9

EUGENE N. PARKER*

A history of the solar wind concept

One presumes that the supersonic solar wind has been blowing since the formation of the Solar System, providing the dynamical plasma and magnetic field throughout interplanetary space. It appears that the early massive solar wind and simultaneous strong solar magnetic fields carried away the initial angular momentum of the young Sun during the first 3×10^8 years or so, leaving the leisurely 25-day equatorial rotation period that we see today (Schatzman 1959, 1973; Biermann 1973). Nowadays the tenuous wind carries no significant angular momentum or mass from the Sun, but it continues to be the dominating condition in interplanetary space, providing the outward sweeping spiral magnetic field, impacting the terrestrial magnetosphere, drastically reducing the intensity of the galactic cosmic rays, and pushing the interstellar gas and galactic magnetic field out beyond the farthest planets. The consequences for the terrestrial environment are profound, but not immediately obvious to us who dwell at the surface of Earth. So the history of the thoughts leading to the recognition of the solar wind extends back more than two millennia. Progress has been paced largely by the development of physics from the days of Gilbert (1544–1603) and Galileo (1562–1642). That is to say, with the exception of the aurora, the effects of the solar wind are not available to our biological senses, but require observations and measurements that can be made only with specially devised scientific instruments. And then it requires an advanced state of physics to infer the implications of the observations. So the history of the solar wind follows both the advances in fundamental classical physics and the studies of the natural phenomena pertinent to the solar wind.

The presence of the solar wind is directly indicated by the continuing fluctuations of the geomagnetic field at the surface of Earth, particularly at high latitudes, the continuing aurora at high latitudes, the varying intensity of the galactic cosmic rays, and the antisolar orientation of the gaseous tails of comets. However, these effects, known individually for decades and centuries, are evidence only of some form of external disturbance, and the scientific challenge over the last century has been to work out precisely what that disturbance really is.

Historically there has been no end of ambiguities and distractions. For instance, starting from the classical point of view that space is completely empty, Kelvin proved that the Sun is not capable of such large magnetic variations as would extrapolate to the observed geomagnetic fluctuations at a distance of 1 AU. From this he asserted that there can be no connection between the activity of the Sun and the magnetic fluctuations at Earth. With the same hypothesis that space is completely empty it was thought possible that the observed variations in the cosmic ray intensity are a consequence of large electrostatic potential differences across interplanetary space. The antisolar orientation of gaseous comet tails was attributed to the radiation pressure of sunlight. The implications of the continuing aurora and high-latitude geomagnetic fluctuations were largely ignored, their origin perhaps meteorological.

To go back to the beginning of the concepts on which the solar wind is based, it must be recalled that the concept of space itself is essential and is only a relatively recent development in human thought. In primitive times the sky was not viewed as the window looking out into space. Rather the sky was the abode of gods and spirits, who looked after, or

*University of Chicago, IL, USA

Eugene N. Parker, The Century of Space Science, 225–255

were represented by, the Sun and the Moon and the various planets and stars, which in turn had a powerful influence on our personal destinies. The sky was "that inverted bowl they call the sky whereunder crawling coop'd we live and die" (Omar Khayyam). Knowledge of geography was as limited as the concept of the sky. Some mythologies would have humans descending to Earth from a beginning in the sky, and not so very long ago at that. The concept of the span of time was as stunted as the localized concept of the sky and space. The idea that the whole world was created around us primarily to accommodate us humans is still alive and well today, demonstrating the reluctance of the human mind to relinquish the egocentric world of fantasy.

So it is not surprising that the realization of the vast space in which the tiny Earth resides developed only slowly and sporadically. We owe great respect and admiration to Aristarchus of Samos (circa 275 BC), as the first ancient philosopher-scientist of record to recognize the central position of the Sun in our local system and the vastness of the surrounding space. Armed with a clear understanding of geometry he realized from the converging umbral shadow of Earth during an eclipse of the Moon that the Sun is larger than Earth. From this he recognized that it is more likely that little Earth orbits the large Sun, rather than vice versa. He viewed the Solar System much as Copernicus did 1800 years later, recognizing the great distance to the stars. Needless to say, the contemporary experts found his ideas disturbing, undermining their infallibility, so he was officially rejected. Fortunately he was not entirely ignored, and his penetrating insights, although lost in the original, have come down to us through the writings of others.

