The Origin of the Solar System

by H. Alfvén, Department of Plasma Physics, Royal Institute of Technology, Stockholm, Sweden

1. Historical survey

The problem how the solar system originated and got its present shape is a classical problem. From the early mythological discussions of it, it began to approach a scientific treatment through Kant. Laplace formulated a qualitative theory of it, which of cause was based on mechanical effect. As we know today, not the least through the space research during the last few years, plasma phenomena are of decisive importance for practically all phenomena associated with the evolution of disperse cosmic clouds. It is not remarkable that Laplace knew nothing about this, but it is very remarkable that most cosmogonists of today still deny that electromagnetic forces were of decisive importance for some important phases of the formation of the solar system. We shall refer to all the theories based on exclusively mechanical forces as "Laplacean".

In some respects there is a general agreement about how the solar system originated. It derived from an interstellar cloud about 4 – 5 billion years ago and the formation included the following phases.

1. Formation and evolution of an interstellar cloud.
2. The formation of the sun in the cloud.
3. The accumulation of matter around the sun.
4. The formation of planetesimals of this matter. (This seems now to be accepted also by most Laplacean theories although there are still a few who claim that planets-satellites were formed by Jeans collapse.)
5. The accretion of the planetesimals to planets.
6. Formation of satellites around the planets. The most straightforward assumption about this process is that it is basically the same as the one by which planets were formed around the sun. Such an approach works very

well (see Alfvén and Arrhenius (1975, 1976) but still the majority of the cosmogonists claim that the satellite formation took place through completely different processes.

We are now witnessing how recent *in situ* measurements of plasmas in the magnetospheres (including the heliosphere or interplanetary space) causes a drastic revision of cosmic plasma physics. This includes also a revision of how interstellar clouds are formed and how stars are formed in them (see H. Alfvén, 1981 a and b).

It would carry us too far to discuss how this causes new approaches to the processes listed above. Insteady we shall concentrate our attention to the problem how *the planets and satellites got their orbital momenta.* This is a serious difficulty with the Laplacean theories which still is not solved in a satisfactory way but simply shuffled under the rug. On the other hand the transfer of angular momentum from a rotating magnetized body to a surrounding plasma *through electromagnetic effects* is a phenomenon which is very well studied by recent *in situ* measurements in the magnetospheres (the terrestrial, the Jovian, the Saturnian magnetospheres, and also in the heliosphere). In order to apply this process to the cosmogonic transfer of angular momentum we must extrapolate it from the present very low magnetospheric densities (~1 particle/cm^3) up to the cosmogonic densities which must have been several orders of magnitude larger.

The first systematic attempt to base a cosmogonic theory on electromagnetic and hydromagnetic effects was done 1942 (Alfvén). It led to the discovery of two important *observational* regularities in the solar system: (1) *the band structure,* and (2) *the cosmogonic shadow effect (the two-thirds fall-down effect).* Both have been discussed in detail in Alfvén and Arrhenius (1975 and 1976). Further, the band structure has more recently been discussed by Alfvén (1981b, p. 115). The second observational regularity, the cosmogonic shadow effect, will be the subject of this paper.

2. Plasma effects. Early theory of the two-third fall-down ratio

The theoretical explanation of the two-thirds effect is very simple. If charged particles (electrons, ions or charged grains) move in a magnetic dipole field — strong enough to dominate their motion — under the action of gravitation and

