RADIATION BELTS AND SUBSTORM PARTICLE INJECTIONS

Xinlin Li
Department of Aerospace Engineering Sciences and LASP, University of Colorado, Boulder, CO 80303-7814, USA

ABSTRACT
The dynamics of radiation belt particles in the Earth’s magnetosphere is briefly reviewed, with a focus on the radiation belt electrons. Diurnal variations observed by satellites at geosynchronous orbit for electrons with different energies are presented and the acceleration mechanisms for radiation belt electrons are discussed.

1. INTRODUCTION
The acceleration of charged particles is of great cosmic significance. Much of the view that we have of the distant universe comes from energetic particles, mostly through their local interactions that produce gamma-ray, X-ray, and radio emission. For example, before spacecraft visited Jupiter, it was anticipated that Jupiter has a magnetosphere with many energetic electrons in it because of its decimetric radio emissions. In fact, the radio emission was used to estimate the radiation environment for the spacecraft design (Blake, private comm., 1998). We also know that the X-rays coming from Pulsars are most likely due to very energetic electrons in very strong magnetic fields. Earth’s magnetosphere is most accessible for us to study charged particle dynamics.

1.1. Motion of Charged Particles
In the Earth’s magnetosphere, a charged particle conducts three distinctive motions, gyration around the magnetic field line, bounce motion along the magnetic field line, and drift motion across magnetic field lines. Ions drift westward and electrons drift eastward. More energetic particles drift faster. The time scales of these three motions are well separated. For example, for an 1 MeV electron with an equatorial pitch angle of 60° at r=6Re, the time periods for the gyromotion, bounce motion, and drift motion are on the order of $10^{-1}$, $10^0$, and $10^3$ seconds, respectively.

There is an adiabatic invariant associated with each of these motions. As long as the magnetic field does not change significantly over one of these time-scales, the corresponding invariant remains constant. The conservation and violation of these invariants are central to understanding the particle’s motion (Roederer, 1970).

If certain space regions are populated with many of these energetic particles, we call them radiation belts.

Next we introduce the radiation belts that we normally know, and we will focus on an unexpected radiation belt created by a strong interplanetary shock about a decade ago. Then we discuss the relation between radiation belt electrons and substorms.

1.2. Brief Review of Radiation Belts
The first and most important discovery in space science was the discovery of the radiation belt, also called as the Van Allen Belt, in 1958. But the first two satellites were launched by the former Soviet Union in late 1957, which also marked the beginning of the space science era.

Figure 1 shows the location of the radiation belts in the meridian cross section. The inner belt consists mostly of energetic protons (> 10 MeV) and very stable, varying on a time scale of a solar cycle; the outer outer belt both...
electrons and protons: hundreds keV to many MeV, and is not stable at all. It varies with solar cycle, semiannual (or seasonal) and solar-rotation time scale, and in particular with magnetospheric storms (Li et al., 2001b). There is a gap between the inner belt and outer belt, called the slot region, where fewer particles can reside normally but it can be filled up during active times. The lighter narrow stripe represents the trapped Anomalous Cosmic Ray particles, the intensity of which varies with solar cycle.

The white lines represent the Earth’s magnetic field lines, approximated as a dipole field which can be well described by the dimensionless parameter L number, which is equal to the central distance in unit of Earth radii at the equator.

After the discovery of the radiation belt, an impressive amount of research was carried out during the 1960s. The source for the radiation belt was believed to be from Galactic Cosmic Ray Neutron Decay. This is true for the inner proton belt, but not true for the electrons (Walt, 1996).

Now we briefly discuss the Anomalous Cosmic Ray belt, based on Mewaldt et al. [1994].

Since the beginning of the space age, it had been known that two main sources of energetic particles pervade interplanetary space: Galactic Cosmic Rays (GCR), originating from supernova explosions, which occur approximately once every 50 years in our galaxy, and Solar Energetic Particles (SEP), usually associated with solar flares or coronal mass ejections (CME).

The Anomalous Cosmic Rays (ACRs) belong to neither of them by definition. In 1973, the anomalous excesses of several elements in low-energy cosmic rays led to the discovery of the so called ACRs. For example, Oxygen exceeds Carbon 30 times in abundance in the low energy range (tens of MeV). Helium is more abundant than Hydrogen. In contrast, in SEP and GCR, C and O are comparable and H is typically 10 times more than He.

