1

Magnetosphere-Ionosphere Coupling, Past to Future

James L. Burch

ABSTRACT

Prior to the 1970s, magnetospheric physics and upper atmospheric/ionospheric physics were separate scientific disciplines with separate space missions and separate theory and modeling programs. This situation led to a certain labeling (of scientific programs, scientific society sections, conferences, and even scientists), and this labeling was limiting scientific advances. Although some of this labeling still persists, it has largely become recognized that the upper atmosphere, ionosphere, magnetosphere, and the nearby solar wind comprise a single coupled system of geospace that must be studied together. This review traces some of the early concepts of magnetosphere-ionosphere (M-I) coupling through the past four decades and makes suggestions for future progress.

1.1. INTRODUCTION

The 1974 Yosemite Conference on Magnetosphere-Ionosphere Coupling was a unique event during which leading scientists in both magnetospheric and ionospheric physics met together in a remote location to examine in a unique way not only the overlap but also the interrelationships of their previously quite separate disciplines. Since M-I coupling as a research field has progressed greatly over the past 40 years, it is perhaps informative to trace some of the instances in which coupled magnetospheric and ionospheric phenomena were just beginning to be appreciated in a meaningful way and describe how these ideas have evolved to the present and into the future.

During this same time somewhat controversial theories for the production of a polar wind, which populates the magnetosphere with ionospheric plasma, were developed and ultimately became widely accepted [e.g., Banks and Holzer, 1968]. In this same era, Vasyliunas [1970] developed a mathematical theory of M-I coupling that formed the basis for many theoretical advances in the field [e.g., Wolf, 1975].

Starting in the early 1970s, satellite measurements began to show that cold ionospheric particles (mainly H+ and He+) are important constituents of the inner and middle magnetosphere [Chappell et al., 1970] and that energetic heavy ions (mainly O+) precipitate into the low-altitude auroral zone during geomagnetic storms [Sharp et al., 1972]. While H+ ions, which dominate magnetospheric plasmas at all energies, can have their origins both in the solar wind and the ionosphere, the widespread prevalence of O+ ions, which are almost exclusively from the ionosphere, suggested that the ionospheric plasma source is important and capable of supplying most if not all of magnetospheric plasma [Chappell et al., 1987].
New data sets and discoveries in that epoch were mainly responsible for the advent of M-I coupling science. One new data set that came on line was generated by the Chatanika Radar facility, which pioneered the use of the incoherent scatter technique to derive large-scale plasma convection patterns [Brekke et al., 1974]. These convection patterns can be mapped into the magnetosphere to help gauge and visualize global magnetospheric dynamics. Another landmark discovery was auroral kilometric radiation (AKR), which was originally referred to as terrestrial kilometric radiation (TKR) [Gurnett, 1974; Alexander and Kaiser, 1976]. Since AKR beams outward from the auroral regions, it was only first observed many years after the discovery of radio emissions from Saturn and Jupiter [Kaiser and Stone, 1975]. In the case of Jupiter, the frequencies are much higher so that the so-called decametric radiation can be observed from the Earth’s surface.

By far the strongest channel for coupling between the magnetosphere and ionosphere is the auroral oval and its extension into space. In the early 1970s, auroral particles first began to be observed from orbiting spacecraft [e.g., Frank and Ackerson, 1971; Winningham et al., 1973]. Sounding rocket measurements of auroral electrons had shown earlier that their energy spectra were monoenergetic and hence consistent with acceleration by an electric field component aligned along the magnetic field [McIlwain, 1960]. Subsequent measurements, however, showed that lower-energy electrons also precipitated into the aurora along with the monoenergetic beams [Frank and Ackerson, 1971]. Some controversy therefore arose about the source of the low-energy electrons, and this controversy was resolved by Evans [1974], who showed that they were backscattered and secondary electrons trapped between the parallel potential drop and the ionosphere. The possibility of Alfvén-wave acceleration of auroral electrons was investigated by Hasegawa [1976]. Later on, measurements from the FAST spacecraft showed that Alfvén-wave acceleration is an important phenomenon especially near the polar-cap boundary [e.g., Chaston et al., 2003].

