# The Early Magnetic Field and Primeval Satellite System of the Moon: Clues to Planetary Formation [and Discussion]

S. K. Runcorn, H. Fechtig and R. Hide

*Phil. Trans. R. Soc. Lond. A* 1994 **349**, 181-196
doi: 10.1098/rsta.1994.0123

# The early magnetic field and primeval satellite system of the Moon: clues to planetary formation

By S. K. Runcorn

*University of Alaska, Fairbanks, Alaska, U.S.A., Max Planck Institute for Chemistry, Mainz, F.R.G., and Department of Physics, Blackett Laboratory, Imperial College of Science, Medicine and Technology, London SW7 2BZ, U.K.*

From the stable remanent magnetization of the Apollo igneous rocks and high-grade breccias the existence of a primeval lunar magnetic field was inferred. The palaeointensities of the samples rise rapidly to a maximum at 3.9 Ga, then decrease exponentially to 3.2 Ga, strongly suggesting that the Moon had a field generated in a core, the existence of which was inferred from its non-hydrostatic figure. Modelling of the Apollo 15 and 16 subsatellite magnetic anomalies, by P. J. Coleman, L. L. Hood and C. T. Russell, gave palaeomagnetic directions of crustal strata. This enabled N pole positions to be calculated, which were empirically found to form three bipolar groups, the mean poles of which define (on the core dynamo hypothesis) three axes of rotation different from the present. These were dated as Pre-Nectarian, Lower Nectarian, and Upper Nectarian–Imbrian. Multi-ring basins of these ages were found to lie close to the corresponding palaeoequators. The impacting bodies were therefore satellites, not asteroids or comets. Their velocities, before collision, can be shown (from basin asymmetries) to be nearly equatorial. The consequent changes in the moment of inertia tensor by basin formation caused these successive reorientations of the Moon relative to its axis of rotation in space. The three mean poles form a 90° spherical triangle. The explanation is that the Moon had three satellites: the orbits of each decayed, they broke up at the Roche limit into smaller bodies, which produced impact basins near the equator. The Moon then reorientated according to Euler's principle before the next group of impacts. Lunar palaeomagnetism, and especially the inferences that the Moon has an iron core that segregated late and had a primeval satellite system, may provide important constraints on theories of lunar and planetary formation.

## 1. Introduction

Perhaps the least expected discovery of the early Apollo missions was that the mare basalts and high-grade breccia returned from them possessed stable natural remanent magnetizations (NRM) (Runcorn *et al.* 1970*a*, *b*; Strangway *et al.* 1970; Helsley 1970). Laboratory experiments showed that the magnetized grains were iron of domain or multidomain size and that their primary magnetization possessed stability similar to that of terrestrial rocks, from which the history of

*Phil. Trans. R. Soc. Lond.* A (1994) **349**, 181–196
*Printed in Great Britain*

TEX Paper

the geomagnetic field has been obtained. Secondary magnetization, usually easily removed by demagnetizing techniques, had in some cases been picked up during or on return to the Earth, but confirmation of stable magnetization in lunar rocks was provided by the discovery, from the Explorer 35 magnetometer mission orbiting the Moon, that solar wind was deflected by localized magnetic fields on the lunar surface (Sonett & Mihalov 1972). These were more frequent on the highlands than on the maria. Ten years before, Luna 1 and 2 had shown that the Moon today possesses no general magnetic field, a discovery that fitted well into the view of the Moon then held; an undifferentiated body, the result of cold accretion, with no core in which a field might once have been generated. Thus the discovery of remanent magnetization led to speculation that it had resulted from the most common physical process at the lunar surface, meteorite impacts (Hide 1972), for an internally generated field seemed to most lunar scientists excluded.

## 2. The figure of the Moon and the existence of a core

However, the existence of an iron core in the Moon had been proposed (Runcorn 1962*b*, 1967). It had been well known, since Laplace's interpretation of Cassini's observational laws of the Moon's rotation, that the lunar figure departed in its second harmonic from that corresponding to hydrostatic equilibrium i.e., it had a non-hydrostatic bulge. This had always been supposed to be a primeval distortion retained until the present by the finite strength of its interior. I argued, however, that this explanation was no longer a plausible one in the light of the accumulating evidence for convection in the Earth's mantle (Runcorn 1962*a*,*c*) at creep rates of about $10^{-14}$ $s^{-1}$: in the Moon a creep rate of $10^{-20}$ $s^{-1}$ would have removed such a primeval bulge. Laplace's analysis had determined the fractional difference in the moments of inertia of the Moon and, assuming uniform density, he obtained a radius towards the Earth 1 km greater than that in the plane of the sky: twenty times the hydrostatic value. I argued that this was dynamically maintained at present by convection and because the ellipsoidal figure implied the existence of a two-cell convection pattern, I concluded, using marginal stability convection theory, that a core of between 100 km and 600 km in radius would have to be postulated. If the core were non-existent, the theory would predict the development of a single-cell convection pattern, which may have been dominant in early lunar history as I will suggest later in this paper.

Three arguments support the convection hypothesis. The first is that the figure of the Moon is more ellipsoidal than Laplace's argument gave. Early astronomers had attempted to verify the existence of the lunar bulge from accurate measurements of the apparent movements of features on the surface, due to the geometrical librations in latitude and longitude; but the results were inconclusive until the simple step was taken to separate points on the highlands and maria (Baldwin 1949). In a series of papers (Runcorn & Gray 1967; Runcorn & Shrubsall 1968; Runcorn & Hofmann 1973) it was established that points on the highlands, the irregular maria and the circular maria each fit ellipsoidal surfaces of nearly the same ellipticities with axes similarly aligned (the long axis towards the mean position of the Earth). These data are noisy because, with ground-based telescopes used, the observations were near the limit of seeing. But they gave correctly the different heights of these surfaces. The Apollo 15 and 16 radar altimeters gave of course much more accurate heights along their tracks (and showed the far-side was further from the centre of gravity than the near-side). These ground-based data remain, until measurements from a lunar polar orbiter are available, the best global data on the shape of the near-side of the Moon. The ellipticities measured were 2–3 times the dynamical ellipticities – those derived from the fractional differences in moments of inertia. This was explained (Runcorn 1967) by the convection hypothesis: the pressure differences associated with convection cause the outer boundary of the convecting region to be distorted but the density is less in the rising current than in the falling one. In a convecting spherical shell, the radial component of velocity and the pressure cannot both vanish on the spherical boundary.

The second argument is that at the time of flooding of the large impact basins, the nearly inviscid basaltic lava would have solidified to form a surface on which the gravitational potential plus a pseudo-potential allowing for rotation, analogous to the geopotential for the Earth, was constant. But, as we have shown, the present shape of the Moon is more ellipsoidal than it would be if it had resulted simply from the present gravitational field. Thus it follows that the second harmonics of the lunar gravitational field between 3.2 Ga and 3.8 Ga were different (when lavas flooded the basins) than now. Thus, the cause of the present non-hydrostatic second harmonic in the Moon's figure must be one that can have radically changed during the last 3 Ga. Solid state convection, being a phenomenon of instability, is therefore attractive.

