analysis. As a final test of the validity of our method, we did a comparison of the figures presented in the present paper, with similar figures with the RMSD criterium replaced by the FSD criterium (not shown). The same patterns clearly emerged, although they appeared less smooth, which is expected from the reduced statistics.

We note that our method differs from the *Carbary et al.* [2003] method, also by the absolute intensity requirement (c_1), which in our case is so low that it has practically no other effect than making sure the Gaussian peak is positive. We also neglected the requirement that the Gaussian should be no smaller than 20% of the background at the peak location ($a_1/(a_4+a_5a_2+a_6a_2^2) > 0.2$). This requirement leads to the rejection of many well-defined boundaries. For instance, if a peak of 400 WIC counts is embedded in a sunlit background of 2000 counts, it would be automatically rejected by this requirement, although such a scenario often involves smooth profiles which are easily fitted by $f(\lambda)$. For this reason, and since sunlight/darkness is an important selection criterium in our analysis, we omit this requirement.

Table 2. Table showing the success rate (upper number in each cell) and the fraction of erroneously accepted boundaries (lower number in each cell) for various combinations of goodness of fit tests. The upper half of the table is based on manual inspection of 2000 boundaries randomly chosen from the full data set used in this paper. The lower half of the table is based on 2000 profiles which include the substorm onset. $\bigcap_{i=1}^{5} c_i$ denotes the requirement that conditions $1-5$ (see text) are all fulfilled. RMSD' denotes the root mean square deviation, divided by the mean of the observed intensity, and FSD is the fractional standard deviation. The upper right corners of the two halves correspond to the requirement that were used in this study.

	RMSD' < 0.2	FSD < 0.2	$\bigcap_{i=1}^{5} c_i$
RMSD' < 0.2	77.8 *18.9*	32.5 *7.9*	63.2 *4.3*
FSD < 0.2		32.7 *7.2*	26.7 *1.6*
$\bigcap_{i=1}^{5} c_i$			78.7 *7.4*
RMSD' < 0.2	88.0 *4.3*	22.5 *0.9*	86.9 *3.5*
FSD < 0.2		22.5 *0.9*	22.0 *0.6*
$\bigcap_{i=1}^{5} c_i$			97.5 *4.5*

Acknowledgments. The authors thank S. B. Mende and the IMAGE FUV team for the use of IMAGE FUV data. We thank C. Smith for the ACE magnetic field data and D. McComas for the ACE solar wind data. We also thank Arne Åsnes for contributing to the processing of the ACE data.

This study was financed by the IPY-ICESTAR project, The Research Council of Norway, contract 176045/S30

References

Akasofu, S.-I., The development of the auroral substorm, *Planet. Space Sci.*, *12*, 273–282, 1964.

Baumjohann, W., M. Hesse, S. Kokobun, T. Mukai, T. Nagai, and A. A. Petrukovich, Substorm dipolarization and recovery, *J. Geophys. Res.*, *104*, 24,995–25,000, 1999.

Blanchard, G. T., L. R. Lyons, and O. de la Beaujardire, Magnetotail reconnection rate during magnetospheic substorms, *J. Geophys. Res.*, *102*, 14,303–24,312, 1997.

Boakes, P. D., S. E. Milan, G. A. Abel, M. P. Freeman, G. Chisham, B. Hubert, and T. Sotirelis, On the use of IMAGE FUV for estimating the latitude of the open/closed magnetic field line boundary in the ionosphere, *Ann. Geophys.*, pp. 2759 – 2769, 2008.

Boakes, P. D., S. E. Milan, G. A. Abel, M. P. Freeman, G. Chisham, and B. Hubert, A statistical study of the open magnetic flux content of the magnetosphere at the time of substorm onset, *Geophys. Res. Lett.*, doi:10.1029/2008GL037059, 2009.

Bristow, W. A., and P. Jensen, A superposed epoch study of SuperDARN convection observations during substorms, *J. Geophys. Res.*, *112*, doi:10.1029/2006JA012049, 2007.

Brittnacher, M., M. Fillingim, G. Parks, G. Germany, and J. Spann, Polar cap area and boundary motion during substorms, *J. Geophys. Res.*, *104*, 12,251 – 12,262, 1999.

Carbary, J. F., T. Sotirelis, P. T. Newell, and C.-I. Meng, Auroral boundary correlations between UVI and DMSP, *J. Geophys. Res.*, *108*, doi:10.1029/2002JA009378, 2003.

Cowley, S. W. H., and M. Lockwood, Excitation and decay of solar wind-driven flows in the magnetosphere-ionosphere system, *Ann. Geophys.*, pp. 103–115, 1992.

Cowley, S. W. H., J. P. Morelli, and M. Lockwood, Dependence of convective flows and particle precipitation in the high-latitude dayside ionosphere on the x and y components of the interplanetary magnetic field, *J. Geophys. Res.*, *96*, 5557–5564, 1991.