Copernicus (1473–1543), beginning in 1530 and finally published in his *De revolutionibus orbium coelestium* in 1543, swept away the Earth-centered crystal spheres of the Ptolemaic geocentric system (first developed by Hipparchus *c*. 150 BC), thereby clearing space of its classical intellectual rubbish and making way for the concept of local terrestrial effects from distant astronomical sources. Equally important, of course, Copernicus, followed by Kepler (1571–1630) and by Galileo with the application of the telescope, began to make sense out of celestial mechanics. This prepared the way for Newton (1642–1727) to develop the Newtonian theory of gravitation and mechanics published in his *Principia* in 1686.

The gradual recognition of the magnetic field of Earth, with the invention of the magnetic compass in China during the Han dynasty about two millennia ago, was another important step on the long road leading to an understanding of the solar wind. The scientific credit here goes particularly to Gilbert (1600) whose ingenious, careful, and quantitative laboratory experiments first established the precise form of the dipole magnetic field around a more or less uniformly magnetized sphere (composed of magnetic iron oxide, magnetite). He recognized that the magnetic field is a special stress system in the space surrounding the magnetized sphere – in other words, that the phenomenon of forces between magnetic objects involves a force-field extending through the space between them. He was the first, then, to recognize the magnetic field as a physical entity throughout the space around a magnet.

The accumulating records of magnetic declination and inclination from extensive ocean voyages were enough for Gilbert to show that Earth itself is a magnetized sphere, and he pointed out that the surrounding space is a special region as a consequence of the magnetic field extending out from Earth. In short, Gilbert recognized the terrestrial magnetosphere, as seen from its lower boundary at the surface of Earth.

GEOMAGNETIC ACTIVITY AND CLASSICAL PHYSICS

It was more than a hundred years after Gilbert's work that Graham (1724) observed a delicately suspended magnetic needle with a microscope and discovered the slight agitation of the needle, implying fluctuations in the magnetic field of Earth. One can appreciate the care that must have been taken to avoid vibrations and air currents to be sure that the agitation was of geomagnetic origin. Celsius (1741) became engaged in similar observations and noticed that the appearance of the aurora was accompanied by enhanced magnetic fluctuations. Celsius and Graham corresponded and soon established that the observed periods of enhanced fluctuation were often simultaneous, thereby showing their large geographic scale.

Wilcke (1777) recognized that the auroral rays are oriented along Gilbert's dipole geomagnetic field. Canton (1759) introduced the first observational connection to the Sun, pointing out that the quiet-time geomagnetic fluctuations are stronger in the summer when the hemisphere is tilted toward the Sun.

It is interesting to note that de Mairan (1754) proposed that the aurora arises from the entry of particles from the Sun into the terrestrial magnetosphere. His idea was based on the contemporary interpretation of the zodiacal light as an extension of the solar corona. Thus Earth orbits through this extended corona, and he proposed that the aurora and the associated geomagnetic fluctuations are the result. It was an inspired conjecture, but he could not do more because the physics of charged particles and magnetic fields was unknown at the time.

Cardan in 1551 had clearly distinguished magnetic and electric effects, and Gilbert emphasized the fundamental difference. In 1733 Du Fay distinguished between positive and negative charge, and a decade later Franklin carried out a number of experiments establishing the conservation of total electric charge. He described positive and negative charge in terms of a surplus or deficit, respectively, of some fundamental electric fluid. We recognize today that electrons make up the mobile electric fluid in most circumstances, so that positive charge actually represents a deficit and negative charge a surplus, but one can see how Franklin advanced the concept.

Priestley discovered the inverse square force law between electric charges in 1767, and Coulomb rediscovered the law in 1785, with the result that today the name Coulomb is associated with the electrostatic inverse square law $1/r^2$. It was not until 1820, however, that Oersted made the experimental connection between electricity and magnetism, demonstrating the magnetic field around an electric current. With this physics in hand, one could understand how electrically charged particles might produce magnetic fluctuations. Only three years later Ampère demonstrated the equivalent magnetic fields of a closed electric circuit and a uniformly magnetized shell whose periphery is bounded by the electric circuit. Today we write Ampère's law in the familiar differential form $c\nabla\times\mathbf{B} = 4\pi\mathbf{j}$, where $\mathbf{j}$ is the electric current density and $\mathbf{B}$ is the magnetic field. By 1831 Faraday had shown that an electric current is driven through a conductor by a time-varying magnetic field, indicating the presence of an electric field whenever $\partial\mathbf{B}/\partial t$ is nonvanishing. The differential form of Faraday's law of induction is the familiar $\partial\mathbf{B}/\partial t = -c\nabla\times\mathbf{E}$, of course. Maxwell completed the electromagnetic equations in 1864, introducing the displacement current $\frac{1}{4\pi}\partial\mathbf{E}/\partial t$ into Ampère's law to obtain $4\pi\mathbf{j}+\partial\mathbf{E}/\partial t = c\nabla\times\mathbf{B}$. Maxwell pointed out that the complete equations predict the electromagnetic wave propagating through vacuum at a speed c given by the ratio of the e.m.u. unit of charge to the c.g.s. unit of charge. Or, as we would say today, the ratio of the units of charge is given by the speed of light, the speed of light being measured directly in the laboratory.