The third argument is that palaeomagnetic evidence exists for reorientation of the Moon, between 4.2 Ga and 3.83 Ga ago, relative to its axis of rotation in space. These reorientations fundamentally involve flow in the lunar interior, so that the hydrostatic equatorial bulge remains perpendicular to the axis of rotation. This process is incompatible with the existence during that time of a non-hydrostatic bulge such as the Moon now has.

## 3. Laboratory studies of the magnetism of lunar rocks

As the Apollo samples were unoriented, the information they gave of the ambient magnetic field, by which they were magnetized, concerned its intensity (Collinson *et al.* 1974). Palaeointensity determinations are the most difficult area of the subject, for the remanent magnetization is a function not simply of the field strength at the time of magnetization, but of the physical and compositional nature of the particles carrying the magnetization and of the magnetizing process. It is a reasonable hypothesis that the lavas and high-grade breccias became magnetized by cooling *in situ* from above the Curie point, the magnetization being acquired from the ambient field only in the relatively short time the temperature remains just below the Curie point. This is called thermoremanent magnetization (TRM). There are three methods of determining palaeointensity from a rock sample.

(i) The Thellier–Thellier method consists of removing the remanent magnetization in steps of rising temperature in zero field and then, after demagnetization, in measuring the PTRM acquired by exposing it to a small laboratory field over different intervals of decreasing temperature. The ratios of NRM lost to PTRM acquired in each temperature interval equals, according to Néel's law, the ancient

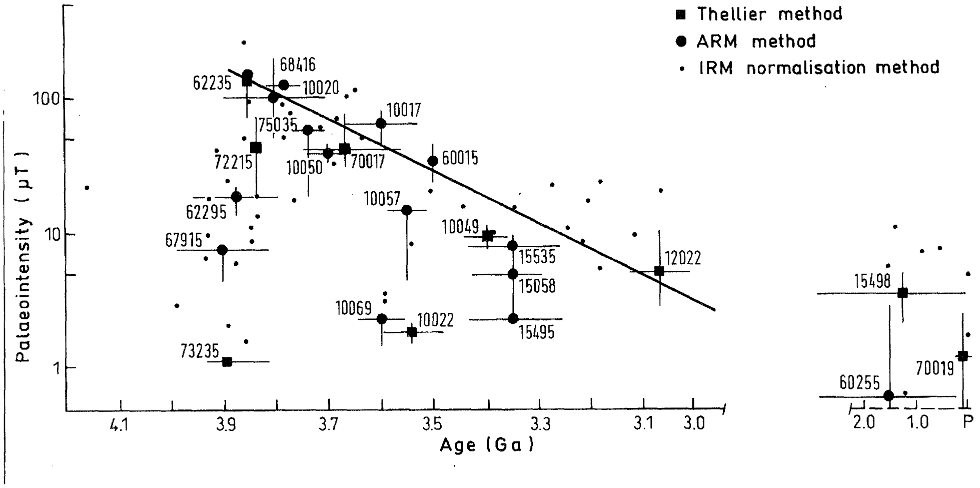


Figure 1. Palaeointensity of lunar magnetic field against age. Apollo sample numbers begin with 10, 12 15, 6, 7, denoting respectively Apollo missions 11, 12, 15, 16, 17. Standard deviations of palaeointensity measurements and radioactive age determinations shown.

field divided by the laboratory field. An important test is whether this ratio is independent of temperature: if so, particles with different blocking temperatures are giving the same information about the magnetizing field (Collinson *et al.* 1974).

(ii) An analogue, the anhysteritic remanent magnetization (ARM) method, has the advantage of not heating the sample, and thereby risking changes in the magnetic minerals (Stephenson & Collinson 1974): it uses the technique of alternating field demagnetization. The loss of the remanent magnetization, as the AC field is increased in steps, is similarly compared with the acquisition of an anhysteritic remanent magnetization in a small laboratory field, as the alternating strong ambient field is increased. Agreement between the two methods has been obtained on a small number of lunar samples of different ages. Many lunar samples, however, fail to give results which can be simply interpreted and much further investigation of palaeointensity methods is needed. The limited number of successful palaeointensity measurements, as was shown by Stephenson *et al.* (1975), lie near an exponential curve (figure 1). The field at 3.9 Ga was about 1 gauss ($10^{-4}$ T) and about 0.02 of this at 3.2 Ga.

(iii) In addition to these palaeointensity determinations, relative estimates of the lunar field can be made by normalizing the remanent magnetization intensity by dividing by a measure of the magnetic mineral content; the isothermal remanent magnetization (IRM) is used (Cisowski *et al.* 1983). These data, which are available for many Apollo samples, though scattered for they are estimates, supports the exponential decay of the field, but it shows (figure 1) a rapid rise in the palaeointensity in the 100–200 Ma before the maximum. These data also give support to the remarkable result that the lunar field reached a value of the order of that of the Earth's present surface field (Fuller 1987).

The palaeointensity data shows that the magnetizing field of the lunar rocks is a function of their age: a result which suggests a global field rather than local fields as the cause of magnetization (Collinson *et al.* 1977; Collinson 1993).

## 4. The magnetic surveys of the Moon

Thus the idea that the remanent magnetization of the lunar basalts and breccias had been acquired at the time of cooling from a global magnetic field, generated by dynamo action in a core, gained ground. The absence today of a global lunar field did not present a difficulty as a small body like the Moon loses its heat energy more quickly than a larger planet and the hypothesis that today its core is frozen, or is not convecting vigorously enough so that its magnetic Reynolds number exceeds the critical value for dynamo field generation, seemed reasonable. The high value of the field 3.9 Ga ago presents a difficulty if the field generated by a core depends only on parameters such as core radius and rotation, but while these enter into the non-dimensional numbers (the former into the magnetic Reynolds number) which must be exceeded for dynamo action, the intensity of the field may only depend on the energy available to drive convection. The fact that the palaeointensity diminishes exponentially from 3.9 Ga to 3.2 Ga seems to support this argument.

However, two serious difficulties remained. On one hand no other evidence for the existence of a core was then available. And, secondly there appeared to be little evidence for the global field supposed to have been responsible for the NRM of Apollo samples. In fact from the Explorer 35 results the anomalies had appeared to be evidence for entirely local magnetizations.

In the interpretation of these results, a simple theory is relevant. A slab of uniform magnetization $I$ and constant thickness $t$ extending to infinity has no magnetic field outside itself. But a finite slab, for example, a disc of radius $a$, produces a field of the order of $2\pi I$ at its edges and $2\pi It/a$ over its surface. It can easily be proved that at a height $h$, the field of a thin disc, if uniformly magnetized, is that of a dipole of strength $\pi a^2 tI$ placed at the lowest point of a sphere, centred at the observation point, circumscribing the disc. The returned Apollo samples gave magnetizations around $10^{-5}$ emu cc$^{-1}$ for the basaltic lavas and for the high-grade breccia around $10^{-4}$ emu cc$^{-1}$. Thus, the magnetic anomalies observed by magnetometer and other surveys fall into two categories: local anomalies, i.e., extending over 1 m–1 km of the order of $100\gamma$ and those detected by satellites, $1\gamma$ over 100 km. The former were seen in the Apollo surface traverses and are 'edge' effects and were not observed by the satellite magnetometers. This explains why the magnetized sheets of mare basalts do not show up in lunar magnetic maps. Most lunar scientists were familiar with the clear gravitational anomalies over the circular maria, called the 'mascons', which arise from the lava sheets filling the circular basins. They therefore found the lack of a similarly clear magnetic signal reinforced their scepticism of the existence of a former global lunar field.