Craven, J. D., and L. A. Frank, Latitudinal motions of the aurora during substorms, *J. Geophys. Res.*, *92*, 4565–4573, 1987.

Elphinstone, R. D., et al., The double oval UV auroral distribution 1. Implications for the mapping of auroral arcs, *J. Geophys. Res.*, *100*, 12,075–12,092, 1995.

Frey, H. U., and S. B. Mende, Substorm onsets as observed by IMAGE-FUV, in *Substorms VIII: Proc. of the 8th International Conference on Substorms*, pp. 71–76, Univ. of Calgary Press, Calgary, 2006.

Frey, H. U., S. B. Mende, V. Angelopoulos, and E. F. Donovon, Substorm onset observations by IMAGE-FUV, *J. Geophys. Res.*, *109*, doi:10.1029/2004JA010607, 2004.

Gjerloev, J. W., R. A. Hoffman, J. B. Sigwarth, L. A. Frank, and J. B. H. Baker, Typical auroral substorm: A bifurcated oval, *J. Geophys. Res.*, *113*, doi:10.1029/2007JA012431, 2008.

Grocott, A., S. W. H. Cowley, J. B. Sigwarth, J. F. Watermann, and T. K. Yeoman, Excitation of twin-vortex flow in the nightside high-latitude ionosphere during an isolated substorm, *Ann. Geophys.*, *20*, 1577–1601, 2002.

Grocott, A., J. A. Wild, S. E. Milan, and T. K. Yeoman, Superposed epoch analysis of the ionospheric convection evolution during substorms: Onset latitude dependence, *Ann. Geophys.*, *27*, 591–600, 2009.

Haaland, S. E., G. Paschmann, M. Forster, J. M. Quinn, R. B. Torbert, C. E. McIlwain, H. Vaith, P. A. Puhl-Quinn, and C. A. Kletzing, High-latitude plasma convection from cluster EDI measurements: Method and IMF-dependence, *Ann. Geophys.*, pp. 239–253, 2007.

Henderson, M. G., et al., Substorms during the 1011 August 2000 sawtooth event, *J. Geophys. Res.*, *111*, doi: 10.1029/2005JA011366, 2006.

Heppner, J. P., and N. C. Maynard, Empirical high-latitude electric field models, *J. Geophys. Res.*, *92*, 4467–4489, 1987.

Huang, C.-S., A. D. DeJong, and X. Cai, Magnetic flux in the magnetotail and polar cap during sawteeth, isolated substorms, and steady magnetospheric convection events, *J. Geophys. Res.*, *114*, doi:10.1029/2009JA014232, 2009.

Hubert, B., S. E. Milan, A. Grocott, S. W. H. Cowley, and J. C. Gérard, Open magnetic flux and magnetic flux closure during sawtooth events, *Geophys. Res. Lett.*, *35*, doi: 10.1029/2008GL036374, 2008.

Jørgensen, T. S., E. Friis-Christensen, and J. Wilhjelm, Interplanetary magnetic-field directions and high-latitude ionospheric currents, *J. Geophys. Res.*, *77*, 1976–1977, 1972.

Khurana, K. K., R. J. Walker, and T. Ogino, Magnetospheric convection in the presence of interplanetary magnetic field B_y: A conceptual model and simulations, *J. Geophys. Res.*, *101*, 4907–4916, 1996.

Laundal, K. M., and N. Østgaard, Asymmetric auroral intensities in the earths northern and southern hemispheres, *Nature*, *460*, 491–493, doi:10.1038/nature08154, 2009.

Laundal, K. M., N. Østgaard, K. Snekvik, and H. U. Frey, Interhemispheric observations of emerging polar cap asymmetries, *J. Geophys. Res.*, *115*, doi:10.1029/2009JA015160, 2010.

Liou, K., P. T. Newell, and C.-I. Meng, Seasonal effects on auroral particle acceleration and precipitation, *J. Geophys. Res.*, *106*, 5531–5542, 2001.

Lyons, L. R., J. M. Ruohoniemi, and G. Lu, Substorm-associated changes in large-scale convection during the november 24, 1996, geospace environment modeling event, *J. Geophys. Res.*, *106*, 397–405, 2001.

Mende, S. B., H. U. Frey, B. J. Morsony, and T. J. Immel, Statistical behavior of proton and electron auroras during substorms, *J. Geophys. Res.*, *108*, doi:10.1029/2002JA009751, 2003.

Mende, S. B., et al., Far ultraviolet imaging from the IMAGE spacecraft. 2. wideband FUV imaging, *Space Sci. Rev.*, *91*, 271–285, 2000.

Meng, C.-I., R. H. Holzworth, and S.-I. Akasofu, Auroral circle - delineating the poleward boundary of the quiet auroral belt, *J. Geophys. Res.*, *82*, 164–172, 1977.

Milan, S. E., M. Lester, S. W. H. Cowley, K. Oksavik, M. Brittnacher, R. A. Greenwald, G. Sofko, and J.-P. Villain, Variations in the polar cap area during two substorm cycles, *Ann. Geophys.*, *21*, 1121–1140, 2003.

