1986ormo.conf..609C

The Impact Theory for Origin of the Moon

A. G. W. CAMERON

Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138

A discussion is given of the essential ideas contained in the suggestion of W. R. Ward and the author that a very large impact on the proto-Earth involving a body with about the mass of Mars or greater led to the formation of an accretion disk about the proto-Earth, and that from the dissipation of this disk the Moon was formed.

The Angular Momentum Argument

In 1975 Hartmann and Davis (1975) noted that the spectrum of bodies bombarding the proto-Earth could have contained a number of quite large objects, having masses extending up to 10^{-1} Earth masses. They also noted that the collision of a large object with the proto-Earth would eject material into space, forming "a cloud of hot dust, rapidly depleted in volatiles." They then stated their expectation that the particles would interact and collapse into the equatorial plane where a satellite could form. Their paper suggested that an impactor of two lunar masses might be adequate, but contains an interesting footnote. When Hartmann first presented these ideas at a 1974 meeting, I objected that such a theory did not solve the angular momentum problem, and that this could only be done with a single collision involving a body having at least 10^{-1} Earth masses (i.e., about the mass of Mars). That idea was then under development by W. R. Ward and me and was presented the next year (Cameron and Ward, 1976). The two theories are often thought to be essentially the same, but in fact there is a profound difference in their approaches to the problem.

The question Ward and I started with was how big an object would have to be such that, striking the proto-Earth a glancing blow, it would impart to the proto-Earth and to any material that might be placed in orbit a total angular momentum equal to the angular momentum of the present Earth-Moon system. The answer was that the projectile should be about 0.1 Earth masses, or about the mass of Mars. Since the geometry of the collision was chosen in this question in an optimum way, the real answer to this question is that the *minimum* mass of the projectile should be about the mass of Mars, and that the actual mass could be substantially larger if the collisional impact was more centrally directed.

Note that we did not ask how much angular momentum should be involved in the collision in order to allow one Moon mass to be in orbit just outside the Earth's Roche lobe. This is only about one-sixth of the present angular momentum of the Earth-Moon system. In order that the Moon can have moved out by tidal interaction to its present orbit, the spin of the Earth must have had at least the other five-sixths of the angular momentum of the present Earth-Moon system. Thus the result required of the collision is not only that enough prelunar material must be placed in proto-Earth orbit, but the proto-Earth itself must be induced to spin rapidly. Since the angular momentum involved here is very large compared to typical planetary spin angular momenta (scaled for the mass of the Earth), this spin is much more likely to be induced in a single large collision than in a series of smaller ones in which the impact points are likely to be spread randomly over the face of the target.

This single collision with a major body became the central theme of the ideas developed by Ward and me. It satisfied the angular momentum problem by definition, and the major question that perhaps one should ask is whether it is plausible that such a major body should have been present in the early solar system. We were also very much concerned with the question of how any of the material that is ejected from the proto-Earth could have gotten into orbit, since all material ejected from the Earth on a trajectory leaving the surface must return to below the surface unless escape velocity has been achieved. We realized that at the collisional velocities involved, large quantities of rock vapor would be produced, and hence that, in addition to the usual gravitational forces in the problem, acceleration of material could take place by means of gas pressure gradients.

On the other hand, Hartmann and Davis started with what they regarded as a plausible planetary formation picture and tried to derive a lunar formation scenario from it. The Hartmann and Davis picture is consistent with ours only in the case where their collision involves a body of the same large mass. They also did not address the need for nongravitational forces in the problem of placing material into orbit. These two issues were the starting points of our own approach.

The Question of Plausibility

Is it plausible that an extra planetary body with about the mass of Mars or more should have been wandering around in the inner solar system at an appropriate time to have participated in our postulated collision? Let us recall that there are two main theories of planetary accumulation.

One of these theories, which has a majority of adherents, assumes that the evolution of the solar nebula reached a stage when any turbulence had died away, so that the small solid particles present in the gas of the nebula would have a chance to drift downward toward the midplane under the influence of gravity. When these particles have formed a rather thin disk, according to the theories developed by Safronov (1972) and by Goldreich and Ward (1973), the thin disk would become unstable against gravitational clumping, and would form objects with roughly the mass of asteroidal bodies. There then starts a process of hierarchical accumulation, in which a spectrum of masses develops in any region of the inner solar system, with the growth of some one body outdistancing the others. George Wetherill has been a leading investigator of this line of thought, and he stated at the Conference on the Origin of the Moon (from which this volume resulted) that it seems quite plausible to him that the second largest body could be 0.1 or more of the mass of the largest body. Thus there seems to be no reason to challenge the plausibility of our suggested impact theory of lunar formation within the framework of this general theory of planetary accumulation.