The most important magnetic maps of the Moon were from the three component magnetometers in the subsatellite orbiters launched during the Apollo 15 and 16 missions (Coleman & Russell 1977). The Apollo 15 subsatellite orbited in a plane at 30° to the equator and a height of 100 km and the Apollo 16 subsatellite in a nearly equatorial plane, but in a decaying orbit – observing some anomalies at about 20 km. Small, $1\gamma$ (1 nT), anomalies observed by these satellite magnetometers extend over 100 km. In the initial modelling of these anomalies by Coleman & Russell, assuming dipole sources, it was found empirically that the dipole had to be placed 50 km below the surface of the Moon (where the Curie point would be exceeded). This result demonstrated that the surface strata producing the anomaly had to be uniformly magnetized over 100 km; later Hood *et al.* (1976) and Hood (1981) fitted the anomalies by surface discs determining their palaeomagnetic directions.

On the hypothesis that the magnetizing field of the rocks was generated by motions in an iron core, then liquid, it is reasonable to assume that the mean field was an axial dipole one. For even in a small core 500 km in radius and even with a period of rotation of 1 month, the Coriolis force is a dominant term in the magnetohydrodynamic equation of the core. In the Earth's case, this fact accounts for the average field being an axial dipole. The central problem in interpreting the lunar palaeomagnetic directions is to determine which strata are the source of the anomalies. Whereas in terrestrial palaeomagnetism, the samples measured are taken from rocks of known age and various field tests can be used to determine when the rocks received their primary remanent magnetization, here the inverse problem confronts us: determining the age of the palaeomagnetic direction. The palaeomagnetic directions from a terrestrial rock strata in one locality, and the poles calculated from them using the dipole formula (or VGP), show a scatter due essentially to the secular variation around the 'dipole' direction. So, poles calculated from rocks (which have not moved) that cluster about an axis is evidence that their magnetizations are of the same age. It was found empirically that the VGPs corresponding to the palaeomagnetic directions, determined as described above from the anomalies observed from Apollo 15 and 16 subsatellites fall into three bipolar groups defining three axes of rotation relative to the Moon (Runcorn 1982). I concluded that the poles calculated and assigned to one epoch, even though few compared with a typical set of results from a terrestrial geological stratum, must be regarded as very significant, because each is determined from the mean magnetization of a large volume of rock. Thus the fundamental problem is to date these three mean poles.

One strong magnetic anomaly (figure 3) correlates well with a light swirl-like feature Reiner Gamma on Oceanus Procellarum. It has been supposed to be an ash layer, but if so, by the shadows cast, its thickness is not more than 1 m. Were this enigmatic feature the source of the magnetic anomaly its intensity of magnetization would have to be orders of magnitude greater than the most intensely magnetized rocks. Nor can the anomaly arise from the magnetization of the thin sheets of lava in Oceanus Procellarum: their magnetizations are known and too small. As we have seen the sources must be rather strongly magnetized strata with variations of thickness and intensity on length-scales of a few 100 km. The geological maps of the Moon (Wilhelms 1987) show extensive sheets of ejecta from the great impact basins filling low areas. In western Oceanus Procellarum there are several anomalies (figure 3) the palaeomagnetic directions of which give grouped pole positions (Runcorn 1982): evidently the sources are the ejecta blankets from the Imbrian impact. It is inferred that the ejecta became magnetized after emplacement during cooling from above the Curie point, or by the precipitation of iron particles, in the global field at the time. Thus these poles (figure 2*c*) are Imbrian in age or 3.83 Ga according to Wilhelms (1987). By similar arguments the dating of the other pole positions can be made when an anomaly is clearly associated with the ejecta from other impact basins. Thus J, M are on ejecta from Nectaris while F and possibly G are on ejecta from Mendeleev. Thus the poles of figure 2*b* are dated as Lower Nectarian.

A clue to the physico-chemical process by which strong magnetization, uniform

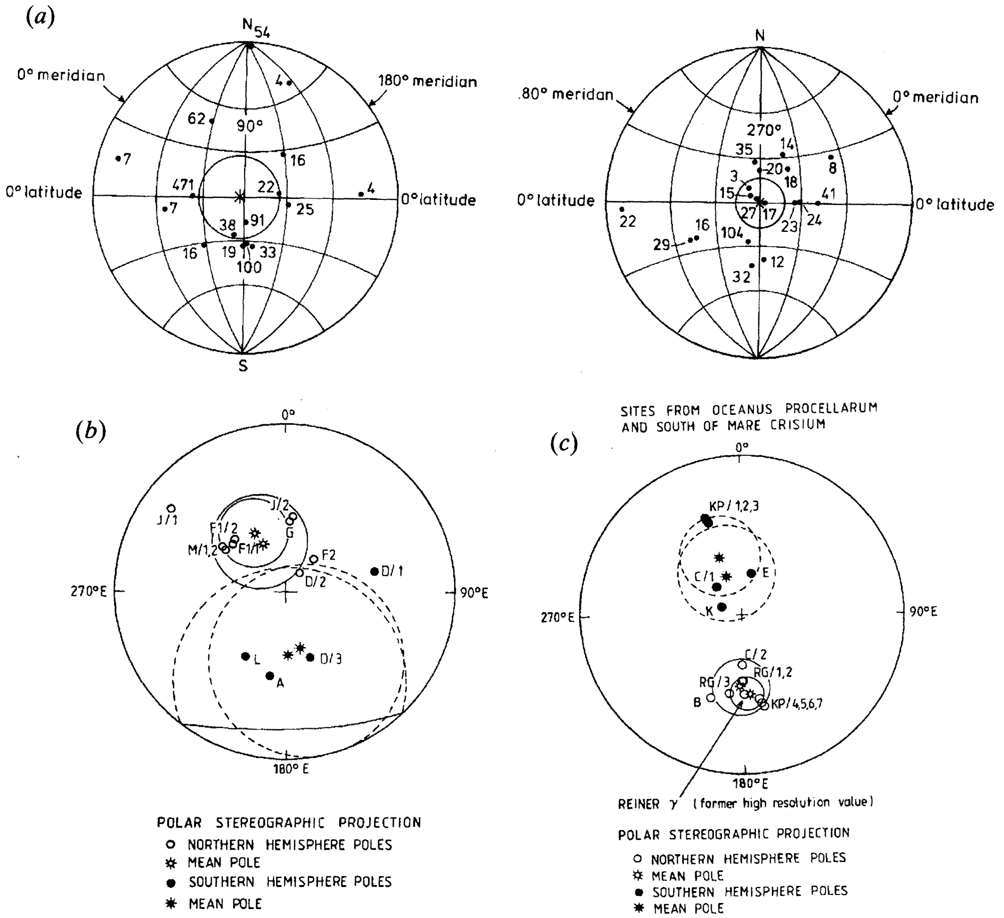


Figure 2. North magnetic pole positions for the three epochs: (*a*) Pre-Nectarian poles E and W, equatorial stereographic projections; (*b*) Lower Nectarian poles and (*c*) Upper Nectarian–Imbrian poles, polar stereographic projections. (Stars are mean poles with their 95% circles of conference.)

