## THE FORMATION OF THE PLANETS,\* PART I

## By GERARD P. KUIPER

## Formulation of the Problem

The problem of the origin of the earth and its sister planets has been of great philosophic interest throughout the ages. It is not a foregone conclusion, however, that the problem has a scientific solution. For instance, an enclosure in which the air has been stirred gives, after some delay, no clue on the nature or the time of the stirring. All memory of the event within the system has been lost.

The solar system is not infinitely old, however, and shows many properties that must depend directly or indirectly on its mode of formation. Its age is, in round figures, 5000 million years, based on analyses of terrestrial and meteoritic materials containing natural radioactive isotopes and their decay products-though some uncertainties remain in their interpretation. The age of the sun cannot at present be accurately determined. Ledoux (1949) computed the value  $5 \times 10^9$  years; but the adopted model was found subsequently to require revision; an appreciable outer convection zone appears to be present while there is still uncertainty about the presence of a convective core (Strömgren, 1953). For these and other reasons it is at present unsafe to estimate the age of the sun directly (Strömgren, 1955). However, it is impossible to conceive that the sun should be younger than the planets, i.e. less than  $5 \times 10^9$ years. Since further, the age of the globular clusters, which is almost certainly greater than the age of the sun, has from the rate of the evolutionary processes been estimated to be about  $5 \times 10^9$  years (Sandage, 1953) or  $6 \times 10^9$  years (Hoyle and Schwarzschild, 1955), it nevertheless follows that the age of the sun is close to  $5 \times 10^9$  years, and hence essentially the same as that of the planets. This compels one to inquire whether the formation of the planets and of the sun cannot have resulted as parts of the same basic process, which must be that of stellar formation in general.

Stellar formation is the process whereby stars form from pre-existing nebular clouds. This process is not merely hypothetical; it appears to occur before our eyes in several nebular regions. The initial dimensions of these clouds must have been nearly as large as the distances between stars. Therefore, an enormous contraction, by about 10<sup>7</sup> times, had to pre-

\*Lecture delivered at a joint meeting of the Royal Canadian Institute and the Royal Astronomical Society of Canada, November 12, 1955.

57

R.A.S.C. JOUR., Vol. 50, No. 2.

58

cede the formation of a stable star. Velocity measures within galactic nebulae as well as dynamical considerations indicate that the pre-stellar clouds will, in general, have had some internal motions. On probability grounds, these internal motions will, in general, not have resulted in zero total angular momentum for a given cloud. Because angular momentum is conserved, contraction will often have been arrested long before the mass attained stellar dimensions. Instead of a single star, a double or multiple star may thus have resulted (cf. Kuiper, 1955). Binary and multiple stars are indeed very common; more than half the stars of the galaxy appear to be members of double or multiple systems. It is therefore suspected that the formation of the planetary system was related to this common process of binary-star formation. Parenthetically, it was this thought that led the writer in the early 1930's, after some frustrated attempts at understanding the origin of the planetary system, to the study of binary-star statistics.

With the age of the planets defined and the general mode of origin suggested, one must next examine whether the planetary system, just after its formation, was physically and dynamically essentially the same as today; or whether important subsequent evolution took place. In other words, there must have been a *formative* period during which the sun and the solar system segregated, and planet-like bodies first developed, followed by what may be called the *geologic* period, which began at the epoch  $T_g$ , some 4–5 billion years ago. Before we can clearly see what a theory of origin must contain, we must get a picture of the evolutionary changes which the planets and their orbits underwent during geologic time.

## Planetary Evolution During Geologic Time

There are several evolutionary effects whose importance we shall examine one by one.

Solar radiation effects. The physical condition on a planetary surface depends in part upon its equilibrium temperature, i.e. on the planet's mean distance from the sun, a, and the solar luminosity, L. If the sun had ever gone through a nova stage—as has sometimes been supposed—drastic changes on the planets would indeed have occurred. Solar evolution is now known well enough to rule out this possibility. After passing through a comparatively rapid contraction phase (Helmholtz-Kelvin time scale) the sun has changed very little. In my earlier papers I used the round figure  $10^8$  years for the duration of this contraction phase. It has now been computed more carefully by Strömgren (1955) who has found  $8 \times 10^7$  years. The last  $3 \times 10^7$  years of this interval have been studied

independently and in considerable detail by Henyey et al. (1955), who published computed evolutionary tracks. The maximum brightness, occurring toward the close of the contraction phase, was found to be almost identical to the present solar luminosity; while the brightness at the very end of the contraction phase was about 20 per cent. lower than at present, this difference having been traversed slowly and gradually during the entire geologic era (Henyey, 1955). This very slow increase is still continuing.

