Lunar surface electric potential changes associated with traversals through the Earth’s foreshock

Michael R. Collier, a, f, H. Kent Hills, a, b, Timothy J. Stubbs, a, c, f, Jasper S. Halekas, d, f, Gregory T. Delory, d, f, Jared Espley, d, William M. Farrell, a, f, John W. Freeman, e, Richard Vondrak, a, f

a NASA/Goddard Space Flight Center, Greenbelt, MD, USA
b Adept Systems, Inc., Lanham, MD, USA
c Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County, USA
d Space Sciences Laboratory, University of California, Berkeley, CA, USA
e Department of Physics and Astronomy, Rice University, Houston, TX 77005, USA
f NASA’s Lunar Science Institute, NASA Ames Research Center, Moffett Field, CA 94035, USA

1. Introduction

The electric potential of the lunar surface has been an active area of investigation from the first realization prior to the Apollo era that the interaction of the Moon with the space environment would cause it to become electrically charged (Singer and Walker, 1962). For example, early work qualitatively described the modern picture of electrostatic dust levitation and transport in the context of lunar erosion (Gold, 1955). Since the Apollo era, the application of formalism derived for spacecraft charging has allowed a quantitative assessment of the lunar surface potential, e.g., Manlo (1973).

In principle, the processes causing lunar surface charging are simple and fundamentally the same as spacecraft charging: When the lunar surface is in sunlight, the dominant process is the photoemission of electrons—solar UV and soft X-ray photons eject electrons off the surface, leaving the surface with a compensating positive potential. When the lunar surface is in shadow in the presence of a plasma, the solar wind electrons, which have a much larger thermal speed than the protons, produce a greater flux onto the lunar surface, resulting in excess electrons and a corresponding negative potential.

This surface charging process produces a variety of phenomena, some of which may impact human exploration. For example, the Lunar Ejecta and Meteorites (LEAM) experiment was placed on the Moon during the Apollo 17 mission in order to directly measure the impact of cosmic dust on the lunar surface (Berg et al., 1976; Colwell et al., 2007). However, the bulk of the events registered by this experiment were not hypervelocity (~10 km/s) impacts by cosmic dust, but were instead due to highly charged dust moving at ~100 m/s. The dust impacts were observed to peak around the terminator regions where the potential transitions from positive to negative (Farrell et al., 2007). The dust is believed to be accelerated by the complex electric field configuration in this region. Because NASA and other countries have indicated interest in exploring the lunar poles, and the polar regions are always in the vicinity of the terminator, this dusty sleet must be well-understood prior to extensive exploration activities of this type.

Lunar electric fields could also pose an electrostatic discharge (ESD) hazard. This danger is particularly acute on the nightside. The nightside charge has no place to dissipate because the plasma environment is so tenuous and the surface is an effective insulator raising the issue of tribocharging and equilibration timescales (Farrell et al., 2008a,b). Given the above interpretation of the LEAM observations, electrostatic transport of dust by lunar electric fields is likely. Because it can obscure vision and affects instrumentation, dust has become a recognized lunar hazard although concerns about dust date back to the Apollo-era: astronauts who walked on the Moon experienced many problems arising from lunar dust. It adhered to clothing and equipment, a health and equipment hazard given its abrasive nature, and it caused respiratory problems (Stubb et al., 2007a; Perk et al., 2008; Liu et al., 2008). It seems likely that differential charging of the lunar surface can result in strong electric fields that are able to eject charged dust into the exosphere (De and Criswell, 1977; Criswell and De, 1977; Borisov and Mall, 2006; Wang et al., 2007) with observable and potentially observable consequences (McCoy and Criswell, 1974; McCoy, 1976; Collier and Stubb, 2009; Stubb et al., 2010).

In this paper, we will describe a study in which we show that the lunar surface potential increases as the Moon traverses the dawn flank of the Earth’s bow shock. This provides an opportunity to understand the lunar surface potential via the use of a very well-defined charge at the bow shock. Analogous to an active experiment in which a laboratory investigator bombards regolith samples and experiment with a top-hat electrostatic analyzer designed to perform a similar experiment, given the well-defined surface potential hazard given its potentially observable consequences (McCoy and Criswell, 1974; McCoy, 1976; Collier and Stubb, 2009; Stubb et al., 2010).

In this paper, we will describe a study in which we show that the lunar surface potential increases as the Moon traverses the dawn flank of the Earth’s bow shock. This provides an opportunity to understand the lunar surface potential via the use of a very well-defined charge at the bow shock. Analogous to an active experiment in which a laboratory investigator bombards regolith samples and experiment with a top-hat electrostatic analyzer designed to perform a similar experiment, given the well-defined surface potential hazard given its potentially observable consequences (McCoy and Criswell, 1974; McCoy, 1976; Collier and Stubb, 2009; Stubb et al., 2010).

In this paper, we will describe a study in which we show that the lunar surface potential increases as the Moon traverses the dawn flank of the Earth’s bow shock. This provides an opportunity to understand the lunar surface potential via the use of a very well-defined charge at the bow shock. Analogous to an active experiment in which a laboratory investigator bombards regolith samples and experiment with a top-hat electrostatic analyzer designed to perform a similar experiment, given the well-defined surface potential hazard given its potentially observable consequences (McCoy and Criswell, 1974; McCoy, 1976; Collier and Stubb, 2009; Stubb et al., 2010).