in direction over distances of 10–100 km, was acquired by lunar strata lies in the map of magnetic fields at the surface of the Moon obtained by the reflected electron techniques (Lin *et al.* 1976). This method gives information over the whole lunar surface, unlike the restricted coverage by the subsatellite magnetometers, but only of field intensity not of direction. There are a few areas of high intensity of magnetization of lunar material which because they reflect electrons must have uniformity over some distance probably of 10–100 km. How did these areas acquire considerably greater intensity of magnetization than other areas? Lin *et al.* (1980) showed that these strong fields were at the antipodal points of the final great impact basins on the Moon, Orientale, Imbrium, Serenitatis, Crisium Nectaris. It is well understood that the shock waves from each impact would have been brought to a focus at the point antipodal to the impact. Hood & Huang (1991) proposed that the solar wind magnetic field was momentarily amplified by a shock wave in the plasma and magnetized the area as a result of stress produced by the focusing of the shock, but there is a simpler interpretation. The energy released at the point antipodal to the impact will heat the rock. Supposing that the Moon's dipole field was present, and the palaeointensity data shows that it was at its maximum at the time of these great impacts 3.9–3.83 Ga ago, an extensive area of rock could have acquired a strong magnetization by one of two processes. If the rock was heated above the Curie point of iron, the magnetization would be acquired by the familiar process of TRM. However the magnetization might have been acquired by the precipitation of iron particles from the silicates, as a result of heating in a vacuum, and as the iron grains grow to dimensions greater than domain size, they would have become magnetized along the ambient field direction. This process, chemical remanent magnetization (CRM), is known to occur in terrestrial sediments and has a theoretical affinity with TRM. This explanation, unlike those which suppose that there was no general field of the Moon present, and rely on a momentarily amplified solar wind field, simply explains the uniformity in direction of the magnetization which must be present to give the anomalies. Moreover, large impacts occurred later, e.g., Copernicus, Tycho, and no intensely magnetized areas antipodal to them are observed. This is simply explained on the dynamo theory, for the dynamo had by then shut down, but there was still the solar wind present!

## 5. Pole positions and their interpretation

The three epochs for which palaeomagnetic observations are available are pre-Nectarian (Runcorn 1978), Lower Nectarian and Upper Nectarian–Lower Imbrian (Runcorn 1982, 1983, 1984). From the grouping of poles I divided the Nectarian era into lower and upper, but this seems geologically acceptable. The assignment of absolute dates – relative ages are not in dispute – are dependent on geologic identification of ejecta from the impact basins, especially Imbrium, Serenitatis and Nectaris at the Apollo sites. Wilhelms (1987) accepts the dating of Nectaris as 3.9 Ga while Stadermann *et al.* (1991) give 3.85 Ga. Dating of Imbrium seemed secure at 3.83 Ga, but the latter give 3.75 Ga, Wilhelms (1987) estimates the younger pre-Nectarian basins as about 4.0 Ga. The most significant conclusions from the pole positions shown in figure 2 are:

(i) The three palaeomagnetic mean pole positions form roughly a 90° spherical triangle on the Moon as shown in figure 3.

(ii) The corresponding palaeoequators place the impact basins of the same age in low latitude as shown in figure 3. The significance of this has been demonstrated by the use of Bingham's distribution, by which the lunar axis can be determined both from the palaeomagnetic poles and from the multi-ring basin distribution for each epoch. Figure 4 shows that the agreement is good.

(iii) For the multi-ring basins that show butterfly wing ejecta, the plane in which the impactor was travelling before collision can be determined, and when other asymmetry can be identified the direction can be found (Wilhelms 1987).

The inference drawn from (ii) is that satellites of the Moon rather than asteroids or comets in heliocentric orbit were the objects that hit the Moon to produce the multi-ring basins (Runcorn 1983). This conclusion is supported by the observed directions in which the impactors were travelling before they hit the Moon: the directions are parallel to the equator at the time as shown in figure 3 (Runcorn 1984).

Turning to observation (i), it shows that the Moon successively reoriented through about 90° with respect to its axis of rotation in space. This is exactly