At present, with an average albedo of 0.36 (Danjon, 1954; Byers, 1954), the earth has a mean radiation temperature\* of 250°K. (-23°C.). Below we find that at  $T_g$  the earth had probably essentially the same mean distance from the sun as today. If, further, the albedo were also the same, the mean radiation temperature would have been 5 per cent. or 12°C. lower, or 238°K. The present mean surface temperature is some 30°C. higher than the radiation temperature because of the "greenhouse" effect of the atmosphere; the chief constituents responsible for this effect are water clouds, water vapour, ozone, and carbon dioxide. Some of these factors may have been different at  $T_g$ : ozone may have been nearly absent and water vapour scarce because of the lower temperature. These effects plus the lower radiation temperature might have caused a lower surface temperature; the polar caps may therefore have been larger than they are now, and the albedo higher than 0.36, which would cause a further reduction in surface temperature. If, on the other hand, water was initially very scarce, as on Mars today, then a somewhat lower albedo and a higher mean surface temperature, approaching 0°C., may have existed. At any rate, the surface temperature on the earth in the early Pre-Cambrian was probably somewhat less than it is today. The picture is, however, made more complex by the occurrence of glacial periods even quite recently, indicating a climatological instability in the difference between the radiation temperature and the mean surface temperature (cf. Brooks, 1951).

More significant than the planetary surface temperatures are, for our purpose, the temperatures of the planetary exospheres; they determine the planetary mass losses. Because of the indicated solar history, the evaporation rates at  $T_g$  will have been somewhat less than at present. Since, astronomically speaking, the rates are negligible now, present planetary *masses* (and compositions) will apply at  $T_g$ .

Internal effects. While the planetary masses and global compositions have not changed since  $T_g$  (with the understanding that  $T_g$  may not be quite

\*The radiation temperature is a measure of the total amount of radiation lost to space; it balances the absorbed solar radiation.

the same for each planet), internal rearrangements must have taken place. The best known of these are the changes in the earth's crust and atmo-hydrosphere, although opinions differ on the extent to which these changes are cyclic or progressive. In other words, have the continents grown from nuclei during geologic time or have they always been roughly of present size and have been merely reworked? The arguments for these respective views are well presented in two recent books (Wilson, 1954; Poldervaart, 1955). There is also uncertainty on the origin of the terrestrial hydrosphere. While the evidence of the whole appears to favour its development from the earth's interior during geologic time, as is true for the atmosphere (Rubey, 1955, and references given there), there is also the result that the water vapour now exhaled by volcanoes is not original (juvenile) but all or nearly all recirculated (Craig and Boato, 1955).

Evidence for other internal rearrangements comes from lunar studies (Kuiper, 1954, and references given there) and from well-known studies of meteorites, fragments of the asteroid ring. These bodies show that widespread melting has occurred which is attributed at least in part to heating by their radioactive constituents (U<sup>235</sup>, U<sup>238</sup>, Th, K<sup>40</sup>, etc.). For the larger planets considerable heating derived also from the potential energy released upon the formation of a high-density core (Urey, 1952). Internal thermal evolution must have affected all planets and all but the smallest satellites (where thermal conductivity prevented accumulation of heat); and one might, of course, regard the development of continents and an atmo-hydrosphere as accompanying features of such a thermal evolution. The atmospheres of Venus and Mars may thus be regarded as evidence for such evolution. The same is probably true for the presence of a planetary magnetic field. Since all planetary interiors presumably are, or at least have been, above the Curie point of iron (about 750°C.), a planetary magnetic field is supposed to require the presence of a liquid core. Thus the moon will probably have no such field while Venus and the Jovian planets probably do. (A field for Venus has been suspected empirically by Houtgast, 1955, but this is not established.) The structure of the Jovian planets must likewise have changed during geologic time. The major outbursts observed on both Jupiter and Saturn every decade or so-unaffected by solar radiation, occurring as they do at random positions along the eccentric orbits-show that the interiors still furnish significant quantities of heat to the atmosphere. This is consistent with the large initial heat content expected for massive planets (e.g. Kuiper, 1952, p. 326). Consequently, early in geologic time these planets must have differed visibly from the present. For instance, the surface temperature of Jupiter may well have been around 500°K., and at least its present type of cloud cover will have been absent. Yet, even these changes do not appear to affect our main problem, that of planetary formation. While the radius of Jupiter will have been slightly larger owing to the higher internal temperature, and hence the velocity of escape slightly less, the solar luminosity was some 20 per cent. less and escape by evaporation would have been no greater than today; that means, it was entirely negligible.