2. History of lunar potential theory and observations

Using early observations of the solar ultraviolet spectrum and space plasma environment, theoretical predictions for the electrostatic surface potential on the lunar dayside were ~ +4 V (e.g. Groisman and Blank, 1969; Walbridge, 1973; Knott, 1973). This was followed by predictions by Manka (1973) of how the lunar surface charging would vary on the dayside as a function of solar zenith angle, as well as the effect of the various plasma environments encountered by the Moon. Manka achieved this by using a set of equations developed for spacecraft charging applications (Whipple, 1959, 1965; Fehrle, 1965). On the observational side, the Charged Particle Lunar Environment Experiment (CPLLE) which was part of the Apollo Lunar Surface Experiments Package (ALSEP) on Apollo 14 measured electrons and protons in 18 different energy bands (for each species) between 40 eV and 20 keV (O’Brien and Reasoner, 1971). CPLLE detected a photoelectron layer above the sunlit surface, but no flux with energies above 200 eV. Reasoner and Burke (1972) took this result to indicate that the sunlit surface was positively charged, and perhaps with a potential of 200 V.

The Suprathermal Ion Detector Experiment (SIDE) was a part of the ALSEP that was placed on the lunar surface during each of the Apollo 12, 14 and 15 missions. Apart from those results, SIDE was able to determine the lunar surface potential for numerous events and at many solar zenith angles (Freeman and Ibrahim, 1975; Freeman et al., 1972). The theoretical predictions were roughly consistent with the Apollo measurements and an overall qualitative picture of the nightside lunar surface potential at ~100 V negative and the dayside lunar surface potential at ~10 V positive.

More recently, the application of electron reflectometry has also contributed significantly to our understanding of lunar surface potentials (Halekas et al., 2002, 2007). The electron reflectometry technique was first developed in order to measure lunar crustal magnetic fields (Goldstein, 1971). In fact, the technique was discovered serendipitously when Apollo subsatellite data revealed electrons returning from the surface of the Moon (Anderson et al., 1975). Once researchers realized that electrons were adiabatically reflected from crustal magnetic fields, they quickly developed techniques to exploit this. Although Apollo reflectometry data had limited energy and angular resolution (5 energy channels ranging from 0.5 to 14 keV and variable pitch angle resolution, but often of the order of ~45°), it proved sufficient to map the broad distribution of lunar crustal fields. Similarly, the use of electron reflectometry to determine surface electric fields was discovered serendipitously by Lunar Prospector (LP) scientists. LPs instrument complement included the electron reflectometer (ER), a top-hat electrostatic analyzer that measured full three-dimensional electron distribution functions (from 10 eV to 20 keV) every 80 s. With LP data, there are two independent approaches to determining surface potential: the energy dependence of the loss cone and the energy of the electron beam (Halekas et al., 2002, 2003). LP data therefore allow a consistent determination of potential differences on the negatively charged nightside and discovered that the nightside potential can become highly negative during solar energetic particle events (Halekas et al., 2005, 2007; Stubb et al., 2007b).

The Moon goes through a variety of different plasma regimes including the Earth’s magnetosphere, both the magnetotail lobes and the plasma sheet, and its own wake. The various predictions and experimental evidence concerning the lunar surface potential in these various plasma regions is summarized, along with references, in Table 1.

3. The suprathermal ion detector experiment

The SIDE instrument measures positive ions and employs curved plate analyzers for energy per charge discrimination in both its sensors, the total ion detector (TID) and the mass analyzer (MA) (Benson, 1975). We will only examine in this paper data from TID because we are concentrating on measurements of the surface potential, which do not require a composition measurement. The instrument is equipped with a ground plane electrode, a circular wire grid 65 cm in diameter in contact with the lunar surface, stepped through a cycle of 24 voltages from -27.6 to +27.6 V. The instrument energy channels range from about 10 to 3500 eV. As reported by Reasoner et al. (1971), photographs of the grid show that it contacts the surface at many points implying good electrical contact facilitated by photoelectrons. This allows determination of the electric potential of the lunar surface by examining the energy spectra of thermal ions born with essentially zero energy accelerated into the instrument by the grid/ground plane voltage in the presence of the lunar surface potential. Fig. 1 shows the SIDE configuration in sunlight when the surface potential is positive, photoions are present, and they are accelerated into the instrument. The Apollo 12 SIDE was placed at Ocean of Storms at 3°S, 23 W, the Apollo 14 SIDE was placed at Fra Mauro at 4°S, 17 W, and the Apollo 15 SIDE was placed at Hadley Rille at 26°N, 4°E. The surroundings were relatively flat except perhaps at Hadley Rille although Mt. Hadley, while providing a striking backdrop for photographs, is still a considerable distance away. So, at all three sites the local terrain near the deployed instruments is not expected to produce any significant effects due to local shadowing (e.g. De and Criswell, 1977; Farrell et al., 2007).

The SIDE fields-of-view are about a 6° square solid angle primarily in the ecliptic plane but canted 15° from the local vertical.

This was done because at the time of the instrument design, the landing sites were not known, but were expected to be near the equator and not at widespread longitudes. With the fields-of-view at 15° from the vertical, there is the possibility of pointing that 15° toward the east or toward the west, so that two instruments can make observations in two directions differing by 30° even if deployed at approximately the same longitude. If deployed at different longitudes, then the difference in longitudes also contributes to the difference in possible look directions. The SIDE look directions are shown in Fig. 2 as a function of lunar phase.

4. SIDE events for determining lunar surface electric potential

Frequently, the SIDE/TID data revealed narrow, low energy (~10 eV) ion flux spectra which show a correlation with the ground plane stepper voltage (e.g. Freeman et al., 1973). These types of observations were named “resonance” events because a “resonant”-like effect between the lunar surface potential and the induced solar voltages of SIDE accelerates initially thermal exospheric photoions into the instrument. Fig. 1 shows a schematic of the potentials during a resonance event. Like the other events to be described, resonance events allow a determination of lunar surface potential and are particularly compelling because the inferred lunar surface potential can be “checked” using multiple energy steps of the analyzer and corresponding multiple grid/ground plane voltages, and as such these events are well-constrained. The events analyzed in this study are all resonance events.