My own view of the planetary accumulation process (which seems to have rather few adherents) is that the history of planetary accumulation is a much richer and more complicated subject. We start with the earliest stages in the accumulation of the solar nebula, in which the infalling gas is cold, turbulent, clumpy, and only weakly gravitationally bound together. Under these conditions gravitational instabilities in the gas are likely to form giant gaseous protoplanets having something like the mass of Jupiter, plus or minus at least a factor of a few (DeCampli and Cameron, 1979). As the mass of the solar nebula increases, the local temperature rises rapidly, and the envelopes of the giant gaseous protoplanets are evaporated (Cameron *et al.*, 1982). However, remnants of planetary mass may remain behind if the condensed particles originally present in the gas of a protoplanet have a chance to precipitate downward to the center of the protoplanet, forming a molten body of very refractory materials. These protoplanetary remnants become the nuclei for further planetary accumulation, and there may be more of them than the present number of inner planets. The solar nebula becomes very hot as the sun is formed from its dissipation, and then it cools and turbulence dies away. As in the "standard" picture, asteroidal-sized bodies can then form, containing elements of medium volatility, and these are primarily collected by the planetary nuclei already formed to complete the planetary formation process. In this picture there are probably no masses within a few orders of magnitude of the planets themselves in the inner solar system, but of course there can be extra planets.

A body in an eccentric Earth-crossing orbit is likely to collide with the Earth after a time on the order of 10^8 years (comparable to typical cosmic ray exposure ages of iron meteorites). That number gives the age of the solar system at which I would expect extra planets to be eliminated by collision. Of course, it is also the time when a major collision would be expected on the "standard" picture. I would expect all of the inner planets, including extra ones, to be formed molten with immediate iron cores, and the accumulation of planetesimals to have provided enough of an atmosphere to have given the planets a relatively cool radiating surface, and thus to have kept them molten over the 10^8 year interval. I believe the proponents of the "standard" picture would believe the planets to have been formed warm but perhaps in general not molten, and maybe that is the major consequence of our different views on the scenario leading to a major collision with the proto-Earth.

Formation of the Accretion Disk

If the center of the impact-induced explosion is at or below the surface of the proto-Earth, then all material emerging from the site of this collision, if on a ballistic trajectory, would either escape from the proto-Earth or fall back onto the surface of the proto-Earth. Neither outcome would place material in orbit and hence no such material could participate in the formation of the Moon. The key to the resolution of this problem lies in the fact that much of the colliding material would be vaporized.

Ahrens and O'Keefe (1972) have discussed the shock vaporization of a variety of materials. The threshold for vaporization in different rocks is predicted by them to lie in the range 7 to 12 km/sec impact velocity, with most values in the range 9 to 11 km/sec. The threshold for iron appears to be a little higher. It is obvious that these thresholds will be significantly reduced if the impact occurs in already molten material, which I believe to be probable. It is expected that the collision of a Mars-sized projectile with the proto-Earth will occur at the escape velocity, 11 km/sec or higher, possibly as much as 14 to 15 km/sec, depending on the eccentricity of the projectile orbit.

Normally, when a small projectile hits a large target, a majority of the impact energy can be dissipated within a large volume in the target. In this case the impact velocity is substantially above the vaporization threshold and the bodies differ only by a factor of about two in radius. Therefore a great deal of vapor should be produced in the collision. The gas in the vapor cloud rising above the surface of the proto-Earth is subject to acceleration by pressure gradients as well as by gravity. This is the crucial difference that makes the formation of an accretion disk around the proto-Earth possible. For example, if the projectile hits the proto-Earth tangentially at 12 km/sec, and if the gas cloud were to receive an equal contribution of mass from the proto-Earth and from the projectile, then the gas would have a mean transverse velocity of 6 km/sec, which is slightly suborbital. The gas on the forward edge of the cloud then need only receive nearly 2 km/sec additional velocity in the direction of the collision due to pressure gradient acceleration to reach orbital velocity. It should also be clear that the expected yield of material in orbit is much less than half of the vaporized mass. It should also be clear that if the gas components derived from the projectile and the proto-Earth do not mix efficiently, then a considerable amount of the gas derived from the projectile may be lost from the system because it exceeds escape velocity, while most of that derived from the proto-Earth will fall back on the surface.

An estimate of the yield of material in proto-Earth orbit as a function of the collision conditions is thus highly desirable. I have recently made estimates of this yield using a particle-in-cell hydrodynamic code; the results were reported at the Conference on the Origin of the Moon and are published (Cameron, 1985).