One problem of rearrangement is, however, directly related to our main theme. Both Jupiter and Saturn now possess a greater angular velocity at the equator than elsewhere. In the case of Jupiter only the bright equatorial zone itself stands out, having a period of rotation 5 min., or nearly 1 per cent., less than the remainder of the planet (minor differences exist between the other zones). In Saturn the period of rotation appears to be a smooth function of latitude, being 11 per cent. greater at 57° than at the equator (Moore, 1939). No convincing explanation of this "equatorial acceleration" has been given on the basis of a global circulation system. Recently the writer found that the explanation may be similar to that proposed by Cowling (1953) for the equatorial acceleration of the sun, namely that a fossil gaseous ring is involved, a remnant of the disk surrounding the body at the time of planetary formation. The time scale of equalization of angular velocity of adjacent gaseous rings by viscosity is  $\rho l^2/\mu$ , in which  $\rho$  is the density, l the dimension and  $\mu$  the coefficient of viscosity. Now  $\rho$  is probably between  $10^{-3}$  and  $10^{-1}$ , l may be limited by the depth of the atmosphere rather than the width of the zone, and be therefore more nearly 108 cm. than 109 cm., while the coefficient of molecular viscosity is about 10-4. The computed time scale is then  $10^{17}$ – $10^{19}$  sec. or  $10^{9\frac{1}{2}}$ – $10^{11\frac{1}{2}}$  years; since the deeper atmospheric layers are the more massive, the higher value is probably more representative than the lower. However, this is an upper limit, since eddy viscosity would be expected to play a role. It is therefore not completely certain that the proposed explanation is correct; but even if the eddy coefficient should be as much as 103 times the molecular value, the explanation seems acceptable. If the eddy viscosity were high it would follow that the initial velocity differences were much larger than they are now, which in fact would be expected for a collapsed ring. While this phenomenon is interesting in its implications, it does, however, not change the total mass or angular momentum of the planets involved.

Finally, a word must be said about the concept of *polar wandering*, as it affects our discussion. Glacial deposits in widely scattered areas of the world, including the tropics, coal deposits near the North Pole, and the

The distribution of continents over the globe may in time give information of more direct astronomic interest. As is well known, this distribution is very uneven, and the analysis of it in terms of spherical harmonics has been used by Vening Meinesz (1951, 1952, 1953) to derive information on the convective regime in the mantle, on the assumption that continents were formed by such convection. (In principle, this could have happened early in the earth's history, when the mantle was perhaps fluid.) He found that a good representation of the mean coefficients of the terms in the second to the tenth spherical harmonics followed from convection, in a mantle of present dimensions; but that the existence of the large first harmonic required a core of less than  $0.18R_E$ . Chandrasekhar (1952) confirmed the numerical conclusions from a rigorous analysis and Urey (1953) found support from these discussions for his hypothesis that the core formed gradually during geologic time. It is this aspect—when did the earth melt, if ever-that bears on the problems of origin. This question has not yet been clarified; but from discussions with Dr. Chandrasekhar it is my understanding that the introduction of rotation into this problem may cause important changes. Now the earth must have had a rapid rotation early in its history, when the moon was much closer and most of the angular momentum of the earth-moon system resided in the rotating earth. Chandrasekhar is currently making a new analysis of the problem, taking into account the rotation.

The moon does not appear to be elastico-viscous, contrary to the earth, because for at least  $10^9$  years it has retained a shape that differs from equilibrium. This perhaps implies that the interior of the moon has cooled substantially below the melting point. Its rotational motion is not free and stability requirements may have caused it to topple over early in its history (Kuiper, 1954, p. 1099). The envelopes of Jupiter and Saturn are so large that they might not have followed even if the interior solid part had turned with respect to the axis of rotation. Such, at least, is suggested by the rotational regime of these envelopes, discussed above.