Note that the resonance events rely on the presence of lunar photoions that begin at thermal energies, essentially zero for our purpose, and are then accelerated downward by the combined effect of the lunar surface potential and the SIDE grid/ground plane voltage as shown in Fig. 1. Thus, the resonance technique is effective only on the dayside where there is sufficient solar photon flux to ionize a significant abundance of exospheric neutral atoms above SIDE (Stem, 1999) to generate observable fluxes at SIDE. This, incidentally, means that SIDE resonance events are highly complementary to Lunar Prospector electron reflectometer observations which require a negative potential and, as such, are observed primarily on the nightside. Also, this means that the absence of a SIDE resonance event indicating that the lunar surface is at a particular potential does not necessarily mean that the lunar surface is not at that potential—there may simply not be enough photoions in the vicinity to produce a detectable signal.

There are also other types of SIDE events that can be used to determine lunar surface potential, for example the “\textit{vxb}” events (Benson et al., 1973; Fenner et al., 1973) which utilize (a) the energy acquired by an ion upon reaching the charged lunar surface and (b) the flux of ions at the surface due to ionization of neutral atmospheric ions by solar ultraviolet radiation. The effect on the energy spectrum is to shift the entire spectrum toward higher energies by an amount proportional to $q \Phi$ where $q$ is the electron charge and $\Phi$ is the lunar surface potential (Benson, 1977). Because these events rely on the induced solar wind electric field, they depend on the orientation of the interplanetary magnetic field. These events get added acceleration from the \textit{vxb} solar wind electric field, hence their name, and result in a broader spectrum.

Still another type is the non-resonant lunar surface potential event (Lundman et al., 1973). In this type of event, a negative lunar surface potential accelerates positive ions into the instrument.
so that there is no apparent relationship between the surface potential and SIDE’s grid/ground plane and electrostatic analyzer voltages. Most of the activity near the terminators represents this type of event.

Using these types of events, Benson (1977) combined his SIDE data with that from Lindeman et al. (1973) to show that the lunar surface potential becomes progressively more negative with solar zenith angle (Benson, 1977, Figure 3). One should note that these results show that the lunar surface can be in sunlight on the dayside but still have a negative potential within 30° of the terminator in agreement with theoretical predictions (Stubbs et al., 2007b; Manka, 1973).

5. Methodology

Because the SIDE resonance events are quite apparent visually in the count rate plots, and they are easily checked for internal consistency by examining multiple energy channels, we have focused in this study on the Apollo 14 data set which appears to contain more resonance events than the other SIDE data sets. Because there exist only specific accelerating grid/ground plane voltage steps as well as observable energy steps in the SIDE instrument, only certain positive lunar surface potentials are observable. Table 2 shows in the first column the three lowest energy channels with the second through fifth columns corresponding to accelerating grid/ground plane voltages of 27.6, 19.8, 16.2 and 10.2 V, respectively. The difference between the accelerating grid/ground plane voltage and the energy of the channel is the observable lunar surface potential which is shown in the table elements. Note that since the dayside lunar surface potential is typically around a few volts positive, most of the resonance events occur in only two or three combinations of grid/ground plane voltages and energy channels.

Fig. 3 shows an example SIDE resonance event from early January 1972. The counts observed in each 1,208 s accumulation interval for the 7 eV channel at 16.2 V accelerating grid/ground plane voltage and the 17 eV channel at 27.6 V accelerating grid/ground plane voltage are plotted on the left hand y-axis. The counts in each instrument cycle for the 7 eV channel at accelerating grid/ground plane voltages of 19.8 and 27.6 are plotted on the right hand y-axis. Because the energy passband of these channels is about 10% of their energy, the central point of the enhancement in each channel will represent approximately the nominal energy. Although the geometric factors converting SIDE count rates to flux units are energy dependent, here we are concerned not with the absolute magnitude of the flux of each SIDE channel, but rather with when the enhancement in a particular channel indicates that potential has been reached. The arrows point to the approximate center of the resonance events indicating a lunar surface potential of 9.2 V (black), 10.6 V (red), and 12.8 V (blue), where the count rates for those particular combinations of accelerating grid/ground plane voltage and energy channel rise up above the other “non-resonant” channels. The 20.6 V resonance event (7 eV channel at 27.6 V acceleration) is not as convincing because it does not show as clear a rise above the “non-resonant” channels and is labeled with a question mark.

Table 2

<table>
<thead>
<tr>
<th>Energy channel (eV)</th>
<th>27.6</th>
<th>19.8</th>
<th>16.2</th>
<th>10.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>20.6</td>
<td>12.8</td>
<td>9.2</td>
<td>3.2</td>
</tr>
<tr>
<td>17</td>
<td>10.6</td>
<td>2.8</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>30</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

In general, because of the importance of the solar wind in establishing the lunar surface potential (as discussed above) for the analysis presented here, if there were no OMNI-2 solar wind observations during a resonance event, that event was excluded from consideration. However, during one event, event 3 (1972 064 04-05), there was an OMNI gap of only 18 h. Based on scatterplots showing the OMNI-2 solar wind density and velocity versus the Apollo 12 SWS observations (Snyder et al., 1970; Clay et al., 1972, 1975; Clay et al., 1972; Neugebauer et al., 1972) which had correlations of 0.82 and 0.78 respectively, we interpolated across this relatively small gap that occurred right at the time of interest.

Because the correlation scale length perpendicular to the solar wind flow of both magnetic field (e.g. Collier et al., 1998, 2000) and plasma (e.g. Richardson and Paularena, 2004) in the solar wind are in the range of ~40–120R_E, solar wind measurements made near the Earth are generally indicative of solar wind conditions at the Moon. The length scale along the solar wind flow direction is significantly larger (J. King, private communication, 2009), so that the evolution of the plasma parcels in the direction of the solar wind flow from the upstream observation point downstream to the location of the Moon will not be significant along the phase plane, and the only concern will be the distance in the phase plane between the upstream observation and the location of the Moon.