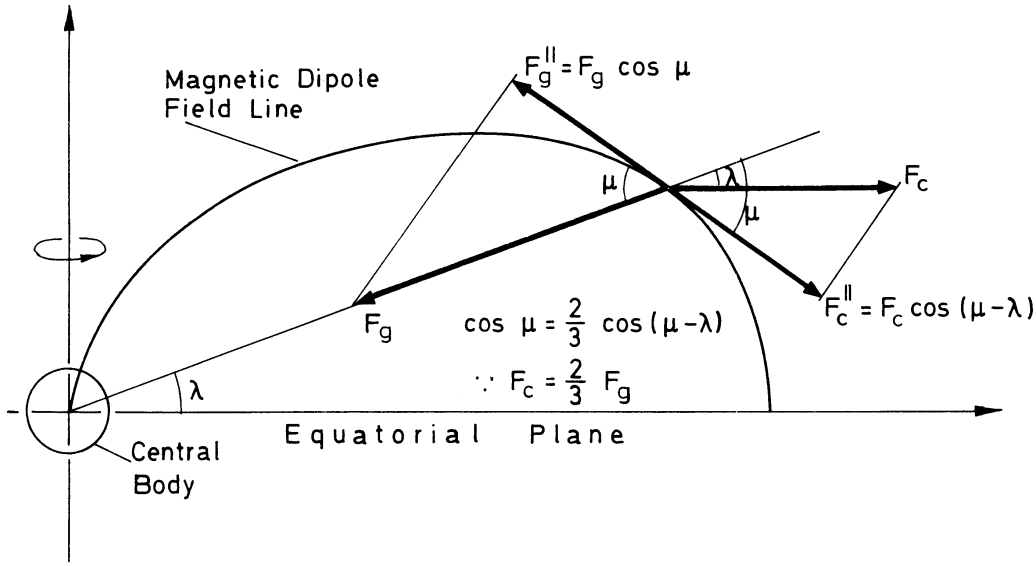


Fig. 1: Charged particles (electrons, ions, charged dust) in an axi-symmetric dipole field around a gravitating rotating body. If their motion is magnetic-field dominated, a quasi-stationary motion requires that the projections of gravitation and centrifugal force on the magnetic field line are equal. As shown by Alfvén and Arrhenius (1975, 1976) this means

$$v_\Theta = (\tfrac{2}{3})^{1/2} v_K$$

where v_Θ is the rotational velocity and v_K the Kepler velocity.

the centrifugal force, they will find an equilibrium in a circular orbit — parallel to the equatorial plane but not in it — if their centrifugal force is ⅔ of the gravitational force (Fig. 1). The consequence of this is that if they become neutralized, so that electromagnetic forces disappear, the centrifugal force is too small to balance the gravitation. Their circular orbit changes to an elliptical orbit with the semi-major axis $a = \frac{3}{4} a_0$ and $e = \frac{1}{3}$ (where a_0 is the central distance where the neutralization takes place [Fig. 2, and 3]). Collisional interaction (viscosity) between the condensed particles will eventually change the orbit into a new circular orbit with $a = \frac{2}{3} a_0$ and $e = 0$. This is the essense of the "two-third fall-down law".

The two-thirds fall-down ratio leads to a "cosmogonic shadow" effect. If, before the fall-down, there is plasma in the region $a_1 - a_2$, after the fall-down we will find this matter in the region $\frac{2}{3}(a_1 - a_2)$. On the other hand, if the plasma in the region $a_2 - a_3$ is absorbed by a planet or satellite, or by small grains orbiting in that region, we will eventually find no (or very little) matter in the region $\frac{2}{3}(a_2 - a_3)$. In other words, matter in the region $a_2 - a_3$ will produce a *shadow* in the