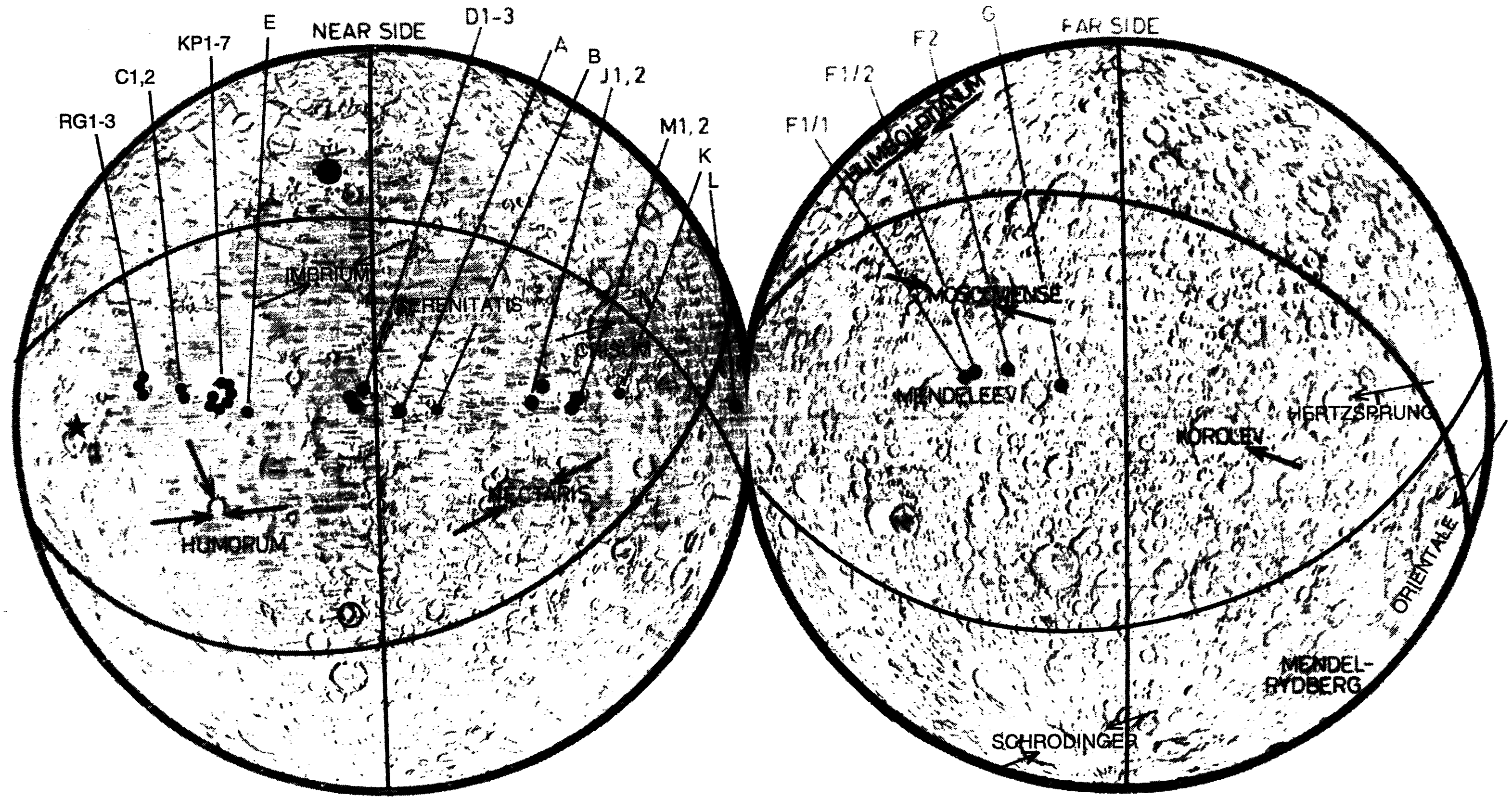


Figure 3. Poles and palaeoequators for Pre-Nectarian, Lower Nectarian and Upper Nectarian–Imbrian. Arrows show the directions, or alternate directions, from which the body came to produce the multi-ring basins. Palaeopoles are shown as follows: star, Pre-Nectarian; large black dot, Lower Nectarian; open circle, Upper Nectarian–Imbrian. Palaeoequators: Pre-Nectarian near prime meridian; Lower Nectarian crosses southern hemisphere on near side (Lower Nectarian basins: Nectaris, Mendel–Rydberg, Moscovience, Korelev, Mendeleev, Humorum). Upper Nectarian–Imbrian crosses north hemisphere on near side; Upper Nectarian–Imbrian basins: Crisium, Serenatatis, Hertzsprung, Imbrium, Schrödinger, Orientale. Small black dots, sources of Lower Nectarian and Upper Nectarian–Imbrian magnetic anomalies. Key as in Hood (1981); see figure 2*b* and 2*c*.

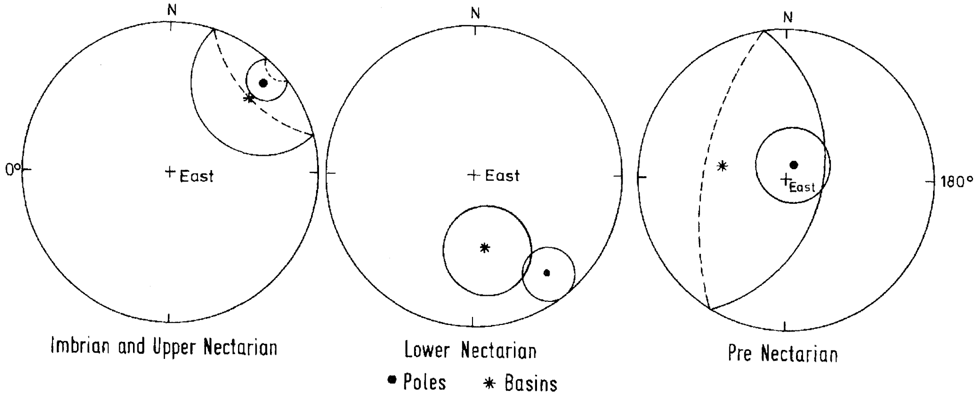


Figure 4. Lunar axes of rotation determined from palaeomagnetic data and multi-ring basin distribution for the three epochs (95% circles of conference).

what Euler's theory requires in the simple case when an impact occurs near the equator: material is excavated and distributed as ejecta over the Moon. Because much of this is highland material, even when isostatic equilibrium is restored by the rising of the denser mantle below, a crater remains and a negative contribution to the moment of inertia tensor results. The axis of maximum moment of inertia is displaced slightly towards the basin and, after the damping out of the Eulerian nutation, this axis becomes the new axis of rotation. The equatorial bulge, then, is no longer exactly at right angles to the axis of rotation, but will reorientate by flow in the interior of the Moon and by flexure of the 'rigid' lithosphere. The reorientation proceeds until the impact basin is at the pole of rotation. Because not one but a number of impacts occur in rapid succession, the reorientation of the Moon proceeds until the impact basins lie along a meridian with respect to the axis of rotation.

The observation (ii) that a series of impacts occurred near each palaeoequator before appreciable reorientation of the Moon took place requires that each group of impacts occurred in a short time compared with that between the epochs, about 100 Ma. Thus the ages of the multi-ring basins should show age gaps during which the Moon's axis reorientates.

A satellite would have broken up at or just within the Roche limit, which for the Moon is about one lunar radius away from its surface, and formed probably 6–12 smaller bodies; these would have been a suitable source for the impacting bodies. The rate at which a satellite is drawn into the Moon is proportional to its mass but inversely to a high power (6 or 7) of its distance to the Moon. Hence this explains why the impacts were closely grouped in time. At the break up the satellite orbits would have departed slightly from circular but on impact their angle of approach to the surface of the Moon would have been small, perhaps less than 5°, which would explain the asymmetries (butterfly wings) in the multi-ring basins. The scenario which explains (i) therefore supposes the existence of three lunar satellites: the first produced the late Pre-Nectarian basins which, after lunar reorientation, became meridional: the second a new equatorial set of basins, the lower Nectarian. Then the new axis of maximum moment of inertia was at the intersection of the two lines of basins and became the new axis of rotation.

## 6. The growth of the lunar core

The existence of a lunar core is now generally conceded. Although the Apollo seismic experiment did not yield conclusive evidence for one, electromagnetic, dynamical and geochemical observations are consistent with an iron or iron–sulphur core. Apollo magnetometer measurements of induction by the solar wind are consistent with an electrically conducting core of 400–500 km radius, though not necessarily metallic (Hobbs *et al.*1984, Russell *et al.*1981). Lunar laser ranging data reveal a slight departure from Cassini's third law, requiring a dissipative, possibly molten, dense core (Yoder 1981). Also the Moon is more depleted in siderophile elements than the Earth's mantle, arguing in favour of distributed free iron in the early Moon (Newsom & Taylor 1989).