6. SIDE resonance event characteristics

Fig. 4 shows SIDE observations for the Apollo 14 resonance event on 4–5 January 1972 introduced in Fig. 3. The top panel shows the count rate of the SIDE 7 eV channel with the grid/ground plane at a photon accelerating voltage of 16.2 V (red) and 27.6 V (black). Thus, the red trace shows the 9.2 V resonance and the black trace shows the lack of a 20.6 V resonance. So these data indicate a lunar surface potential of about 10 V during this event. The middle panel shows the higher energy SIDE channels: 500 eV, 1 keV, 2 keV, and 3 keV. At these higher energies, the rates will be insensitive to the setting of the grid/ground plane. Note that the plot is logarithmic in count rate and that at about the time the resonance event begins, the count rate of these higher
energy channels drops by a couple orders of magnitude. As will be discussed later, this is interpreted as the Moon passing from the magnetosheath shock-heated plasma where the particle distributions are hot into the high Mach number solar wind. When this happens, the look direction changes correspondingly from into the sheath flow to an angle not aligned with the solar wind flow.

The third panel shows the upstream (dotted) omni data solar wind density (blue) and velocity (red) along with the density and velocity measured by the solar wind spectrometer (SWS) on the surface of the Moon. Figs. 5-10 show the rest of the events listed in Table 3 in the same format as Fig. 4. Note that these are all ∼10 V resonance events.

7. Proximity of the Moon to the bow shock during SIDE events

In this section, we compare the location of the Moon during the SIDE charging events with predicted locations of the bow shock and the magnetopause. We use the Shue et al. (1998) magnetopause model, which is dependent upon solar wind dynamic pressure, $P$, and the interplanetary magnetic field (IMF) $z$-component, $B_z$. The distance to the magnetopause from the Earth center is given by

$$R_{mp} = R_{mp0} \left( \frac{2}{1 + \cos \theta} \right)^{7/6},$$

which is a function of the angle $\theta$ from the apparent solar wind flow vector (which is close to the solar zenith angle but has aberration due to the Earth–Moon orbit around the sun and solar wind flow removed), distance to the subsolar magnetopause in Earth radii

$$R_{mp0} = (10.22 + 1.29 \tanh(0.184(B_z + 8.14))) \cdot 10^{1.65},$$

and

$$\gamma = (0.58 - 0.007 \cdot B_z)(1 + 0.024 \ln(P)).$$

Assuming cylindrical symmetry, this may be converted into $x = p$
coordinates using
\[ x_{mp} = R_{mp} \cos \theta, \]
\[ \mu_{mp} = \sqrt{y_{mp}^2 + z_{mp}^2} = R_{mp} \sin \theta. \]  
(4)

To determine the location of the bow shock, we use the Howe and Binsack (1972) model which has been modified in order to allow it to vary with \( P \) and \( B_z \) (Stubbs et al., 2004). We used the Shue et al. (1988) model and the fact that typically the distance to the subsolar bow shock \( R_{bso} = 1.3R_{mpo} \) (Fairfield, 1971). Thus, we get
\[ \rho_{bs} = \sqrt{y_{bs}^2 + z_{bs}^2} = H \left( \frac{D + R_{mp} - x_{bs}}{D} \right) - 1 \]  
(5)

where \( H \) and \( D \) are empirically determined constants and have values \( H = 56.7R_E \) and \( D = 177.7R_E \).

The models in the form given above are rotationally symmetric about the \( x \)-axis, and so we need to include aberration of the bow shock and magnetopause boundaries due to (1) the orbit of the Earth–Moon system about the Sun, 30 km/s in the negative GSE \( y \)-direction, which causes the magnetotail to align itself on average 4° off the Sun-Earth line in the positive GSE \( y \)-direction and (2) the non-radial component of the solar wind flow, which can sometimes result in the flow vector being a few degrees off the Sun-Earth line, but is typically near zero. These non-radial components are typically defined by two angles, the flow latitude \( \psi_0 \) and the flow longitude \( \psi_6 \).

The magnetotail responds much like a "wind sock" to these effects. However, when comparing the location of the Moon \((x_m, \rho_m)\) with these boundaries, it is more convenient to invoke an aberration of the location of the Moon \((x_m', \rho_m')\) in order to account for the above effects and make the comparison in the axi-symmetric "model" frame.

We are interested in finding the closest approach (i.e. during these events) of the Moon to both the bow shock and magnetopause boundaries, so we search for a minimum in both the following equations:
\[ \Delta_{bs} = [(x_{bs} - x_m)^2 + (\rho_{bs} - \rho_m)^2]^{1/2}, \]
(6a)
\[ \Delta_{mp} = [(x_{mp} - x_m)^2 + (\rho_{mp} - \rho_m)^2]^{1/2}. \]
(6b)

Because the preliminary results indicate that the SLE resonance events are associated with the bow shock rather than the magnetopause (Collier et al., 2008), we have focused on

calculating the uncertainties in $d_{ml}$ in an attempt to verify this. The solar wind parameters acquired from OMNIWeb are hour-averaged with values for the mean and standard deviation $\sigma$. Table 4 lists the uncertainties used in each of the parameters that go into calculating the model magnetopause and bow shock locations. We propagate the uncertainties in Table 4 assuming that the variables are uncorrelated.

Fig. 11 shows a comparison with the bow shock and magnetopause models for SIDE event 1. This figure shows the effects of the aberrations, as well as the points on the bow shock and magnetopause corresponding to the minimum distance to the Moon at the time of the event. The same plots have been generated for events 2–7 and Fig. 11 is a typical example. Fig. 12 shows the minimum bow shock distance results for just the fully aberrated Moon locations for each event, along with the uncertainties. The closest approach to the magnetopause for each event is also indicated in Fig. 12 for comparison. The overall mean and standard deviations of the closest approach to the bow shock and magnetopause for all seven events are plotted on the right side of the figure. This analysis shows that, according to the models, these events are at the bow shock.