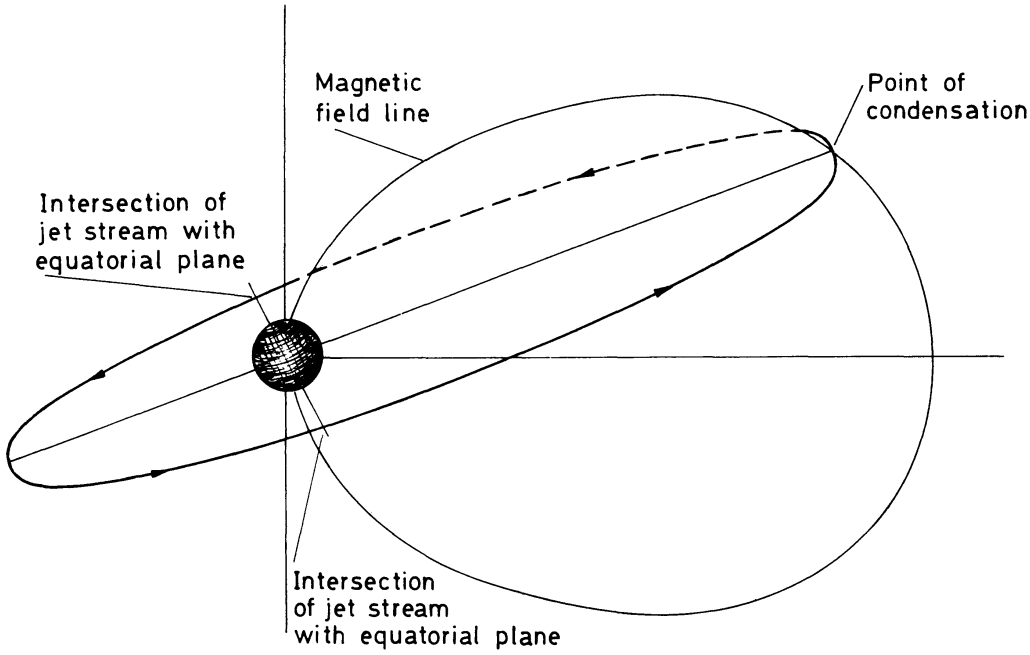


Fig. 2a

Fig. 2 (a), (b): Vanishing magnetic forces give a transfer into elliptic orbits. If the magnetic field or the particle charge suddenly disappears, the particles at the central distance a_0 will orbit in ellipses with semi-major axis $a = \frac{3}{4} a_0$, and eccentricity $e = \frac{1}{3}$. They will collide mutually when they reach the nodes in the equatorial plane at $a = \frac{2}{3}$. Collisions will transform their orbits, eventually to circular orbits in the equatorial plane with $a = \frac{2}{3} a_0$, $e = O$.

region $\frac{2}{3}(a_2 - a_3)$. What this means is shown in Fig. 5–10. As we shall see in this paper, this shadow effect can explain the macro-structure of the Saturnian rings (SR) and the main belt asteroidal region (AR).

Although it is possible — or even likely — that this effect was of importance at the formation of planets and satellites in general, we cannot expect to find any clear trace of it except in these two regions, because the accretion of planetesimals to planets/satellites has obliterated the primary structure. The reason why this has not occurred in the Saturnian ring region is that this is located inside the Roche limit. In the asteroidal region the density is so small that the accretion of planetesimals to large bodies has not yet led to the accumulation of all mass into one or a small number of bodies. It still gives essential information about the planetesimal state.

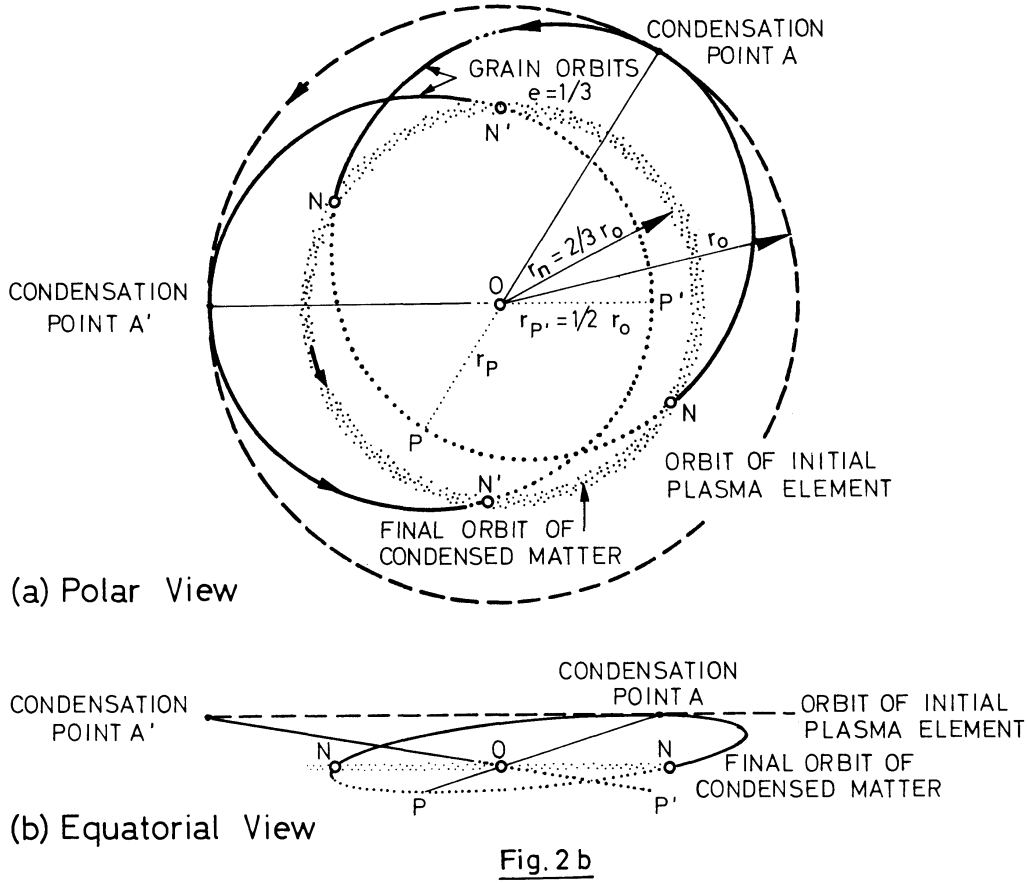


Fig. 2 b

See caption on opposite page.