Another auroral phenomenon associated with M-I coupling is the stable auroral red (SAR) arc, which appears at mid-latitudes during magnetic storms. These arcs are produced either by Coulomb collisions between ring current particles and plasmaspheric electrons, electron acceleration by resonant wave interactions along magnetic field lines, or possibly precipitation of energetic electrons [Hoch, 1973]. These possibilities started to be examined closely during the early 1970s, and later satellite measurements combined with auroral imaging triggered further work in the 1980s, but research on the source of SAR arcs is still ongoing [Kozyra et al., 1997].

Starting from these early observations, the following sections trace progress and consider future directions in a subset of important M-I coupling phenomena. Related M-I coupling phenomena are also described that are observed at other planets, particularly Saturn, which, while vastly different, may in fact be the closed analog to Earth’s magnetosphere.

### 1.2. Stable Auroral Red Arcs

In his review of ground-based observations of SAR arcs, Hoch [1973] noted that a few hours after the Earth’s magnetic field has been disturbed by a strong increase in the solar plasma flux, two glowing red zones are often detected, occurring approximately along lines of constant geomagnetic latitude in mid-latitude regions. These glowing zones, which occur simultaneously, one in each hemisphere, are caused by emission from the neutral atomic oxygen atom. He noted further that the arcs are subvisual and are detected only at night with photometric and photographic equipment. Based on the spatial occurrence of SAR arcs approximately along the plasmapause and their temporal relationship with large geomagnetic storms, Hoch suggested the ring current as the energy source and the interaction of the ring current with the plasmasphere as the energy transfer mechanism. Mechanisms suggested by Hoch [1973] included the following:

1. **heat flow**: transfer of kinetic energy by Coulomb collisions
2. **transfer of ring current proton kinetic energy to hydromagnetic waves**, which are damped by the electrons in the SAR arc region
3. **direct influx of energetic electrons into the SAR arc region**

Later measurements from spacecraft confirmed his observations based on global imaging as shown in Figure 1.1 and allowed further research to be done regarding the three possible mechanisms suggested in his review. The most current review of SAR arc formation is by Kozyra et al. [1997], who showed modeling results consistent with the energy source being Coulomb drag energy losses from ring-current O⁺ ions (Figure 1.2). The mechanism for transferring this energy downward along field lines is still not settled. Even though heated electron inflow into a SAR arc was observed by Gurgiolo et al. [1982], the transport mechanism of the electrons from the ring current to the ionosphere is still to be determined. Because of the relative rarity of SAR arcs and their subvisual nature, imaging from orbiting spacecraft with sensitive wave and electric field measurements will be needed for an eventual understanding of this fascinating phenomenon that populates one of the important interfaces between the ionosphere and the magnetosphere.
1.3. PLASMAPHERE DRAINAGE PLUMES

The early 1970s saw the first synoptic satellite measurements of cold plasma in the equatorial region of the inner and middle magnetosphere. Comprehensive studies of the morphology and dynamics of the plasmasphere, which is produced by filling of magnetic flux tubes by ionospheric plasma via diffusive equilibrium, were reviewed by Chappell [1972]. Erosion of the plasmasphere during magnetic storms, a typical bulging of the plasmasphere into the dusk hemisphere, and detached blobs of plasma in the afternoon sector were some of the prominent features discovered in the equatorial region by the OGO-5 spacecraft. During the same time, models of the response of the plasmapause to geomagnetic activity as reflected by changes in the convection electric field in a dipole magnetic field were described by Grebowsky [1970] and Chen and Wolf [1972]. Examples of the results are shown in Figures 1.3 and 1.4. The Chen and Wolf model (Figure 1.4) predicts that the plume will wrap around the
Earth if, after a period of intensification the convection electric field drops to a lower value and remains there for an extended period of time. Chen and Wolf [1972] referred to this predicted evolution as the “wrapping up of the plasmasphere.”