The formation of the Moon's core has not hitherto been discussed. If the Moon came together as a molten mass as in one scenario, i.e., the impact of a Mars-sized body on an already differentiated Earth, then the iron core would have formed at the time of origin of the Moon. If, however, it came together as the coalescence of cold planetesimals in orbit around another body, their relative velocities thus being small, the accreted Moon would have been cold with iron bodies uniformly distributed throughout its volume. It is then not likely that the lunar core would have formed very early. For the Earth the gravitational energy released by the formation of the core from a body with a uniform mixture of silicates and iron is well known to be about twice that required to melt the whole Earth, supposing no losses. The corresponding energy released by the formation of the lunar core is about $10^{-4}$ of that for the Earth. Thus for the Earth, core formation was a run-away process, as radioactive heat released began to allow flow. In the Moon the differentiation of the anorthositic highlands, which was complete by 4.4 Ga, requires the postulate of a magma ocean of uncertain depth but likely to have been 200 km deep. There is no geochemical evidence requiring melting below this ocean until later. The radioactivity of U, Th and K distributed throughout the Moon at this early stage began to heat the Moon beneath the differentiated crust and at temperatures about half the melting point, solid state creep would have allowed iron to sink towards the centre. During this stage it is $U^{235}$, with a half-life of 710 Ma, which would have determined the time-scale of the formation of the core. That the core formed about 3.9 Ga ago and not 4.55 Ga is an entirely new idea but there are two arguments which lend it support.

Firstly the curve of palaeointensity of the Moon's magnetic field that has attracted attention, controversy and latterly support (Fuller 1987) showed that about 3.9 Ga the strength of the lunar field was comparable to that at the surface of the Earth today and diminished exponentially to about 0.02 of this value 3.2 Ga ago, subsequently to disappear. However, the data before this, mainly from the IRM method, which should give reliable relative values of the field intensity, shows a rapid rise of the field intensity from small values to this maximum over 100–200 Ma. If the core was forming over this time, there is a simple explanation for this sudden rise in palaeointensity. Secondly, the present second harmonic in the Moon's figure has been explained by a two-cell convection pattern, the development of which requires a core. But three arguments suggest that this particular convection pattern was not present early in lunar history. The shape of the surfaces through the maria clearly show the present distortion of the Moon was not present when the lava flooded the impact basins. The reorientations of the Moon with respect to its axis of rotation between 4.0 Ga and 3.83 Ga (or between 3.9 Ga and 3.75 Ga) following the impacts forming multi-ring basins near the equators would not have been possible unless the convection pattern was then different. A second harmonic in its gravity field stabilizes the axis of rotation relative to the Moon. The Apollo 15 and 16 radar altimeters showed that the far-side highland surface is further away from the lunar centre of gravity than the near-side: the centre of figure is displaced about 1 km away from the centre of gravity in the direction of the far-side. The inference drawn is that the highland crustal shell, less dense than the mantle, is thicker on the far-side. The lava which flooded the multi-ring basins on the near-side, 3.8–3.2 Ga rose through a height determined by the pressure in the magma source and its density relative to surrounding rock. This height was not great enough for the lava to extensively flood the far-side, thus explaining the striking hemispherical asymmetry of the Moon. The asymmetry in the thickness of the highland shell is indicative of the presence of a single-cell convection current in the early differentiation of the Moon.

In previous papers (Runcorn 1984), I have supposed that the Pre-Nectarian poles were dated at 4.2 Ga and the Lower Nectarian poles at 4.0 Ga, but from the above discussion by Wilhelms (1987) dates of 4.0 Ga and 3.9 Ga seem the earliest reasonable ones. The magnetic anomalies are then consistent with the interpretation of the palaeointensity curve. In planetary science; we have become familiar with the primary role of a planet's potential fields, its gravitational field and its magnetic field, in the development of knowledge of its interior. The Moon is unique among bodies of the solar system in possessing not only a record of its primeval magnetic field and its changes from extremely early times but also important information about early changes in the low harmonics of its gravitational field and shape.

## 7. Implications for formation of planets

These conclusions from the analysis of lunar palaeomagnetism may be a constraint on theories of the formation of the Moon providing observational evidence in a subject so far based on speculation. Firstly, because the pole positions, derived from the palaeomagnetic directions, can be explained on Euler's theory, this gives strong support to the interpretation of lunar magnetism by a core dynamo. The origin of the Moon must therefore involve core formation and evidence that this occurred late implies a cold origin for the Moon. Secondly, it solves the problem very early posed: where were the objects stored which hit the Moon in the final large bombardment, once termed the cataclysm? Were the Earth and Moon alone in space, the restricted three body problem would apply: around both there would be space in which satellites would be stable, apart from tidal friction. Thirdly, it suggests that there were satellites around the Earth for a time after its initial formation. Their life would have been shorter than the lunar satellites as the rate of decay of such orbits is proportional to the mass of the Earth. Could these satellites have been the source of the secondary addition of primitive solar system material, which Wänke concludes is necessary to explain the volatiles in the Earth? The composition of the primitive satellite therefore becomes a key issue. The physics of great impacts is not well understood but if the multi-ring basins were formed in this way there should be diagnostic evidence for impactors with low velocities (2 km $\mathrm{s}^{-1}$) rather than smaller bodies, comets or asteroids, with high velocities (15–20 km $\mathrm{s}^{-1}$). The slight asymmetry (butterfly wings) of the multi-ring basins cannot be explained by high-velocity impacts as the impactor would be buried and the energy released would have resulted in a circular basin. Also bodies in heliocentric orbit would not in general hit the Moon at low angles to the surface. Because the outer part of the Moon did not, subsequent to these impacts, change, except for the volcanism which covered the near-side with thin sheets of basalt, some component of the regolith must be material from the satellites. On the hypothesis that the Moon condensed from material flung out from the collision of the already differentiated Earth by a Mars-sized body, the primeval satellites would have been material left over and consequently of composition largely similar to the Earth's mantle. Were the satellites of primeval solar system composition, the origin of the Moon–Earth system would require rethinking. A study of the provenance of the regolith becomes a key problem of lunar science.

## References

Baldwin, R. B. 1949 *The face of the Moon*. University of Chicago Press.

Cisowski, S. M., Collinson, D. W., Runcorn, S. K., Stephenson, A. & Fuller, M. 1983 A review of lunar palaeointensity data and implications for the origin of lunar magnetism. *J. geophys. Res.* **88** (Suppl.) A691–A704.

Coleman Jr, P. J. & Russell, C. T. 1977 The remanent magnetic field of the Moon. *Phil. Trans. R. Soc. Lond.* A **285**, 489–506.

Collinson, D. W. 1993 Magnetism of the Moon: a lunar core dynamo or impact magnetization. *Surv. Geophys.* **14**, 89–118.

Collinson, D. W., Stephenson, A. & Runcorn, S. K. 1974 Lunar magnetic field palaeointensity determinations on Apollo 11, 16 and 17 rocks. *Geochim. cosmochim. Acta* Suppl. 5 **3**, 2859–2871.

Collinson, D. W., Stephenson, A. & Runcorn, S. K. 1977 Intensity and origin of the ancient lunar magnetic field. *Phil. Trans. R. Soc. Lond.* A **285**, 241–247.

Fuller, M. 1987 Lunar palaeomagnetism. *Geomagnetism*, vol. II (ed. J. A. Jacobs), pp. 307–448. Academic Press.

Helsley, C. E. 1970 Magnetic properties of lunar 10022, 10069, 10084 and 10085 samples 1970. *Proc. Apollo 11 Lunar Sci. Conf.* **3**, 2213–2219.