The particle-in-cell method is a rather crude technique for determining the hydrodynamic behavior of a gas with an irregular three-dimensional geometry. The procedure can at best make semiquantitative predictions about the flow of gas in a system, and it is difficult to represent nonideal behavior in the equation of state of the gas. Nevertheless it is a reasonable method to use in the first investigation of a problem that exhibits relatively little symmetry. The available space is subdivided into cells, and there should be enough particles present so that in the active region of the problem there are several particles in each cell.

In any one time step of a problem, the positions of the particles are advanced with the particles traversing ballistic orbits subject to acceleration due to the gravitational field. At the end of a time step the position and motion of the center of mass in each cell is computed, and the energy associated with random motions with respect to the center of mass is extracted from the cell. The lowest energy state of a cell involves rigid rotation around the center of the mass. The extracted energy from a cell is then fed back into the particles randomly but in a way that preserves the angular momentum. Thus at the end of a time step the mean motion of the fluid in each cell is preserved but the motions of the particles are randomized.

Imagine two adjacent cells sharing a common boundary but with different numbers of particles. The randomization of the motions of the particles will then send more particles across the common boundary from the rich cell to the poor cell than will be sent in the reverse direction. In this way the scheme acts to accelerate mass down pressure gradients.

This scheme was used to predict the behavior of the colliding system for a wide variety of initial conditions. The velocity of impact was varied from 11 to 14 km/sec. The initial state of the shocked vapor was varied from "hot" (odd numbered particles given the velocity of the proto-Earth and even numbered particles given the velocity of the projectile) to "cold" (all particles given the same mean velocity). The source region of the gas was varied from just the region of overlapping hemispheres of the colliding bodies to essentially the entire volume of the projectile, including the overlapping region. The prediction of interest was the amount and angular momentum of the gas that settled into orbit about the proto-Earth, neither escaping from the system nor falling back onto the surface of the proto-Earth. Because the disk that is formed is subject to dissipation, with mass flowing both inward toward the proto-Earth and outward toward the Roche lobe, a "successful" disk formation was deemed to be one in which the mass in orbit was at least twice that of the Moon and the angular momentum was at least twice that of a lunar mass in orbit just beyond the Roche lobe.

The results showed that these conditions can be achieved under a wide variety of conditions. For the best yields the gas should be neither too "hot" nor too "cold," the collisional velocity of the projectile should be a few km/sec greater than the escape velocity minimum, and most of the volume of the projectile should be part of the source region for the vapor that is produced. In all cases the majority of the mass that goes into orbit following the collision is derived from the projectile.

These initial conditions are optimally met for a projectile in orbit about the sun in a proto-Earth-crossing trajectory of significant eccentricity. This is consistent with a time scale for the collision on the order of 10^8 years after the formation of the solar nebula.

The collision should lead to the loss of the original atmosphere of the proto-Earth (Cameron, 1983). Judging from the fact that Venus has an atmospheric content of rare gases that is very large compared to the Earth, one can conclude that the loss of the atmosphere occurred after the bulk of the accumulation of volatile-containing planetesimals had occurred. This is another argument in favor of the collision happening about 10^8 years after formation of the solar nebula.

Dissipation of the Accretion Disk

Once the accretion disk has been formed, it is subject to severe dissipation. The column density of the disk is great enough that self-gravitation is important, and there is instability against local clumping due to gravitational instabilities (Ward and Cameron, 1978). However, inside the Roche limit, the gravity of the proto-Earth is strong enough to shear apart any such clumping that occurs. This dissipates a lot of energy and hence leads to a spreading inward and outward of the mass in the disk. To the extent that the temporary mass fluctuations can raise small tides in the proto-Earth, there can also be a small amount of angular momentum transfer from the spin of the proto-Earth to the orbital motion of the disk material.

Ward and I estimated the dissipation time of the disk to be months to years. However, Thompson and Stevenson (1983) pointed out that such a rapid dissipation would release energy at a sufficient rate to completely volatilize the material in the disk. This would obviously slow the dissipation rate to a value that would allow some of the material to remain in condensed form. Thompson and Stevenson also pointed out that in a mixture of the gaseous and condensed phases of a substance ("froth"), sound speed is greatly reduced relative to the value in either phase alone. This enhances the gravitational instability. They estimated the dissipation time of the disk to be on the order of a century.