The presence of the predicted drainage plumes could not be confirmed until plasmasphere imaging became available with the Imager for Magnetopause-to-Aurora Global Explorer (IMAGE) mission [Burch et al., 2001]. An image of the plasmasphere taken in 30.4 nanometer (nm) extreme ultraviolet (EUV) light is shown in Figure 1.5. This emission is caused by resonant scattering.
of sunlight by helium ions, which comprise about 15% of the plasmasphere density. Also noted in Figure 1.5 are other features that appear at or near this wavelength including the aurora and the helium geocorona. The shoulder feature noted in Figure 1.5, which was discovered by IMAGE-EUV, is caused by northward turnings of the interplanetary magnetic field (IMF) [Goldstein et al., 2002].

Figure 1.6 shows the evolution of the drainage plume as observed by IMAGE-EUV during a period of multiple substorms on June 10, 2001 [Sandel et al., 2003]. As noted in Figure 1.6, the plume wraps around the Earth in the manner predicted by Chen and Wolf [1972] creating a channel, which is often observed in the global images (Figure 1.5).

It is interesting to compare plasmasphere dynamics at Earth with similar phenomena at rotation-dominated planets such as Jupiter and Saturn. Saturn is roughly ten times as large as Earth and rotates more than twice as fast (10.7-hour rotation period). It has a spin-aligned dipole magnetic field that is much weaker than Jupiter’s but nevertheless about 580 times stronger than Earth’s. Except for a magnetotail, Saturn’s magnetosphere is essentially a plasmasphere but with internal plasma sources (predominantly Enceladus) and ubiquitous interchange instabilities [Burch et al., 2005, 2007; Hill et al., 2005]. An example of interchange events observed within the E ring of Saturn is shown in Figure 1.7. Colder high-density plasma is replaced by much hotter but lower density plasma from the outer magnetosphere. This process is important at Saturn because of the planet’s rapid rotation with centrifugal force taking the place of gravity in the closely related Rayleigh-Taylor instability on Earth.

### 1.4. RING CURRENT DECAY

As the cause of global magnetic disturbances during geomagnetic storms, the ring current is one of the most powerful of magnetospheric phenomena, involving ions...
with energies of 10s of kiloelectron-volt (keV). Nevertheless, the ring current is strongly mediated and eventually lost by interactions with the upper atmosphere and ionosphere. Resonant interactions with whistler-mode waves were shown to be important for the precipitation of ring current ions, particularly near the plasmapause, where the ring current and plasmasphere overlap [Williams and Lyons, 1974]. On a global basis, however, charge exchange with exospheric hydrogen atoms and Coulomb collisions within the loss cone of the ring-current ions have been shown perhaps to be more important.

As shown in Figure 1.8, recent comprehensive models of the loss of ring-current ions due to charge exchange and Coulomb collisions have produced results that are consistent with both in situ measurements and imaging of ring-current ions [Fok et al., 2010]. Nevertheless, there still is strong evidence for the importance of wave-particle interactions as an ion precipitation agent in regions of overlap between the ring current and plasmapause. One of these regions is associated with the detached proton auroral arcs that were sometimes observed by IMAGE in the afternoon sector as shown in Figure 1.9 from Burch.
et al. [2002]. Spasojevic et al. [2005] investigated a number of the detached proton auroras and found that they were spatially associated with plasmasphere drainage plumes. In the event shown in Figure 1.10, measurements from the Polar spacecraft, which was located in a nearby region, showed the presence of intense electromagnetic ion cyclotron waves of the type that are predicted to grow in regions of enhanced cold plasma density.

Figure 1.9 Ultraviolet auroral image mapped to invariant latitude and magnetic local time. A detached arc is centered on 15:00 MLT. Selected from Figure 6 of Burch et al. [2002].

Figure 1.10 EUV plasmapause locations (black dots) are shown along with the mapped proton precipitation region (open squares) from ultraviolet images like the one shown in Figure 1.4. The diamonds show the track of the Polar spacecraft over which strong electromagnetic ion cyclotron waves were observed [Spasojevic et al., 2005].