Solar Dynamical Effects. If during geologic time the sun had either lost or gained mass by a certain factor, the scale of the planetary orbits would have been changed by the same factor, but in the opposite sense. That is, a one-per-cent. mass increase would have caused a to decrease one per cent. Incidentally, the solar luminosity would by this change have increased about three per cent. so that at the planet the sun would have appeared 5 per cent. brighter; the planetary radiation temperature would have increased 1.2 per cent., or 3°C. for the earth.

There is no clear evidence, however, that the sun has been gaining mass. It is often assumed that at least the zodiacal light (caused in part by reflection of sunlight on dust particles of cometary origin and in part by scattering by electrons of solar origin) is evidence for accretion, since the cometary particles spiral inward under the Poynting-Robertson effect. However, before they merge with the sun itself they reach a zone where they will evaporate. It is readily shown that they remain in this zone much longer than is required for complete evaporation. A large particle remains in this zone longer than a small one, proportionally to the particle diameter; but evaporation, likewise, takes proportionally longer. Since the evaporation can be shown to take place for any chosen size, it happens for all sizes. The vapours become rapidly dissociated and ionized, whereupon solar protons and electrons sweep them out of the solar system, as is true for comet tails. The enormous display of the Comet of 1882 shows what happens to cold matter penetrating the solar corona! A similar fate would result for approaching gas. Interplanetary magnetic fields will also inhibit accretion of interstellar gas (Opik, 1954). Thus, solar accretion must be very nearly zero.

Solar mass losses are probably more important, though still very small. The electron density near the earth appears to be about 500 per cm.<sup>3</sup>

(comet tails, zodiacal light, radio whistlers) and the proton density must be similar. The mean outward velocity of these particles is perhaps 300 km./sec. If the solar emission were isotropic, this would, if at a uniform rate, entail a solar mass loss of 0.5 per cent. during geologic time; but since the solar emission depends on latitude, as is indicated e.g. by the isophotes of the polarized component of the zodiacal light (Elsässer, 1953), the loss will be perhaps 0.1 per cent. A still lower value, 0.001 per cent., corresponds to the rate estimated by van de Hulst in 1950 from the appearance of the corona.

Planetary Dynamical Effects. The dynamical properties of the planetary system, evolving under the multiple internal forces resulting from Newton's law of gravitation, can be found with only limited accuracy for the past, the accuracy decreasing as the time interval increases. If the computations were made by numerical integration of the equations of motion, the limit could be made to depend only on the precision of the starting values, which must be found empirically and can never be exact. For very long intervals of time, however, even machine computation would be impossibly slow and costly. Analytical methods must then be used which study the variations of the orbital elements by a series of approximations. These approximations may be said to arise as follows: first consider all planets but one to move in undisturbed Kepler orbits, with the one perturbed by the others which are now located at known positions. Each planet is so treated in turn, whereupon the second approximation is found by considering the perturbing planets at their new, perturbed, positions; and so on. The actual series developments for the elements of a planet contain terms in the first, second, third, . . . powers of the masses of other planets,\* with coefficients containing, among others, products of infinite series in both the orbital eccentricities and the mutual inclinations. Large numbers of periodic terms of significant amplitude result for each orbital element-which one may ignore as long as each is small enough. In addition, "secular" terms arise, which are those having the time, t, as a factor; these terms are the "dangerous" ones because for large-enough time intervals they will upset the present order of the planetary system. The question of the "stability" of the system revolves largely around the significance of these terms.