So far there has not been a careful examination of what happens at the outer edge of the disk as it recedes past the Roche limit. One possibility is that a gravitational instability occurs that forms a body of not very great mass, and that this continues to accrete matter as it emerges or is perturbed past the Roche limit. Another possibility is that a small body forms, as before, and then recedes from the proto-Earth because of tidal interaction with both the proto-Earth and the accretion disk. Then another body can form and recede in its turn, and so on. If the last body to form in this way were to be the most massive, then it would recede fastest and proceed to sweep up its predecessors. Otherwise this last scenario might be in trouble.

Discussion

The general theory outlined above requires much more detailed examination to expose its strengths and weaknesses and to fill in the gaps.

The iron core of the projectile is probably not significantly vaporized. Assuming, as stated above, that projectile is a differentiated planetary body, then the iron core should have concentrated and extracted the siderophile elements from throughout the body of the projectile. Hence the accretion disk and the Moon should be depleted in these elements, as is observed to be the case.

The deposition of mass into the proto-Earth by the collision should release a great deal of energy, which I have estimated to raise its surface temperature to about 6000 K (Cameron, 1983). Dissipation within the accretion disk probably maintains the temperature near 2000 K in the vicinity of the Roche lobe. Hence when the Moon forms it should do so in a temperature field that will have prevented most of the elements of medium volatility from condensing with any significant abundances. Again, this is observed to be the case.

It should be noted that the accretion disk need not form precisely in the equatorial plane of the proto-Earth initially, nor should one expect the orbit of the Moon to lie in that plane today. The accretion-disk-forming collision adds a very large angular momentum vector to the proto-Earth, which will nevertheless add to the previous angular momentum vector. Thus the initial plane of the disk will differ from that of the equator. There will be a large equatorial bulge, and there will thus be a tendence to align the two angular momentum vectors. The rate at which this happens needs to be investigated. It is likely that a great deal of matter escapes from the proto-Earth system after the collision (so my simulations have generally indicated), but this matter will go into independent Earth-crossing orbits and will before too long probably recollide with the proto-Earth, further disturbing the angular momentum vector of the proto-Earth.

It has generally been assumed that the equality of the oxygen isotope ratios in the Earth and the Moon indicates that the Moon was formed out of material that condensed at the Earth's orbital distance from the sun. Until the origin of the variations in oxygen isotope ratios is properly understood, this principle will remain empirical, but it seems reasonable to accept it provisionally. It seems to me that this requirement is plausible with either of the two models of planet accumulation.

There are a number of other aspects of lunar composition, such as the iron oxide content of its mantle, that were discussed at the Conference on the Origin of the Moon. In particular, Ringwood felt that if the bulk of the Moon could not be made out of terrestrial material, then the projectile should be considerably larger than Mars to meet chemical restraints. I see no problems with that; the calculations that I did were for a Mars-sized body, but this is only a lower limit, and the projectile could perhaps be as large as one-third of the mass of the Earth without getting into trouble with the dynamics.

Acknowledgments. This research was supported in part by the National Aeronautics and Space Administration under grant NCR 22-007-269.

References

Ahrens T. J. and O'Keefe J. D. (1972) Shock melting and vaporization of lunar rocks and minerals. *The Moon, 4*, 214–219.

Cameron A. G. W. (1983) Origin of the atmospheres of the terrestrial planets. *Icarus, 56*, 195–201.

Cameron A. G. W. (1985) Formation of the prelunar accretion disk. *Icarus, 62*, 319–327.

Cameron A. G. W. and Ward W. R. (1976) The origin of the Moon (abstract). In *Lunar Science VII*, pp. 120–122. The Lunar Science Institute, Houston.

Cameron A. G. W., DeCampli W. M., and Bodenheimer P. H. (1982) Evolution of giant gaseous protoplanets embedded in the primitive solar nebula. *Icarus, 49*, 298–312.

DeCampli W. M. and Cameron A. G. W. (1979) Structure and evolution of isolated giant gaseous protoplanets. *Icarus, 38*, 367–391.

Goldreich P. and Ward W. R. (1973) The formation of planetesimals. *Astrophys. J., 183*, 1051–1061.

Hartmann W. K. and Davis D. R. (1975) Satellite-sized planetesimals and lunar origin. *Icarus, 24*, 504–515.

Safronov V. S. (1972) *Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets*, translated from the Russian. Israel Program for Scientific Translations, Tel Aviv.

Thompson A. C. and Stevenson D. J. (1983) Two-phase gravitational instabilities in thin disks with application to the origin of the Moon (abstract). In *Lunar and Planetary Science XIV*, pp. 787–788. Lunar and Planetary Institute, Houston.

Ward W. R. and Cameron A. G. W. (1978) Disc evolution within the Roche limit (abstract). In *Lunar and Planetary Science IX*, pp. 1205–1207. Lunar and Planetary Institute, Houston.