8. Downstream lunar potential

Spacecraft charging is frequently used as an analogy for the charging of the lunar surface and, indeed, the first predictions of lunar surface potential were based on equations developed for spacecraft charging. Thus, it might be illuminating to examine if and how the potential of spacecraft change as they traverse the bow shock at or near $60R_e$.

Fig. 13 shows the Wind spacecraft potential, electron density, and electron temperature observations as it traversed the terrestrial bow shock on February 27, 1999 (day 58), at about 09:43 UT at 68.8$R_e$. The potential is determined iteratively to find the potential at which the incident electron current as measured by 3DP and the spacecraft photocurrent balance. Note that the spacecraft potential rises by a couple volts as it passes through

the bow shock. This might suggest that the processes that cause the spacecraft potential to increase as Wind traverses the shock are the same as those that cause the lunar surface potential to increase during its shock crossings. However, the Wind spacecraft potential changes suddenly and is therefore consistent with the spacecraft potential responding to the discontinuous change in plasma parameters across the shock. Note that the time scale of the x-axis is quite different for the Wind data in Fig. 13 which shows only 3 h and the SIDE data in Fig. 3 which shows six days. In Fig. 13, the Wind potential change occurs in less than a minute while in Fig. 3, the potential change as determined by the center of the count rate enhancement occurs over about three days. In the case of the SIDE data, there must be a different process at work because the time scale of the potential change is of the order of a day (e.g. Fig. 3).

Given that observationally these SIDE resonance events appear to be closely associated with the Moon's traversal of the terrestrial bow shock, one might ask why the lunar surface potential, at least at the location of the Apollo 14 SIDE, might increase in conjunction with these traversals and, further, why the potential would approach 10 V positive. The first step in addressing such a question might be an evaluation of the lunar surface potential both downstream and upstream of the bow shock.

Determining the equilibrium potential of the lunar surface, or for that matter, of any body immersed in a plasma, becomes an exercise in identifying and evaluating the currents into and out of the body and requiring that they be equal (e.g. Whipple, 1981). To zeroth order then, traditionally the lunar surface potential on the dayside can be thought of as a balance between the current densities due to solar wind electrons, \( I_{\text{sw}} \), solar wind protons, \( I_{\text{swp}} \), and that due to photoemission, \( I_p \), so that

\[
I_{\text{sw}} + I_{\text{swp}} + I_p = 0. 
\]

In the presence of a surface potential \( \phi \), the photocurrent at a solar zenith angle \( \theta \) is given by (Manka, 1973)

\[
I_p = I_p \cos \theta \exp \left( -\frac{\phi}{k_B T_p} \right). 
\]
where $e$ is the electron charge, $i_p$ is the lunar photocurrent under normal incidence and $k_b T_e$ is the thermal energy of the emitted photoelectrons.

Feuerbacher et al. (1972) measured the photoelectron current from a sample of lunar fines assumed to be representative and estimated the density and electric field distribution in the sheath by approximating the electron distribution by a Maxwellian. They found a maximum photoelectron yield per incident photon of about 7% at 900 Å wavelength, much lower than typical yields for insulators which is of the order of 30%. They attributed this to the material being in the form of a fine grained powder leading to a higher probability of photoelectron reabsorption than on a flat surface, an effect that also causes the Moon's low albedo. Photoelectron energy distributions were shown to peak around 2-3 eV with a mean kinetic energy of 2.2 eV and a very weak tail. They also estimated the lunar photocurrent under normal incidence to be $4.5 \times 10^{-6}$ A/m² under typical conditions. Consequently, we take $i_p = 5 \times 10^{-10}$ A/cm² and $k_b T_e \approx 1.5$ eV.

The current density onto a positively charged surface due to solar wind electron flux is given by Manka (1973) as

$$I_{swe} = -ne^2 \frac{\sqrt{k_b T_e}}{2 \pi m_e} \left(1 + \cos \theta \right)$$

(9)

where $n$ is the electron density taken to be about $10$ cm$^{-3}$, $k_b T_e$ is the thermal energy of the solar wind electrons, taken to correspond to $T_e = 1.4 \times 10^5$ K (Newbury, 1996) or about 12 eV (a reasonable value under most conditions), and $m_e$ is the electron mass. This yields an $I_{swe} = 9.3 \times 10^{-11}$ A/cm².

The current due to solar wind protons is generally lower than that due to solar wind electrons because the protons are supersonic with a bulk flow speed less than the typical electron thermal speed. Consequently, we can evaluate the solar wind proton current as

$$I_{swp} = ne V_{sw} \cos \theta = 5.1 \times 10^{-11} \text{ A/cm}^2$$

(10)
where \( v_{sw} \) is the solar wind convection speed, taken to be 500 km/s, and the other variables are as defined above with the solar zenith angle taken to be 50°, typical for the Apollo 14 resonance events considered here. In principle, Eq. (10) has a cut-off, too, but because the 1000 eV nominal energy of solar wind protons is much greater than the lunar surface potential, this term is unity.

Using Eq. (7) above and plugging in various expressions, we get

\[
9.3 \times 10^{-11} \text{ A/cm}^2 = 5 \times 10^{-10} \text{ A/cm}^2 \cdot \cos(50°) \cdot \exp\left\{-\frac{\phi}{1.5 \text{ V}} \right\}
\]

\[
+ 8.0 \times 10^{-11} \text{ A/cm}^2 \cdot \cos(50°).
\]  

(11)

Note that there is no \( \phi \) term, i.e. no lunar surface potential term, in the solar wind electron and proton currents in Eq. (11) because the electron flux toward the surface is conserved for positive potentials (Liouville's Theorem) and because the protons are at such a high energy that the small positive potential does not affect them.