Soon after this discovery, Fisk et al. [1974] proposed that they represent a sample from interstellar particles. The ACRs originate from interstellar neutral materials, mostly He, N, O, etc, which can easily penetrate into our heliosphere. Ionized materials would be deflected at the bow shock. These elements are relatively difficult to ionize, with the first ionization potential (FIP) $\geq 13.6$ eV. But when they get to close to the Sun, they are ionized by Solar UV radiation or by charge exchange with solar wind particles. After losing a single electron, the ion is picked up by the solar wind and heads toward the solar wind termination shock, where it can be further accelerated. Some of the accelerated ions can come toward the Earth. A singly charged ion can be further stripped of its electrons when it happens to skim the Earth’s atmosphere. Since the gyroradius is inversely proportional to the number of the charge, its gyroradius is reduced many times and the ion can become trapped by the Earth’s magnetospheric field. This scenario was predicted far in advance by Blake and Friesen (1977). The first evidence for trapped ACRs in the magnetosphere was provided by a team of Russian and US scientists using observations from a series of COSMOS satellites from 1985-1988. SAMPEX, a NASA satellite launched in July 1992 to a polar orbit with an altitude of 600 km, pin pointed the location of the narrow belt of ACRs, within the Inner Van Allen belt (Cummings et al., 1993).

The ACRs, both in interplanetary space and trapped in the magnetosphere, are strongly modulated by the solar activity. Their anti-correlation with solar activity is quite evident (Selesnick, 2001). ACRs have to work their way close to the center of our solar system, just like salmon swim against a current. As for the absolute flux, the ACR belt is much weaker compared with the inner belt.

We have discussed the radiation belts that normally exist. Now we will focus on the new radiation belt created by a strong interplanetary shock and recent enhancements in our understanding of the outer radiation belt as a whole.

1.3. Sudden Formation of New Radiation Belts

By the end of the 1970s, the radiation belt was regarded as a well understood subject by many people, and ra-
radiation belt research had in fact moved out from planet Earth to planet Jupiter and beyond (Roederer, 1996). The driving aspiration was to push forward magnetospheric studies to the other planets. Indeed, Pioneer and Voyager missions revealed or confirmed that all of the outer giant planets-Jupiter, Saturn, Uranus, Neptune-have well-developed magnetospheres, and all of them have trapped radiation belts. The intensity of radiation belt particles in Jovian magnetosphere is much stronger, while the other three planets have comparable radiation belts to the Earth. Radiation belt research had been dormant until a wake-up call finally came on March 24, 1991.

Figure 2 is a radiation dose rate sorted by L-shell and plotted vs time from the CRRES satellite, which was in geo-transfer orbit. The inner belt is relatively stable, outer belt varies with time, usually on a scale of hours and days. But on March 24, 1991, a new belt with a much greater intensity formed in the slot region in about one minute, totally beyond expectation. The rest mass of an electron is only 0.511 MeV. A 13 MeV electron is moving in gyration at 99.9% of the speed of light. Where did these >13 MeV electrons come from? Several experimentalists who were involved with the data analysis literally lost several nights’ sleep (Blake, Korth, and Vampola, private comm., 1994).

What happened on that day and before could be traced back to the Sun. There was a major CME one day before, which resulted in a very fast interplanetary shock, with a speed of >1400 km/s, coming toward the Earth. In front of the shock, there were energized solar energetic particles, mostly protons, due to the 1st order Fermi acceleration (Blake et al., 1992).

On March 24, 1991, as CRRES moved toward its perigee, it witnessed the prompt enhancement by several orders of magnitude of multi-MeV electrons. The left column of Figure 3 shows the actual measurement of electrons in different energy channels, the electric field and magnetic field. The first peak is the initial enhancement. The subsequent peaks are due to the same electrons coming back to the satellite several times before the satellite left the region. These are called drift echoes. The magnetic field shows mainly one pulse, corresponding to compression and relaxation. The background field has been subtracted, so only the variation is shown. The electric field shows mainly two pulses. The first is associated with compression and the second one is associated with relaxation. CRRES was at post midnight; the fields would be much stronger at dayside. Questions: Where are the electrons coming from? How are they energized to such high energy and so promptly.

The right column shows the model fields and simulated electron drift echoes in the same energy ranges. The simulation has incorporated the satellite’s motion and detectors’ actual responses (Li et al., 1993).

In short, the interplanetary shock severely compressed the magnetosphere and resulted in a huge inductive electric field, which energized some pre-existing electrons (1-2 MeV) at larger L and brought them into the lower L (=2.5) region, or the slot region. The whole process took just one minute, which is completely out of the picture of classical radial diffusion theory.