Laplace and Lagrange found that in the series containing only the first powers of the disturbing masses, m, no secular terms in a occur and that the secular terms in e and i are in reality sums of periodic terms of long period ( $10^5-10^7$  years, roughly). Poisson found that in the terms containing  $m^2$  there are no secular terms in a but that there are "mixed"

<sup>\*</sup>The masses are expressed in terms of the solar mass and are therefore  $<10^{-3}$ .

periodic terms, with the amplitude proportional to t. Eginitis (1889) found that in the terms containing  $m^3$  there are secular terms in a but that in the coefficients of these terms the secular changes in e and i of the first order occur, which were found to be in reality periodic, of long period. The same is therefore true for these terms in  $m^3$ , though their amplitudes increase with time. Eginitis computed for the earth and Saturn the present rate of change in a due to the secular terms in  $m^3$ . Extrapolating these back to  $T_g = -4.5 \times 10^9$  years is, of course, most hazardous, if only because the coefficient of t is in reality a fluctuating quantity; but if we do, we find:

Earth,  $\Delta a/a = +3.3 \times 10^{-5}$ , Saturn,  $\Delta a/a = +33$ . Since the theory probably breaks down when  $\Delta a \approx 0.1$  a, the result for Saturn becomes meaningless for  $\Delta t > 10^7$  years.

Somewhat more representative may be the calculations of the secular variations of the mean motions, made by Newcomb (1895) for the terrestrial planets and by Brouwer and van Woerkom (1950) for the Jovian planets. Dr. Brouwer has kindly derived the numerical values for the Jovian planets in the form used by Newcomb for the terrestrial planets. The data are combined in Table I, second column. In the third column the value of  $\Delta n/n$  has been computed, again by extrapolation to  $T_g = -4.5 \times 10^9$  years. The precise meaning of  $\Delta n/n$  is not known; if

TABLE I
SECULAR CHANGES IN THE CENTENNIAL MEAN
MOTIONS FOR 1950

Planet	Secular change	$\Delta n/n$
Mercury Venus Earth Mars Jupiter Saturn Uranus Neptune	$ \begin{array}{c}                                     $	+0.004 $-0.002$ $+0.014$ $-0.011$ $+0.013$ $-0.10$ $+1.07$ $+0.05$

it were, the problem of the long-range stability would be solved. If some or even most of the secular terms are actually parts of periodic terms (such as the series for  $\sin ct$ ), as has been suspected,  $\Delta n/n$  may give an upper limit to the true variation. In that case the system of the Jovian planets has not altered appreciably in  $5\times 10^8$  years and nothing can be said beyond that.

These numerical data illustrate what has already been expressed by Brown (1932), that the existing planetary theory, for all its incredible complexity, is not able to make dependable statements for intervals more remote than at most  $10^8$  years. Now consider for a moment systems like our planetary system, but with planetary masses either much smaller or much larger. In the former case the perturbations would be much less, the series in  $m, m^2, m^3 \dots$  being much more rapidly convergent, and the stability would be assured for much longer intervals of time. In the opposite case, with masses approaching the solar mass, no stability whatever could exist, even for short times.

The data of Table I suggest that the mean distances of the terrestrial planets are less subject to change than those of the Jovian planets. This is perhaps not surprising because the masses of the terrestrial planets are some 100 times the smaller, and each planet is perturbed principally by its immediate neighbours, with the asteroid gap isolating the two planet groups to some extent. Since the system of the Jovian planets is known to be stable for at least 107-108 years, one is tempted to conclude that the system of the terrestrial planets will not have changed appreciably for 109-1010 years, i.e. the age of the solar system, provided at least that the orbits of the Jovian planets have not changed greatly. Now there are reasons to believe that this condition may be fulfilled. The Jovian planets carry all but 0.16 per cent. of the angular momentum of the planetary system and this must have been conserved. Further, they carry all but 3.1 per cent. of the (negative) energy, and this must have been conserved. The energy among the Jovian planets is very unevenly divided; Jupiter has 84.4%, Saturn 13.8%, Uranus 1.0%, and Neptune 0.8%. Since the energy is proportional to 1/a, energy exchanges between the planets will not have been able to change Jupiter's distance from the sun appreciably, except if Saturn would have been much closer to Jupiter. This possibility is ruled out later in this paper. It is noted that the simultaneous conservation of angular momentum and energy has a great constraining effect on changes in a; because one varies very nearly as  $+a^{1/2}$  and the other as  $-a^{-1}$ , so that in this approximation simple exchanges, between pairs of planets, are not possible. (The case becomes more complex when slightly eccentric orbits with small mutual inclinations are considered, and the energy and momentum of the sun and the energy of interaction are not neglected; these statements will correspond to the successive theorems on the constancy of a.)