The top panel of Fig. 14 shows in red the currents out of the surface or the left hand side of Eq. (11) and in blue the currents into the surface or the right hand side of Eq. (11). The solid red and blue lines are the total current out of and into the surface, respectively, and the intersection of the two lines which is around \( \phi \sim 3 \text{ V} \)

(12)
Fig. 11. A comparison between the modeled bow shock and magnetopause boundaries for SILE event 1 on 4–5 January 1972. The results are plotted in an axi-symmetric coordinate system which accounts for the aberration from GSE caused by the Earth–Moon orbit about the Sun and the non-radial component of the solar wind flow. The effect of this aberration on the location of the Moon with respect to these boundaries is indicated. Also shown are the closest approaches of the Moon to these boundaries, which reveals that it was much nearer the bow shock. The magnetopause boundary is calculated using the Shue et al. (1998) model for \( P = 2.3 \text{nPa} \) and \( B \text{= 2.2 nT} \), while the bow shock boundary is calculated using a modified version of the Howe and Bieseck (1972) model (see Stubbs et al., 2004).

Fig. 12. The distance of the Moon from the predicted bow shock location during each of the seven SILE events listed in Table 1 [black dots]. These distances are calculated following the example shown in Fig. 11 for event 1, with the positive direction representing positions in the solar wind. Each of the SILE events is shown to be associated with the bow shock, with the estimated uncertainties (see Table 4). The mean distance to the bow shock for all events is 0.3Re with a standard deviation of 3.2Re (open circle). Also shown is the distance of the magnetopause from the bow shock at closest approach (crosses), which has a mean value –18.4Re with a standard deviation of 4.7Re (open square).

Fig. 13. Data from a Wind spacecraft bow shock crossing at about the orbit of the Moon. Panel A are Wind electron moment calculations including a spacecraft potential calculation. Panels B and C are the electron density and temperature, respectively.

the solution to Eq. (11). So, the lunar surface potential in the downstream region is a few volts positive.

9. Upstream lunar potential

The region upstream of the bow shock but downstream of the interplanetary magnetic field line tangent to the bow shock is known as the foreshock (e.g. Fuselier, 1995; Anderson et al., 1979). The plasma in this region is different from that in the downstream magnetosheath because it includes particles reflected and energized by the bow shock through shock drift acceleration. The presence of these energetic particles in the upstream region requires a modification to the calculation presented above for the downstream magnetosheath. In particular, the shock produces beams of energetic electrons moving away from the shock along the magnetic field lines (Fitzsimmons et al., 1996). These beams peak at about 300 eV (~10,000 km/s) and have densities ~3–5% of the solar wind density (A. Vithas, private communication, 2009). As discussed earlier, because of the secondary emission of electrons from the surface produced by this population of energetic primary electrons, we must add another term to the right hand side of Eq. (11) representing an additional source of current into the surface.
Although the upstream solar wind electron beams are formed at the shock through reflection by the jump in the magnetic field and electric potential in the shock transition layer (Leroy and Mangeney, 1984), this process is most effective near the tangent normal to the shock scientists. As the Moon approaches the Earth, the incident beam will be reflected from the lunar surface (Wieser et al., 2009). However, these particles are reflected as neutral atoms and therefore have no surface charging consequences and so will be ignored in this analysis.

As far as η values go, at low energies, measured η values have particularly large scatter (ranging from 0.1 to 0.6), perhaps due to surface conditions and vacuum environment. Furthermore, the backscatter coefficient increases as the incident beam goes from normal to oblique incidence (El-Gomati and Assaad, 1998). So, given the geometry of the SIDE events, η is likely substantially larger than for normal incidence. For the calculation and discussion below, we take η to be 0.3 which is a typical value well within the spread.

We will estimate the net current into the surface in the upstream foreshock region due to energetic electron beams, I_{beam}, as the sum of the energetic electron beam current itself (negative), the backscattered current (positive), and the secondary electron current (positive):  
\[ I_{beam} = n_0 F_{beam} \{-1 + \eta + \exp(-\frac{e \phi}{T_{eff}})\} \]  
where \( n_0 \) is the solar wind electron density taken to be 10 cm\(^{-3}\), \( F \) is the fraction of the solar wind density in the foreshock electron beams, taken to be 0.05 (5%), \( e \) is the secondary electron yield, taken to be 3, \( e \) is the electron charge, \( v_{beam} \) is 300 eV or 10,000 km/s, \( \eta \) is the reflection coefficient taken to be 0.3, and \( T_{eff} \) is the effective temperature of the secondary electrons, taken to be 11 eV. In expression (13), the −1 in the brackets represents the

When irradiated by an electron beam, solid surfaces emit low energy secondary electrons. Typically, to quantify the emission process, one defines the total secondary yield as the ratio between the emitted and incident number of electrons. This ratio can be greater than unity for primary electron energies below a few keV, in which case more electrons leave the surface than impinge on it and the entire process represents a net current into the surface (Pintado and Hessel, 2000; Horanyi et al., 1998).

The shape of the total electron yield curve with primary electron energy is zero at zero energy, rising rapidly to a peak at typically around 300 eV (Pintado and Hessel, 2000) and then exhibiting an extended power law tail at higher primary energies. (Note that Nitter et al. (1998) exclude secondary electron emission in the solar wind on the basis that there exists a low energy cut-off at about 10 eV below which secondary electron emission is negligible.) The peak yield varies by material (e.g. Seller, 1983), but is typically 2–3 for silica, for example. The energy distribution of the secondary electrons also appears to be quite wide, as high as ~15 eV or so with power law tails, and essentially independent of the energy of the primary electrons (e.g. Scherer and Hoelzl, 1972). Of course, these parameters also depend on the surface topography, which could be very relevant in the case of the Moon.