Hide, R. 1972 Comments on the Moon's magnetism. *The Moon* **4**, 39.

Hobbs, B. A., Hood, L. L., Herbert, F. & Sonett, C. P. 1984 Low frequency electromagnetic induction in the Moon: linearized inversion theory and lunar core calculations. *Geophys. J. R. astr. Soc.* **79**, 691–696.

Hood, L. L. 1981 Sources of lunar magnetic anomalies and their bulk directions of magnetization: additional evidence from Apollo orbital data. *Proc. Lunar Planet. Sci. Conf. 12th*, pp. 817–830.

Hood, L. L. & Huang, Z. 1991 Formation of magnetic anomalies antipodal to lunar impact basins: two dimensional model calculations. *J. geophys. Res.* **96** (Suppl. B6), 9837–9846.

Hood, L. L., Russell, C. T. & Coleman Jr, P. J. 1976 Evidence for a non-random magnetization of the Moon. *Geophys. Res. Lett.* **5**, 305–308.

Lin, R. P., Anderson, K. A., Bush, R., McGuire, R. E. & McCoy, J. E. 1976 Lunar surface remanent-magnetic fields detected by the electron reflection method. *Geochim. cosmochim. Acta* Suppl. 7 **3**, 2691–2103.

Lin, R. P., El-Baz, F., Hood, L. L., Runcorn, S. K. & Schultz, P. H. 1980 Magnetic anomalies antipodal to large impact basins. *Lunar planet. Sci. Abstr.* **XI**, 626–627.

Newsom, H. E. & Taylor, S. R. 1989 Geochemical implications of the formation of the Moon by a single giant impact. *Nature* **338**, 29–34.

Runcorn, S. K. 1962*a* Toward a theory of continental drift. *Nature* **193**, 311–314.

Runcorn, S. K. 1962*b* Convection in the Moon. *Nature* **195**, 1150–1151.

Runcorn, S. K. (ed.) 1962*c* *Continental drift*. Academic Press.

Runcorn, S. K. 1967 Convection in the Moon and the existence of a lunar core. *Proc. R. Soc. Lond.* A **296**, 270–284.

Runcorn, S. K. 1978 The origin of lunar paleomagnetism. *Nature* **275**, 430–432.

Runcorn, S. K. 1982 Primeval displacements of the lunar pole. *Phys. Earth Plant. Int.* **29**, 135–147.

Runcorn, S. K. 1983 Lunar magnetism, polar displacements and primeval satellites in the Earth Moon system. *Nature* **304**, 589–596.

Runcorn, S. K. 1984 The primeval axis of rotation of the Moon. *Phil. Trans. R. Soc. Lond.* A **313**, 77–83.

Runcorn, S. K. & Gray, B. M. 1967 The figure of the Moon. In *Mantles of the Earth and terrestrial planets* (ed. S. K. Runcorn), pp. 165–173. London: Interscience.

Runcorn, S. K. & Hofmann, S. 1973 The shape of the Moon. In *IAU Symp. No. 47* (ed. H. C. Urey & S. K. Runcorn), pp. 22–31. Dordrecht: Reidel.

Runcorn, S. K. & Shrubsall, M. H. 1968 The figure of the Moon. *Phys. Earth planet. Int.* **1**, 317–325.

Runcorn, S. K., Collinson, D. W., O'Reilly, W., Stephenson, A., Greenwood, N. N. & Battey, M. H. 1970*a* Magnetic properties of lunar samples. *Science* **167**, 695–697.

Runcorn, S. K., Collinson, D. W., O'Reilly, W., Battey, M. H., Stephenson, A., Jones, J. M., Manson, A. S. & Readman, P. W. 1970*b* Magnetic properties of Apollo 11 lunar samples. *Proc. Apollo 11 Lunar Sci. Conf.* **3**, 2369–2387.

Russell, C.T., Coleman, P. S. & Goldstein, B. E. 1981 Measurements of the lunar induced magnetic moment of the geomagnetic tail: evidence for a lunar core? *Proc. Lunar Planet Sci. Conf. 12th*, pp. 831–836.

Sonett, C. P. & Mihalov, J. D. 1972 Lunar fossil magnetism and perturbations of the solar wind. *J. geophys. Res.* **77**, 588–603.

Stadermann, F. J., Heusser, E., Jessberger, E. K., Linguer, S. & Stöffler, D. 1991 The case for a younger Imbrium basin; new $^{40}$Ar-$^{39}$Ar ages of Apollo 14 rocks. *Geochim. cosmochim. Acta* **55**, 2339–2349.

Stephenson, A. & Collinson, D. W. 1974 Lunar magnetic field palaeointensities determined by an anhysteretic magnetization method. *Earth planet Sci. Lett.* **23**, 220–228.

Stephenson, A., Collinson, D. W. & Runcorn, S. K. 1975 On changes in the intensity of the ancient lunar magnetic field. *Geochim. cosmochim. Acta* Supp. 6 **3**, 3049–3062.

Strangway, D. W., Larson, E. E. & Pearce, G. W. 1970 Magnetic studies of lunar samples – breccias and fines. *Proc. Apollo 11 Lunar Sci. Conf.* **3**, 2435–2451.

Wilhelms, D. E. 1987 The geologic history of the Moon. U.S. Geol. Survey Prof. paper 1348.

Yoder, C. F. 1981 The free librations of a dissipative Moon. *Phil. Trans. R. Soc. Lond.* A **303**, 327–338.

## *Discussion*

H. Fechtig (*Max-Planck-Institute for Nuclear Physics, Heidelberg, F.R.G.*). If Professor Runcorn's assumption of the disruption of three distinct lunar satellites is correct, one should see the cratering events in the crater statistics against time for the lunar surface. I do not remember that this has been reported.

S. K. Runcorn. I agree that relative ages of the multi-ring basis, as measured by crater counting on contemporaneous surfaces, should be grouped into three 'spikes', with time intervals between them. If these time gaps are present, they will be somewhat obscured by the uncertainties of each counting age. The data collected by G. Neukum (habilitation thesis, University of Munich, 1983) ought to be examined with this objective in view.