Milan, S. E., G. Provan, and B. Hubert, Magnetic flux transport in the dungey cycle: A survey of dayside and nightside reconnection rates, *J. Geophys. Res.*, *112*, doi: 10.1029/2006JA011642, 2007.

Milan, S. E., A. Grocott, C. Forsyth, S. M. Imber, P. D. Boakes, and B. Hubert, A superposed epoch analysis of auroral evolution during substorm growth, onset and recovery: Open magnetic flux control of substorm intensity, *Ann. Geophys.*, pp. 659–668, 2009.

Østgaard, N., S. B. Mende, H. U. Frey, T. J. Immel, L. A. Frank, J. B. Sigwarth, and T. J. Stubbs, Interplanetary magnetic field control of the location substorm onset and auroral features in the conjugate hemisphere, *J. Geophys. Res.*, *109*, doi: 10.1029/2003JA010370, 2004.

Østgaard, N., N. A. Tsyganenko, S. B. Mende, H. U. Frey, T. J. Immel, M. Fillingim, L. A. Frank, and J. B. Sigwarth, Observations and model predictions of substorm auroral asymmetries in the conjugate hemispheres, *Geophys. Res. Lett.*, *32*, doi:10.1029/2004GL022166, 2005.

Østgaard, N., S. B. Mende, H. U. Frey, J. B. Sigwarth, A. Aasnes, and J. Weygand, Auroral conjugacy studies based on global imaging, *J. Atmos. Terr. Phys.*, *69*, 249–255, 2007.

Oznovich, I., R. W. Eastes, R. E. Hoffman, M. Tur, and I. Glaser, The aurora at quiet magnetospheric conditions: Repeatability and dipole tilt angle dependence, *J. Geophys. Res.*, *98*, 3789–3797, 1993.

Provan, G., M. Lester, S. B. Mende, and S. E. Milan, Statistical study of high-latitude plasma flow during magnetospheric substorms, *Ann. Geophys.*, *22*, 3607–3624, 2004.

Richmond, A. D., Ionospheric electrodynamics using magnetic apex coordinates, *J. Geomag. Geoelectr.*, *47*, 191–212, 1995.

Ruohoniemi, J. M., and R. A. Greenwald, Dependencies of high-latitude plasma convection: Consideration of interplanetary magnetic field, seasonal, and universal time factors in statistical patterns, *J. Geophys. Res.*, *110*, doi: 10.1029/2004JA010815, 2005.

Scholer, M., On the motion of artificial ion clouds in the magnetosphere, *Planet. Space Sci.*, *18*, 977, 1970.

Song, P., V. M. Vasyliunas, and X.-Z. Zhou, Magnetosphere-ionosphere/thermosphere coupling: Self-consistent solutions for a one-dimensional stratified ionosphere in three-fluid theory, *J. Geophys. Res.*, *114*, doi:10.1029/2008JA013629, 2009.

Wang, H., H. Lühr, S. Y. Ma, and H. U. Frey, Interhemispheric comparison of average substorm onset locations: Evidence for deviation from conjugacy, *Ann. Geophys.*, *25*, 989–999, 2007.

Weimer, D. R., Correction to "Predicting interplanetary magnetic field (IMF) propagation delay times using the minimum variance technique", *J. Geophys. Res.*, *109*, doi: 10.1029/2004JA010691, 2004.

Weimer, D. R., D. M. Ober, N. C. Maynard, M. R. Collier, D. J. McComas, N. F. Ness, C. W. Smith, and J. Watermann, Predicting interplanetary magnetic field (IMF) propagation delay times using the minimum variance technique, *J. Geophys. Res.*, *108*, doi:10.1029/2002JA009405, 2003.

Wing, S., P. T. Newell, D. G. Sibeck, and K. B. Baker, A large statistical study of the entry of interplanetary magnetic field y-component into the magnetosphere, *Geophys. Res. Lett.*, *22*, 2083–2086, 1995.

Wu, Q., T. J. Rosenberg, L. J. Lanzerotti, C. G. Maclennan, and A. Wolfe, Seasonal and diurnal variations of the latitude of the westward auroral electrojet in the nightside polar cap, *J. Geophys. Res.*, *96*, 1409–1419, 1991.

H. U. Frey, Space Sciences Laboratory, University of California, 7 Gauss Way, Berkeley, CA 94720-7450, USA. (hfrey@ssl.berkeley.edu)

K. M. Laundal, University of Bergen, Dept. of Physics and Technology, Allegaten 55, 5007-Bergen, Norway. (karl.laundal@ift.uib.no)

N. Østgaard, University of Bergen, Dept. of Physics and Technology, Allegaten 55, 5007-Bergen, Norway. (nikolai.ostgaard@ift.uib.no)

J. M. Weygand, Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California, USA. (jweygand@igpp.ucla.edu)