In addition to the secondary electrons, there is also a population of elastically backscattered and elastically reflected electrons which have significantly higher energies than the secondary electrons. Backscattered electrons are typically considered to be those with energies greater than 50 eV while secondary electrons are considered to be those with energies <50 eV. The ratio of the number of electrons that leave the sample with energies >50 eV to the total number of incident electrons is denoted by \( \eta \) (El-Gomati and Assaad, 1998; Seller, 1983; Whipple, 1981).

Although it was once assumed that the regolith completely absorbs the impinging solar wind, those engaged with the lunar community recognize that the first observations of neutral atoms from the Moon made by the Sub-keV Atom Reflecting Analyzer (SARA) instrument on the Indian Chandrayaan-1 spacecraft showed that up to 20% of the impinging solar wind protons are reflected from the lunar surface (Wieser et al., 2009). However, these particles are reflected as neutral atoms and therefore have no surface charging consequences and so will be ignored in this analysis.

current due to the incident beam while $\delta$ and the exponential represent the current due to electrons the incident beam releases from the surface, that is the secondary electron yield. $\eta$ represents the fraction of the incident beam that is backscattered. Thus,

$$I_{eb} = 8.0 \times 10^{-11} \left( -0.7 + 3 \cdot \exp\left(-\frac{\phi}{11V}\right) \right) \text{A/cm}^2,$$

(14)

which is positive for small values of $\phi$ causing the surface to adopt a greater positive potential. Of course, this should only be viewed as a qualitative estimate of these additional currents using nominal values. To be more precise, one would need more accurate values for the lunar surface regolith at the Apollo 14 site than are available.

One interesting aspect of the secondary electron emission is that, because of the small escape depth of secondary electrons, changing the incidence angle from normal to shallower angles increases the path length of the primary electrons near the surface, increasing the excitation of secondary electrons, such that the yield increases approximately as $1/\cos \theta$ where $\theta$ is the angle measured relative to the surface normal or, in our case, solar zenith angle (Sailer, 1983; Nishimura et al., 1994). Because the flux of a beam of electrons hitting a surface decreases approximately as $\cos \theta$ where again $\theta$ is the angle measured relative to the surface normal, the current due to secondary electron emission will be relatively independent of the angle at which the field aligned beam hits the lunar surface.

The bottom panel of Fig. 14 shows the same quantities as the plot in the upper panel but with the contribution due to secondary electrons added as the dash-triple dot curve. The solid blue curve is the new total current into the surface which now intersects the red curve, the current out of the surface, at a potential of about 10 V. So, plausibly, at least, secondary electron emission from primary electrons in the foreshock field-aligned beams can raise the lunar surface potential to $\sim 10$ V.

As mentioned above, it is important to note that the transition shown in Fig. 14 in the bottom panel will occur gradually as the Moon moves through the foreshock rather than suddenly as it moves through the shock. This is not to say that there will not be a sudden potential change as the Moon moves through the shock, as well. In fact, it is likely that a sudden potential change due to the change in plasma parameters at the shock does occur at SIDEn as well as at Wind. However, the potential change, if similar to that observed by Wind, would not create a resonance event because the potential would have to exceed about nine volts to be observed.

The presence of the electron beams themselves implies that the solar wind terms from the previous analysis of the downstream potential must change, at least slightly. For example, to maintain neutrality, the solar wind electron density in the upstream must be 5% lower to compensate for the electron beam density. However, because we present primarily a qualitative analysis, this is a small effect, and for ease of presentation, we have neglected any modification to the solar wind terms due to the presence of the energetic electron beam in this analysis.

Along similar lines, it should be emphasized that the calculations here are qualitative, and we have assumed that the solar wind currents in Eq. (7) as shown in the top panel of Fig. 14 remain the same in going from the downstream to the upstream region. Although this will be approximately true (e.g. Manka, 1973, Table 1), various parameters do change across the shock as Fig. 13 illustrates and these could effect a small change in potential. The Wind spacecraft for example goes from about six to about eight volts as it crosses the shock in Fig. 13, but as mentioned above, this is a smaller and quicker change in potential than necessary to explain the observations (e.g. Fig. 3).

This raises the issue of the geometry of these events. Fig. 15 shows the typical geometry associated with the SIDEn resonance events discussed here. In this figure, the lunar orbit is shown as the dotted line, the Howe and Binsack (1972) nominal bow shock is shown with the thick solid line, and the nominal Parker spiral interplanetary field orientation is shown with the thin solid line. Note that if secondary electrons are the cause of these events, then this location would be close to ideal to observe them. This is because SIDEn is magnetically connected to the bow shock, or at least very close to connected, for the nominal IMF configuration. Furthermore, SIDEn is far enough away from the terminator to have the potential to be positive and hence observable as a resonance. That is, if SIDEn were closer to the terminator, the potential would be negative and if SIDEn were closer to the subsolar point, it would almost never be connected to the shock so we would not observe the effect of the secondary electrons.

However, from Fig. 15, it appears because the direction of the IMF fluctuates so that the field line from the bow shock will be connected to the SIDEn location without having to cross through the Moon less than half the time. So given this geometry, it is unclear why these events would so consistently occur when the Moon crosses the bow shock. One consideration is the gyroradii of the electrons. The gyroradius, $r$, of $v_e = 10,000$ km/s electrons in a $B = 10$ nT field is

$$r = \frac{m v_e}{q B} = \frac{9.1 \times 10^{-31} \text{kg} \cdot \text{m/s}}{1.6 \times 10^{-19} \text{coul} \cdot 10^{-5} \text{T}} = 5.7 \text{ km},$$

(15)

where $m$ is the electron mass and $q$ is its charge. So, the gyroradii are not large enough so that they will intersect the surface even if they are on a field line that passes well above the surface. Of course, the bow shock is in actuality a dynamic boundary whereas we use here a highly simplified averaged model.