Related to this problem is an important recent study by Hart and Guier (1955) of the statistical properties of N weakly interacting particles in a Newtonian potential, moving in two dimensions. The energy and angular momentum of the planetary system were considered as given and the ultimate radial distribution of N particles derived. A fair agreement with the radial mass distribution among the Jovian planets was found. This

should perhaps be interpreted to mean that (a) the time scale, which was not derived, was insufficient to affect the terrestrial planets (cf. our Table I); (b) the energy and momentum of the Jovian planets are almost equal to the total amounts for the solar system so that the comparison may be regarded as applying to the Jovian planets alone; (c) since the dependence of the time scale on N was not evaluated it may be that the mass distribution computed refers to the pre-planetary solar nebula rather than to the four planets.

We may summarize by saying that two problems must be clearly distinguished: the practical one, of the stability of the planetary orbits during the past  $5 \times 10^9$  years, and the philosophical one, of the stability of the system for all time to come. One may hope that at least the answer to the former question can be found.

Brown (1932) has pointed out another significant fact. There are certain near-commensurabilities among the periods of the (Jovian) planets ( $5 \text{ J} \cong 2 \text{ S} \cdot 2 \text{ U} \cong \text{N}$ ;  $3 \text{ N} \cong 2 \text{ P}$ ), just as there are such relations in the satellite systems. Motions in these pairs appear to possess a certain stability, i.e. resistance to change by perturbations. This indicates that even if there have been no major changes among the distances of the Jovian planets during geologic time, minor adjustments toward low commensurabilities may have occurred.

In passing we note the very puzzling fact that the four outer satellites of Jupiter on the one hand have orbits that are unstable on the basis of the Jacobi constant (resulting from their peripheral position in Jupiter's sphere of influence), but on the other hand have orbits that are still distinctly similar, indicating a common origin, although they must date back to the formative stage of the planetary system. A similar problem arises with the next group of three stable (i.e. permanent) satellites. These two satellite groups give empirical evidence of amazing long-range stability in the orbital elements that has implications for the general problem of planetary stability which have not yet been evaluated.

The variations of e and i cause these quantities to fluctuate within definite limits during intervals below  $10^8$  years. What happens for longer intervals is not known. Figure 1 shows the fluctuations of these elements according to Stockwell's integrations, as listed by Charlier (1927). Dots are present values. For Pluto no data are available. With modern values of the masses the fluctuations would be somewhat different but not markedly so.

We have already commented on the uncertainty of the original distances in the satellite systems, owing to the presence of librations which, as Brown has remarked, suggest that the precise initial distances may be even undeterminable. Two satellites, Phobos and the moon, which

the writer (1953d, 1954) has considered elsewhere, present special problems. In both cases the distance to the planet has greatly altered. An important new investigation on the *tidal history of the earth-moon system* has just appeared (Gerstenkorn, 1955).

The planetary obliquities and the periods of rotation will not have changed appreciably during geologic time except for the earth where tidal friction has introduced large changes. According to Darwin (1908) the initial obliquity of the earth was probably about 12° but the initial

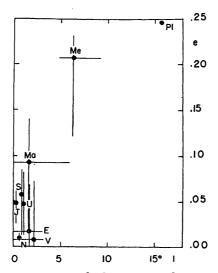


Fig. 1—Inclinations and eccentricities of planetary orbits. *Dots* indicate present values; *lines*, total range of secular variations according to Stockwell.

period of rotation is not known, owing to the uncertain place of origin of the moon; it may have been 5<sup>h</sup>-10<sup>h</sup>. See also Gerstenkorn (1955).

The principal reasons for assuming that no major changes have taken place in the planetary distances are, at present, general. First there are the conservation laws, of energy and momentum which, as we have seen, greatly limit any rearrangements within the system. Second, there are three empirical facts, the nearly geometric distance law (Bode's law) and the present small e and i values, which do not appear to be the result of any tendency toward equipartition, as would be expected if energy and momentum exchanges had been important. These three properties definitely appear to be original, antedating the development of planets as near-point masses; and to stem not from interactions between a few major planets originally moving more or less at random, but from regularized motion, such as can develop only in a flat gaseous disk. We shall therefore assume present values for all dynamical quantities unless a reason is known to do otherwise. The problem will then be to account, at least in a general way, for the system and subsystems so defined.

(To be continued)