It is an interesting, and not entirely unexpected, historical fact that Maxwell's extension of Ampère's law, based on a direct analogy with an electric current in a dielectric, and the consequent prediction of electromagnetic waves, was largely ignored in England for the next several decades. Kelvin declared that he did not understand Maxwell's electromagnetic theory (which was certainly the case), did not need Maxwell's electromagnetic theory because light and electromagnetic waves could just as well be explained by an elastic aether, and could see no benefit from Maxwell's theory (an immediate consequence of his lack of understanding of the theory). For instance, no reference is given to Maxwell's electromagnetic theory in the extensive articles on electricity, magnetism, and light to be found in the 1894 edition of the *Encyclopaedia Britannica*, except for some sentences in Maxwell's biography where the writer remarks on Maxwell's contributions to electricity and magnetism, citing Maxwell's *Treatise on Electricity and Magnetism* (1873) and noting that the primary test of Maxwell's theory of electromagnetism is whether the velocity of light is accurately given by the ratio of the units of charge in e.m.u. and c.g.s. A few, such as Heaviside, recognized the importance of Maxwell's complete electromagnetic equations, as did Helmholtz and Rowland, and particularly Hertz whose experiments on electromagnetic waves eventually convinced the world of the fundamental importance of Maxwell's equations. Lorentz (1895) demonstrated the implications of Maxwell's equations for transforming $\mathbf{E}$ and $\mathbf{B}$ from one moving coordinate frame to another, which became part of Einstein's (1905) special relativity theory 10 years later.

This is perhaps an appropriate place to note an additional fundamental contribution of Faraday, who, in the course of his epic experiments on magnetic induction, recognized the central importance of the concept of field lines, or magnetic and electric lines of force, as the most direct and effective means for visualizing the form and behavior of a field. The concept is widely applicable today, particularly in magnetohydrodynamics (MHD), wherein the magnetic field is transported bodily with any medium that cannot support a significant electric field in its own moving frame of reference. The magnetic field and, of course, the magnetic energy and stress as well, move with the fluid, as if each field line were a thin line of ink entrained in the fluid, thereby providing a direct visualization of the deformation and transport of the magnetic field.

The theoretical basis for this magnetic transport arises from the researches of Faraday and Lorentz, already mentioned. Beginning with the nonrelativistic Lorentz transformations for the electric field $\mathbf{E}'$ and magnetic field $\mathbf{B}'$ in the frame of reference moving with the fluid velocity $\mathbf{u}$ relative to the laboratory frame, in which the fields are $\mathbf{E}$ and $\mathbf{B}$, we have

$$\mathbf{B}' = \mathbf{B} - \frac{\mathbf{u}\times\mathbf{E}}{c}, \qquad \mathbf{E}' = \mathbf{E} + \frac{\mathbf{u}\times\mathbf{B}}{c}$$

Then since $\mathbf{E}'\cong 0$ in the conducting fluid, it follows that $\mathbf{E} = -\mathbf{u}\times\mathbf{B}/c$, so that $\mathbf{B} = \mathbf{B}'$ upon neglecting terms second order in u/c compared to one, and Faraday's induction equation reduces to the familiar MHD result (Alfvén 1950)

$$\frac{\partial\mathbf{B}}{\partial t} = \nabla\times(\mathbf{u}\times\mathbf{B})$$

Poynting (1884, 1885) showed the mutual compatibility of Newtonian mechanics and Maxwellian electromagnetic theory, with the concept of the electromagnetic stress tensor $-\delta_{ij}(E^2+B^2)/8\pi+(E_iE_j+B_iB_j)/4\pi$ representing the isotropic pressure $(E^2+B^2)/8\pi$ and the tension $E^2/4\pi$ along the electric field lines and the tension $B^2/4\pi$ along the magnetic field lines. Thus, static equilibrium of an electric field involves a balance between the pressure gradient and the tension along curved field lines. The same applies for static equilibrium of a magnetic field. The energy densities of the electric and magnetic fields are $E^2/8\pi$ and $B^2/8\pi$, respectively. At this point in time one could begin to appreciate Gilbert's concept that a magnetic field represents a physically real medium – or stress system (Gauss 1839). The field is not a ponderable medium in that it is not made up of matter, but it is an elastic physical medium with its own energy and, therefore, its own mass density. Thus, for instance, the energy of Gilbert's geomagnetic field above the solid surface of Earth is 8×10^{24} erg, contributing 9 kg to the inertial and gravitational mass of Earth – equivalent to a sack of flour.