It is conceivable that the surface potential could be equilibrating over a relatively large region, so the energetic electrons are intersecting nearby, but driving the surface potential more positive over a large region of the dayside particularly because the conductivity of lunar fines is much higher in sunlight. Furthermore, it is possible that there are still some energetic electrons on flux tubes that no longer connect to the bow shock as...
they convect downstream, perhaps from some of them reflecting off upstream structures, etc. So, it is quite possible that the upstream field lines may not need to directly connect the bow shock to the lunar surface for the lunar potential to be affected by energetic electrons.

10. Discussion and conclusion

The lunar surface potential is an important factor in the overall lunar environment affecting such aspects as dust transport (Stubbs et al., 2006) and the availability of resources (Farrell et al., 2010; Crider and Vondrak, 2003). The analysis here shows that the lunar surface potential increases to about 10 V and higher as the Moon traverses the terrestrial bow shock from the magnetosheath into the foreshock. Because the Moon is a paradigm for similar bodies throughout the solar system, these results are surely applicable to the general problem of airless bodies, such as Mercury, asteroids, and Kuiper belt objects interacting with shocks. The lunar exosphere/surface interaction occurring in the flowing solar wind represents a fundamental physics problem. This analysis of in situ lunar data suggests that secondary emission from field-aligned electron beams in the foreshock region can significantly affect the surface potential of solar system bodies.

Because the onset of these events corresponds very closely to passages through the bow shock, the length of these events may be related to the size of the foreshock region in which the electron beams impact the site of the Apollo 14 SIDE experiment. However, the change in potential of the surface is still very rapid. Rapid changes in potential could result if the equilibrium potential is an unstable equilibrium. This can occur in situations in which secondary emission is important. For example, Prokopenko and Laframboise (1990) have shown that when the environmental plasma is not Maxwellian, an almost universal condition for space plasmas (e.g., Collier, 1993), large, sudden changes in the surface potential can be produced by gradual changes in the external environment.

As further evidence for the potential importance of field-aligned electron beams for lunar surface charging, Vogt et al. (1976) showed that for two substorm events observed by ATS-5 in which the current densities from the field-aligned currents were about 1% of the omnidirectional current density that the charging of spacecraft insulators could be determined by the field-aligned fluxes. Vogt et al. (1976) suggest that field-aligned electron beams may explain why spacecraft charging anomalies such as command changes and surpising noise do not occur on a one-to-one basis with geomagnetic activity. Although not specifically addressing the issue of secondary electron emission, the Vogt et al. (1976) work does illustrate the importance of electron beams play in charging space objects even in situations in which the current density in the beams is significantly lower than that of the omnidirectional flux. Of course, in the lunar cases considered here, these currents densities appear comparable.

No SIDE Apollo 14 resonance events were found on the dusk flank. This is likely a consequence of two factors. First, given the Parker spiral configuration of the IMF, it is less likely to geometrically connect the Moon with the bow shock close to the nose on the dusk flank as compared to the dawn flank. Second, the location of the Apollo 14 SIDE places it on the night side of the terminator on the dusk side where potentials are typically negative so SIDE would be unable to see resonance events that require a positive potential.

This study has concentrated on the Apollo 14 SIDE resonance events. We intend to examine both the Apollo 12 and Apollo 15 SETA data sets, as well, to search for events similar to the ones discussed here. However, a cursory look shows that, in comparison to the Apollo 14 site, relatively few of these types of resonance events were observed at Apollo 12 or Apollo 15 sites. Because these SIDE resonance events rely on the presence of photoions, it is possible that the Apollo 14 SIDE was the only one at the correct lunar longitude to observe the events on the lunar nightside during the shock crossings, a possibility which, as discussed earlier, is supported by Fig. 15.

Also possible, if secondary electron emission is the cause, is that some regions of the lunar surface may be better connected or have a more favorable orientation to the interplanetary magnetic field than others. For example, the magnetic field at the Apollo 12 location is relatively large, about 38 nT, which can alter the direction of the field line and hence the electron beam and may suppress the electron beam current density that arrives at the lunar surface. Furthermore, if secondary electron emission in the electron foreshock is a prominent effect, then the nightside which will be readily magnetically connected to the shock in the foreshock may have its negative potential moderate if the flux of backstreaming electrons that can overcome the potential barrier is significant.

If this should be correct, then, because photoelectrons may be neglected in calculating the lunar surface potential at \( \sim 10 \) V in the foreshock, the physical interpretation of the increase in the observed lunar surface potential at the bow shock is that the Moon, or at least the portion of the Moon where the Apollo 14 SIDE makes its measurements, undergoes a transition from a region where photoelectrons dominate to a region where secondary electron emission dominates. If the secondary electron temperature is much higher than that of the photoelectrons, then this will in general be the case in regions where the lunar surface potential is much greater than the photoelectron temperature (=2 eV).

Acknowledgments

Special thanks to David Simpson for help with the programs to read the SIDE data, as well as Jan Merka and Adam Szabo for the listing of Wind bow shock crossings. Also thanks to Davin Larson for the Wind spacecraft potential data and Rob Kilgore for drafting some of the figures. This material is based upon work supported by the National Aeronautics and Space Administration under LASER 132438.01.15.01.06 issued through the Science Mission Directorate/Planetary Science Division and Exploration Systems Mission Directorate/Advanced Capabilities Division, through LRO Participating Scientist Grant no. NNX08AB76G (T.J.S.), and through the Dynamic Response of the Environment At the Moon (DREAM) NASA Lunar Science Institute team.

References