Sounding-rocket measurements of nearly monoenergetic keV electrons focused attention on an electrostatic acceleration mechanism in the topside ionosphere [McIlwain, 1960]. Further sounding-rocket measurement showed that the spectrum extended to low energies of a few tens of eV [Westerlund, 1968]. These low-energy electron measurements began to cast doubt on the electrostatic acceleration mechanism because if all of the auroral electrons originated in the magnetosphere they should all arrive at the beam energy, and this doubt persisted until 1974. By that time orbiting satellites had shown the monoenergetic beams to have a characteristic inverted-V shape in energy and latitude [Frank and Ackerson, 1971]. Using energy spectra from the Frank and Ackerson publication, Evans [1974] produced a model of the acceleration of auroral electrons with a field-aligned electrostatic potential drop and the interaction of the electrons with the upper atmosphere. Elegant in its simplicity, this model was able to fit the observed electron energy spectrum with the low energy electrons being auroral backscattered and secondary electrons, which are trapped between a magnetic mirror point in the atmosphere and the electrostatic potential at high altitudes (see Fig. 1.11). “The possibility
that upward-going backscattered and secondary electrons, produced by a primary beam incident upon the atmosphere, would reappear as precipitating electrons was not appreciated” [Evans, 1974].

Further measurements of auroral particles by orbiting spacecraft showed that not all the electrons appear in inverted-V structures. In some regions of the auroral oval often, but not exclusively, near the polar-cap boundary, field-aligned and counterstreaming electrons, with broad energy distributions (<10 eV up to a few keV) are observed as shown in Figure 1.12. The broad energy range and bi-directionality suggest stochastic acceleration by Alfvénic parallel E-fields [Chaston et al. 2003a, 2003b], but other observations indicate that resonant Landau acceleration by inertial Alfvén waves propagating downward from high altitudes is also at play (Wygant et al. 2002). Field-aligned currents in the downward-current acceleration region are carried by upflowing superthermal electrons.
One of the great surprises in 1970s magnetospheric physics was the discovery of precipitating keV-range oxygen ions by Shelley et al. [1972], but what comes down also must have gone up. This was confirmed by Shelley et al. [1976], who discovered copious amounts of keV-range oxygen and hydrogen ions flowing out of the ionosphere along magnetic field lines. This result was followed by the discovery by Sharp et al. [1977] of ion conics, particles moving out of the ionosphere at pitch angles of 130° to 140°, which were interpreted to have been accelerated in a direction normal to the magnetic field at a lower altitude with the magnetic mirror force and magnetic moment conservation accounting for the “folding up” of the distribution toward the magnetic field direction. Although this interpretation is probably correct for some of the ion conics, it has been difficult to find the presumed source region where the pitch angles would be near 90°. Also the observation of conics over a wide range of altitudes shows similar conic angles, suggesting that the acceleration is not limited to a narrow altitude range but instead occurs all along magnetic field lines.

Together, the field-aligned energetic ions (ion beams) and ion conics add up to a massive outflow of particles into the magnetosphere. While hydrogen cannot be used as a tracer of the solar wind and ionospheric sources, O⁺ surely can. The fact the O⁺ ions are observed throughout the magnetosphere over a wide range of energies leads to the conclusion that the ionosphere is a very important source of plasma to the magnetosphere [e.g., Chappell et al., 1987].

Prior to 1972 a common assumption in magnetospheric physics was that all of the energetic plasma came from the solar wind and that it was only the cold plasma of the plasmasphere that originated in the ionosphere. This notion was so strong that for many years no keV range mass spectrometers were ever designed into
magnetospheric missions. It is interesting that the breakthroughs on the ionospheric keV ion source were made from rather obscure low-altitude defense department satellites rather than from mainstream magnetospheric physics missions. The lesson is always to be trying something new and different even if it is against the conventional wisdom, but it is usually not possible on expensive missions that are designed by committees and must guarantee results.