One point needs to be emphasized here. The upper panel of the left column of Figure 3 shows the count rate of the electrons and the electrons have about the same drift period, which suggest that these electrons have almost the same energy. Producing mono-energetic particle flux cannot be achieved by Fermi acceleration, and cannot be achieved by betatron acceleration. It has to be associated with some resonance. In this case, it is a drift resonance. The idea is that when a magnetic field is compressed, an inductive electric field will be generated, which also propagates through the magnetosphere. It selectively accelerates some pre-existing particles whose drift speed happens to be comparable to the wave propagation. We call this drift resonance, which is analogous to wave surfing.

Energetic protons were also injected to low L region. The proton sudden enhancements and subsequent drift echoes were simulated and reproduced as well, based on the
same field model, but using a different source population. They are mostly the solar energetic particles which temporarily filled up magnetosphere at larger L-shells (Hudson et al., 1995). It should be noted that SEP observed in the magnetosphere are usually transient particles. These energetic ions have large gyro-radii and cannot be trapped by the magnetosphere. However, if there is a strong interplanetary shock following them to the Earth, some of them will be pushed into much smaller L-shells and be trapped.

The comprehensive measurements and our successful simulation of this event have revolutionized our understanding of the formation of radiation belts. This test particle simulation was based on a simple analytical model. Hudson et al. (1997) started to trace the particles on the fields generated from a global MHD simulation. They reproduced most of the observations, revealed other details, and consolidated this shock-associated acceleration mechanism. In these simulations, the sources of the new radiation belt electrons are 1-2 MeV electrons in the larger L region (L>7), which were assumed to be pre-existing before the arrival of the shock, based on limited observations. The question is, where do these 1-2 MeV electrons come from in the first place?

1.4. Source of MeV Electrons in the Magnetosphere
The intensity of 1-2 MeV electrons usually peaks around L=4-5 because intensity is usually plotted as differential flux. A more physically meaningful parameter is phase space density. If the phase space density, sorted by first and second adiabatic invariants, is plotted, it usually increases toward larger L, indicating that the source of these electrons is further out. Electrons can be energized by inward radial transport through the violation of the third adiabatic invariant. The energy that can be gained by radial transport, whether in the form of radial diffusion or fast injections, is limited by the difference of magnetic field magnitudes within the region of radial transport. Thus radial transport as an energization mechanism normally requires a source population of electrons that is already hot.

Although the average temperature of solar wind electrons is relatively low (Te~10 eV), the solar wind also contains a much hotter ‘halo’ and an even hotter ‘superhalo’ population of electrons (Lin, 1998). This superhalo population varies with solar and solar wind activity and has a temperature (Te~5 keV) whose high energy tail is conceivably sufficient to produce some of the electrons in the radiation belt if energized within the magnetosphere solely by radial transport. However, Li et al. (1997a) examined this hot superhalo population and showed that it did not have sufficient phase space density to supply the radiation belts without additional heating processes within the magnetosphere. This result implies that the source population for the electron radiation belts must be created by some heating processes within the magnetosphere. For a given value of the first and second adiabatic invariants, the space density usually increases with increasing L for L=3-6.6 and beyond (Selesnick and Blake, 1997). Thus there should be a region outside of geosynchronous orbit, and within the magnetosphere, where the phase space density at constant first and second adiabatic invariants peaks. Indeed, a study of several Wind perigee passes in conjunction with POLAR measurements suggests that the phase space density for given first adiabatic invariant continues to increase toward larger radial distances (~11-14 Re) and precipitously decreases once Wind goes out of the magnetosphere (Li et al., 1997b).

1.5. Solar wind correlation with MeV electrons in the magnetosphere
While the solar wind does not provide the direct source for the MeV electrons in the magnetosphere, it strongly modulates or even controls the variations of the MeV electrons. Paulikas and Blake (1979) showed that there is a good correlation between the solar wind velocity and the MeV electron flux at geosynchronous orbit. Since geomagnetic activity and substorms are known to be controlled more strongly by the polarity of the interplanetary magnetic field, the better correlation of the radiation belt electrons with solar wind velocity had been mysterious.

Recently we have reexamined this issue using several years of almost continuous solar wind data from the Wind and ACE satellites and confirmed the results of Paulikas and Blake (1979). Larger solar wind velocities may also drive fluctuations at the magnetopause and produce ULF waves within the magnetosphere (Engebretson et al., 1998; Mathie and Mann, 2000).

An obvious refinement is to see if we can do an even better job of predicting the electron flux on the basis of measured solar wind parameters. Figure 4 demonstrates our current results at geosynchronous orbit (Li et al., 2001a). It is a comparison of five and a half years of daily averages of the MeV electron flux measured at geosynchronous orbit with our prediction, based solely on measurements of the solar wind. Both the shorter time scale and the longer seasonal effects, such as the overall reduction in the electron fluxes in the middle of 1996, are reproduced. These years include the last part of the previous solar cycle, the solar minimum, and the current solar cycle close to the solar maximum.