The magnetic field becomes a ponderable elastic medium, of course, when a collisionless plasma or equivalent highly conducting medium is present. In particular, Poynting showed that electromagnetic energy is transported with a flux density $c\mathbf{E}\times\mathbf{B}/4\pi$, which becomes $\mathbf{u}_\perp B^2/4\pi$ in MHD, where $\mathbf{u}_\perp$ is the fluid velocity perpendicular to $\mathbf{B}$ and $B^2/4\pi$ represents the magnetic enthalpy. Thus the electromagnetic transport under MHD conditions ($\mathbf{E}'=0$) is the convective transport of the magnetic field and the enthalpy of that field, representing the convective transport of the magnetic energy density $B^2/8\pi$ plus the work done by $\mathbf{u}_\perp$ pushing the magnetic pressure $B^2/8\pi$ against the fluid ahead. In the nonrelativistic case the electromagnetic momentum density $\mathbf{E}\times\mathbf{B}/4\pi c$ is small to second order in u/c compared to the momentum density in the moving fluid, so that it can be neglected.

We can see how toward the end of the nineteenth century the basic classical physics was falling into place for interpreting the geomagnetic fluctuations. The final piece of physics was the realization that electric charge is not a fluid but consists of discrete particles. The quantized nature of electric charge was indicated by Faraday's experiments with electrolysis around 1831. A small integer times a basic total charge is always associated with the electrolysis of a mole of any substance. Then, noting that the electrolysis experiments showed that matter is composed of discrete particles, Helmholtz in 1881 emphasized that the electric charge, intimately associated with the discrete ions, must also come in discrete units. Johnston Stoney was the first to apply the term "electron" to the basic charge in 1874. In 1897 J.J. Thomson used a Crookes tube (cathode ray tube) to show that the particles emitted by the cathode have a mass that is the fraction 1/1840 of the mass of the hydrogen atom. Thus the idea of the individual lightweight electron and the relatively massive ion as individual free particles was born. The actual structure of the atom was not established until scattering experiments by Hans Geiger and Robert Marsden, using alpha particles from the newly discovered radioactive decay of radium, were performed in 1909, from which Rutherford pointed out in 1911 that the distribution of deflected alpha particles indicated that the atoms from which they scattered contained a massive concentrated central nucleus surrounded by a diffuse electron cloud. Millikan's first measurements of the charge of the electron (4.8×10^{-10} esu) were carried out in the same year.

THE NATURE OF AURORAE AND GEOMAGNETIC ACTIVITY

At this point the physics was ready for interpreting the aurora and the geomagnetic fluctuations, although the scant observational data, limited to instruments at the surface of Earth, held out no brilliant guiding light. Franklin in about 1750, Dalton (1828, 1834), and Gauss (1839) had all come to the view that the aurora is an electrical phenomenon, based on analogy with such electrostatic phenomena as high-voltage coronal streamers and lightning discharges. By about 1890 the shifting ray structure of the aurora reminded physicists of the cathode ray streamers in the partially evacuated Crookes tube, from which Fitzgerald and others suggested that the aurora is a similar electrical discharge. That is to say, the aurora was recognized as being caused by fast particles, and it was presumed that the associated geomagnetic fluctuations were a product of the same particles through Oersted's effect and Ampère's law.

Fortunately the observational studies were moving ahead during the nineteenth century, however slowly. Gauss (1839) became interested in the geomagnetic anomalies and fluctuations, and, beginning in 1832, established the standard magnetic observatory instrumentation. Backed by Humboldt he began organizing the building of magnetic observatories around the world over the next two decades. His instruments used a small mirror attached to a magnetic compass needle, onto which a collimated beam of light was directed. The reflected pencil of light intersected a suitable distant screen, thereby vividly displaying any agitation of the needle without resorting to the cumbersome close-up microscopic scrutiny (see Chapman (1967) for a review of the geomagnetic studies).

Humboldt publicized Schwabe's discovery of the decadal periodicity of sunspots to the scientific world (see von Humboldt 1858). Sabine (1852) pointed out the important fact that geomagnetic disturbances at Toronto are strongly correlated with the number of spots on the Sun. Lamont noted the same correlation for the magnetic variations at München.

The connection of geomagnetic activity with sunspots was reinforced when the four-week variation of the activity was recognized by Broun (1858, 1874), and shown by