The science of ion beams and conics has progressed very rapidly and by now is a science discipline of its own. Recent data and modeling results by Lynch et al. [2002] show how various types of conical ion distributions occur in various auroral conditions, illustrating the complexity of this field of study and the many unsolved problems that still exist. Figure 1.13 shows four different H+ distribution functions, which all fall into the general description of ion conics. Only Figure 1.13(b) fits the original concept of conical distributions while the others contain mixtures of parallel and perpendicular acceleration and wave heating.

The global nature of ion outflow is illustrated in Figure 1.14, which shows outflowing <1 keV O+ ions (right panel) along with electron precipitation power (center panel) and downward Alfvénic Poynting flux (left panel). These three parameters are generally correlated, especially in the pre-midnight region of ionospheric flow reversal (the Harang discontinuity), indicating that ion energization is closely coupled to convection especially in fast flow channels.

1.7. AURORAL KILOMETRIC RADIATION

That the Earth is a powerful radio source was surprisingly unknown prior to the observations made from outside the magnetosphere by Gurnett [1974], Kaiser and Stone [1975], and Alexander and Kaiser [1976]. The generation and beaming of AKR was explained by a comprehensive theory published by Wu and Lee [1979]. The cyclotron maser theory of Wu and Lee has been successful in predicting the X-mode radiation, the beaming of waves upward from an auroral plasma density cavity, and the polarization of the waves, which is opposite in the northern and southern hemispheres. The predictions of the theory have been confirmed in the case of Saturn kilometric radiation (SKR) as well as for Jupiter’s decametric radiation, which by virtue of its much higher frequency was discovered through ground-based observations in 1955 [Burke and Franklin, 1955].

The cyclotron maser theory is based on electron velocity-space gradients that occur in the auroral regions. These gradients were originally identified with the well-known loss cone, which is caused by atmospheric absorption of energetic particles but has since been associated with electron “hole” distributions that develop in the downgoing auroral electron population. An example of the simultaneous occurrence of both of these gradient regions is shown in Figure 1.15 from Menietti et al. [1993]. A schematic representation of the cyclotron maser interaction is shown in Figure 1.16, in which a flux tube depleted of plasma by a field-aligned electric field forms a resonant cavity for Doppler resonance of electromagnetic waves with auroral electrons. The density gradients that occur at the ionosphere and at the walls of the cavity both trap the waves and allow them to escape upward. The electron interaction explains the right- and left-hand polarizations that occur in the two hemispheres of Earth, Saturn, or Jupiter.

1.8. SATURN MAGNETOSPHERIC PERIODICITY

Although SKR and its periodicity of about 10.7 hours was observed by the Pioneer and Voyager spacecraft, it was not until the Cassini orbital mission that the evolution of the periodicity and its appearance in all plasma,
energetic particle, and magnetic field measurements began to be observed. Figure 1.17 shows the evolution of the periodicity of the northern and southern hemisphere components of SKR for the first six years after Cassini's orbital insertion. Prior to the Cassini mission the SKR periodicity was taken as the best measurement of Saturn's rotation rate. However, the discovery of two periodicities, both of which are slower than measurements based on the gravity field [Anderson and Schubert, 2007] and cloud motions [Read et al., 2009], raised new questions. The rotation deficit was associated with slippage between the ionosphere and magnetosphere, which varied in a seasonal manner since initially the slower rotation occurred in the summer (southern) hemisphere and the more rapid rotation in the northern (winter) hemisphere. Numerous ideas and models have been proposed for the periodicity of SKR and the many other plasma and field phenomena observed in the Saturn magnetosphere (Figure 1.18). Some of the ideas have involved magnetospheric phenomena such as magnetic cams [Southwood and Kivelson, 2007], plasma cams [Burch et al., 2009], plasma tongues [Goldreich and Farmer, 2007], or interchange modes [Gurnett et al., 2007], while others have involved ionospheric sources such as long-lived vorticities [Jia and Kivelson, 2012]. The search has been complicated by the fact that the clear hemispheric separation between the two periodicity modes has not been re-established since the apparent crossover in 2010.