Our MeV electron prediction model (Li et al., 2001a) for geosynchronous orbit is based on the standard radial diffusion equation (Schulz and Lanzerotti, 1974). The inner and outer boundary are set at L = 4.5 and L = 11, though these values can be adjusted. Since we have so far used the model to determine the MeV electron flux at geosynchronous orbit, the exact values of the inner and outer boundary do not significantly affect our results. The dif-
Figure 4. (Extended version of Figure 1 of Li et al. [2001a]) A comparison of five and half years of daily averages of the MeV electron flux measured at geosynchronous orbit with predicted results based solely on measurements of the solar wind. The red line shows the observed electron fluxes and green line shows predicted results. Horizontal axis shows the day of the year.

The electron data are daily averages of the 0.7-1.8 MeV electron flux from Los Alamos National Laboratory (LANL) detectors at geosynchronous orbit. The model results are based solely on solar wind density, velocity, and magnetic field data from the Wind and ACE spacecrafts. Using this method, we have achieved a prediction efficiency of 0.81 and a linear correlation of 0.90 for the two years 1995 and 1996 for the logarithm of average daily flux of electrons with energies of 0.7-1.8 MeV (Li et al., 2001a). The same model has been used to predict higher energy electrons (1.8-3.5 MeV and 3.5-6.0 MeV) at geosynchronous orbit with even greater success in terms of the prediction efficiency.

2. RADIATION BELT AND SUBSTORM INJECTIONS

2.1. Diurnal Variations

Energetic electrons measured at geosynchronous orbit show local time dependence, as demonstrated in Figure 5. Since the magnetosphere is compressed at dayside, a geosynchronous orbit satellite is at different L-shells at different local time—at smallest L-shell at dayside and largest L-shell at nightside. For lower energy electrons, 50-300 keV, their fluxes peak at night side as a result of substorm injections. However, there is an upper energy cut-off (≈ 300 keV) of injected electrons associated with substorms, which has been noted before (Baker et al., 1989; Li et al., 1996). The fluxes of higher energy electrons, >500 keV, peak at dayside because the fluxes of the higher energy electrons peak at smaller L-shells.

Figure 5 shows a superposed epoch analysis of the local time dependence of electrons of different energies measured at geosynchronous orbit for three selected periods. The minimum Dst index was -56, -43, and -117 nT for May 29, Sept. 7, and Nov. 26, respectively (WDC-C2, Kyoto, Japan). Figure 6 further demonstrates that lower energy electrons peak near midnight and higher energy electrons peak near noon during different geomagnetic activities.
Figure 6. Superposed epoch plot of electrons at different energies measured at geosynchronous orbit for three periods of interest. The vertical dotted lines indicate local midnight.

2.2. Substorm injections and further energization

Radiation belt electrons are formed by accelerating less energetic electrons. There are two possible sources of less energetic electrons: One source is electrons at larger L that can be energized by being radially transported inward. This is usually called ‘radial diffusion’ and is usually considered to be the main process. Another source is less energetic electrons at the same location that can be energized by wave-particle interactions. Both possible sources usually have a substantially larger phase space density than the radiation belt electrons and thus either of them could be a source of radiation belt electrons.

Baker et al. [1979] showed that only 20% of substorm injections include an increase of electrons with energies greater than ~300 keV. However, these substorm injected electrons (some of them do not reach geosynchronous orbit) may well be the ‘seed population’, subsequent radial diffusion can further energize them to higher energies (e.g., Baker et al., 1998). By analyzing plasma wave and particle data from CRRES satellite for three case studies, Meredith et al. [2002] suggest that the gradual acceleration of electrons to radiation belt energies during geomagnetic storms can be effective only when there are prolonged substorm activities during the recovery phase of the storm. They argue the prolonged substorm activity provides sustained VLF waves which in turn accelerate some of the substorm injected electrons to higher energies. On the other hand, prolonged substorm activities may also be correlated with enhanced ULF waves, which enhance the radial transport of the electrons (Hudson et al., 2000). The relative effectiveness of these acceleration mechanisms has not been quantified.

ACKNOWLEDGMENTS

The author gratefully acknowledges M. Temerin, D. N. Baker, J. B. Blake, M. K. Hudson, I. Roth, J. Wygant, G. D. Reeves, S. G. Kanekal, and D. Larson for their contributions to the published work (Li et al., 1993, 1997a, 2001a). This work was supported by NSF/ATM-9901085 and NASA/NAG5-9421.

REFERENCES