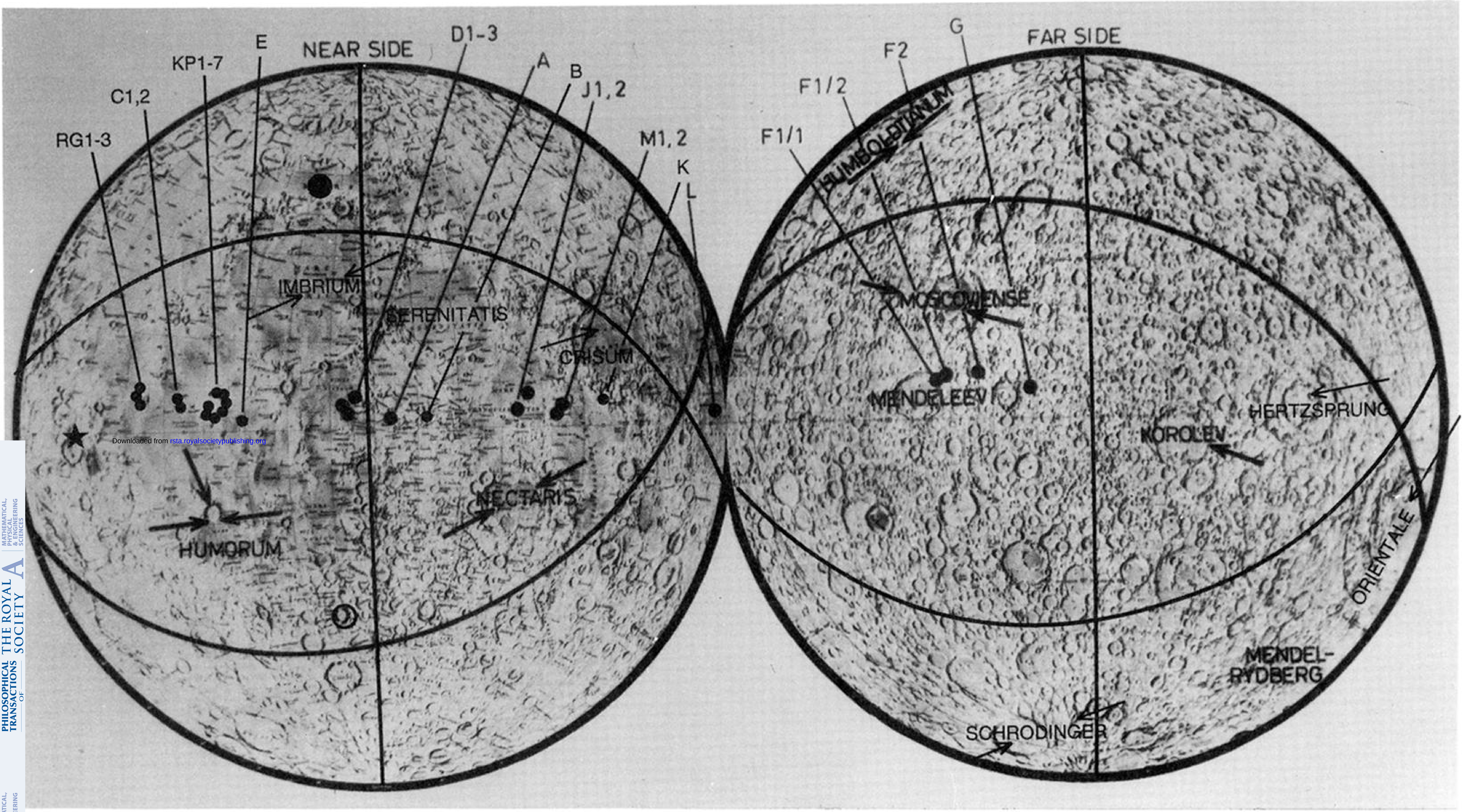
R. HIDE (*Oxford University, U.K.*). Determinations of the fossilized magnetic properties of magnetizable material in terrestrial rocks have been used to good effect to trace the history of the main geomagnetic field, which is produced by the self-exciting magnetohydrodynamic (MHD) dynamo mechanism operating in the liquid metallic core fo the Earth. Much more controversial have been attempts to interpret the observed magnetization of the Moon and other 'small' bodies (asteroids, meteorites) in the Solar System in terms of ancient dynamos and magnetic fields associated with the solar nebula, such as those of Professor Runcorn and Dr Collinson.

Having listented to these papers today and previously studied some of the extensive literature on the magnetization of small bodies in the Solar System, that has arisen in the past twenty years, I am struck by the reluctance of many workers in this field to take seriously the possibility that these bodies may have acquired their magnetization when they experienced hypervelocity impacts. A substantial amount of material in the colliding bodies would undergo a novel 'thermoremanent' remagnetizing process in any ambient magnetic field present. This material would be located far enough away from the point of impact to remain solid and unfragmented, but close enough to experience rapid heating to temperatures above the Curie point (but below the much higher melting point) followed by rapid cooling by expansion to temperatures below the Curie point, in response to the passage of a strong shock wave followed by a rarefaction wave. I discussed this process more than 20 years ago in the context of the magnetism of the Moon and meteorites, but its implications do not seem to appear to have been understood by critics of the idea that impacts may have played more than a negligible or subsidiary role.

The dominant contribution to the ambient magnetic field in which the material becomes remagnetized as its cools (by expansion!) through the Curie point on the very short time scale of interest is likely to come from strong electric currents generated by various processes, possibly even charge separation as in terrestrial lightning, in the mixture of dust and ionized material produced in the most intensely shocked region of all, close to the point of impact, and ejected into space at very high velocity. The detailed study of these processes will in my view be a necessary step towards the interpretation of the magnetism of 'small' bodies in the Solar System, just as much further research on MHD dynamos will be needed in efforts to interpret the detailed observations of magnetic fields of 'large' bodies in the Solar System such as the Earth, Jupiter, Saturn, Uranus and Neptune. These are important and exciting areas of physics and neither of them should be neglected as we plan future research of relevance to the study of the Solar System.

It is stated in the literature that lunar rocks show 'evidence' of having been exposed to magnetizing fields for a few hours at least, and that this 'evidence' rules out shock magnetization. I have not yet discovered whether the basis of the statement amounts to anything more than a calculation of cooling rates in which the possibility of cooling by expansion (following shock compression) is implicitly ignored.

S. K. Runcorn. The papers reporting the discovery of stable remanent magnetization of the lunar samples (*Proc. Apollo 11 Conf. 1970*) all mentioned meteorite impacts as a possible cause. Professor Hide's more detailed suggestion (1972) stimulated discussion, experiments and models: however, no tests of the idea against the varied observations of meteorite or lunar magnetism have resulted. By contrast, the hypothesis that the remanent magnetization of Apollo samples and lunar strata were acquired from a primeval core dynamo field has been tested against palaeointersity and palaeomagnetic direction data. In fact this theory is what Karl Popper calls a strong theory: it has many aspects which could be confronted by observation – and may be falsified. For instance if a full seismic experiment on a future lunar mission proved conclusively that the Moon does not have a core, the theory would be abandoned.



igure 3. Poles and palaeoequators for Pre-Nectarian, Lower Nectarian and Upper Nectarian–Imbrian. Arrows show the directions, or alternate irections, from which the body came to produce the multi-ring basins. Palaeopoles are shown as follows: star, Pre-Nectarian; large black dot, ower Nectarian; open circle, Upper Nectarian–Imbrian. Palaeoequators: Pre-Nectarian near prime meridian; Lower Nectarian crosses southern emisphere on near side (Lower Nectarian basins: Nectaris, Mendel–Rydberg, Moscovience, Korelev, Mendeleev, Humorum). Upper Nectarian–mbrian crosses north hemisphere on near side; Upper Nectarian–Imbrian basins: Crisium, Serenatatis, Hertzsprung, Imbrium, Schrödinger, rientale. Small black dots, sources of Lower Nectarian and Upper Nectarian–Imbrian magnetic anomalies. Key as in Hood (1981); see figure 2*b* nd 2*c*.