Although the cause of the SKR and magnetospheric periodicity at Saturn remains a mystery, it is nonetheless one of the most dramatic manifestations of M-I coupling in the solar system. Future missions to Saturn that are designed to investigate these specific phenomena, most likely with multiple spacecraft and enhanced atmospheric and magnetospheric imaging, will be needed.

1.9. FUTURE CAPABILITIES: MODELING AND NEW MISSIONS

Progress in understanding the geospace environment is dependent not only on new measurements but on accurate modeling, which only recently began to include M-I coupling phenomena such as ion outflow. Figure 1.19 illustrates the results of a model of magnetospheric sawtooth oscillations both with and without ionospheric outflow [Brambles et al., 2013]. Inclusion of the outflow is clearly necessary for the sawtooth events to appear in the model.

Much progress has been made in the assimilation of data into ionospheric models [e.g., Schunk et al., 2004] but much less so in magnetospheric models. With the dramatic advances in the accuracy of magnetospheric models, the role of data has to evolve from something to be explained, to targeted inputs, to models that establish boundary conditions and end states.
Figure 1.17 The top panel shows the latitude of Cassini, and the bottom panel shows a frequency-time spectrogram of the normalized peak-to-peak power of the SKR modulation as a function of UT and the rotational modulation rate, $\omega$. The white dashed horizontal lines are the internal rotation rates of Saturn inferred by Anderson and Schubert (2007) and Read et al. (2009). Note that the rotational modulation rates are all less than the inferred internal rotation rate, which implies that the magnetospheric is slipping with respect to the rotation of Saturn’s interior [Gurnett et al., 2010].

Figure 1.18 The fluxes and number densities of low-energy electrons (first and second plots) and magnetic field (third through sixth plots) measured by the Cassini spacecraft during 17 July 2006 to 21 July 2006 [Khurana et al., 2009].
The first Jupiter polar orbiter mission, Juno, is set to arrive at the planet on July 4, 2016. With auroral imaging and a full set of plasma, energetic particle, and wave and magnetic field measurements, Juno is equipped to investigate M-I coupling in the Jovian environment. In the case of Earth, the Magnetospheric Multiscale (MMS) mission is now performing a detailed experiment on magnetic reconnection in the outer magnetosphere. While not specifically designed to investigate M-I coupling, MMS will nevertheless obtain the first detailed measurement of the process that transmits solar-wind energy into the magnetosphere and ionosphere.

Future proposed magnetospheric and ionospheric missions generally involve clusters or constellations of spacecraft equipped to map out the flow of mass, energy, and momentum throughout geospace with both imaging and in situ measurements. Because of the sheer size and dynamic behavior of geospace, such missions will have to involve intrinsic modeling components because it will not be possible to measure everything on a closely spaced grid but instead will require a computational web to connect many of the measurement points.

While the challenge for M-I coupling at the Earth is to obtain global and dynamic coverage of geospace, full understanding of M-I coupling requires further exploration of its occurrence in other planetary environments. Up until now magnetospheric and ionospheric measurements have been carried upon planetary missions, but the time is coming when the traditional boundaries of heliophysics need to be expanded toward their natural limits.

1.10. CONCLUSIONS

As summarized in this review, the early 1970s clearly was a watershed period for M-I coupling. The many new measurements that were made over only about half a decade resulted in the realization of the importance of M-I interactions. The discussions held at the 1974 Yosemite Conference on Magnetosphere-Ionosphere Coupling led eventually to the implementation of a dedicated space mission, Dynamics Explorer, which resulted in vast new knowledge of how the polar magnetosphere and ionosphere behave as one coupled system.

But now there is a crossroads, with measurement requirements expanding while resources are mostly stagnant. More now than before, the relevance and excitement of M-I coupling and other important heliophysics phenomena need to be demonstrated in the context of a mature science rather than a new science, which is clearly more difficult. There is no easy answer, but the lessons from the past often illuminate paths to the future, and the lesson of the birth of M-I coupling research four decades ago is that a large and diverse scientific community working together while appreciating each other’s science can lead to great success.

REFERENCES