3. Correction of the Fall-Down Theory

When general interest in cosmic electrodynamics began to avalanche in the 1950s, it was thought appropriate to develop the cosmogonic theory further. A monograph on the origin of the solar system (Alfvén, 1954) was published. In this it was shown that in the SR, but not in the AR, the cosmogonic shadow produced a "load" (or "reaction") on the particles producing the shadow, with the result that the ratio 2 : 3 = 0.67 should be reduced by about 5 % (Fig. 5). If density function is approximated as $p = \text{const}\, a^n$, the reduction depends on the value of n. The value of n was estimated from the assumption that it was essentially the same in the whole region of the inner Saturnian satellite (ring – Rhea) and from this it was derived that the fall-down ratio should be reduced to $\Gamma = \frac{2}{3}_{\text{corr}} = 0.63 - 0.65$ (Alfvén, 1954, p. 86 – 92).

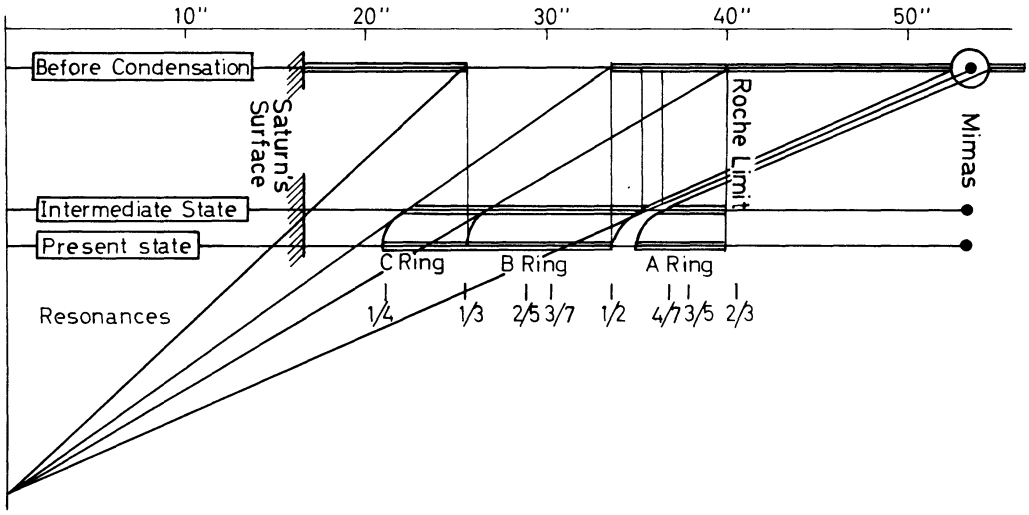


Fig. 3: Correction of the ⅔ fall-down ratio. The upper line, marked "Before Condensation", shows the plasma density in the equatorial plane of Saturn, the centre of which is at the left end. The next line, marked "Intermediate State", represents the density of the grains produced by condensation of the gas. As these move at ⅔ of the distance of the gas out of which they are produced, the density distribution is obtained by a geometrical construction reducing the central distances to ⅔. The lower line, marked "Present State", represents the density distribution into which the "Intermediate State" is transformed by the "shadow reaction", or "shadow load".

4. Is it reasonable that the present SR und AR data are of cosmological relevance?

Before going into a detailed analysis of the observational results, it is necessary to clarify whether any present signatures in the SR and AR could in principle be fossils from cosmogonic times 4-5 10^9 years ago. There are three reasons for not excluding this possibility.

(1) *The resonances* (with Jupiter in the AR and with Mimas, and possibly also other satellites in the SR) form a sort of obstacles blocking the free transfer of mass in the radial direction. This is especially conspicuous in the AR with the Kirkwood gaps, which are empty regions dividing the rings into a number of "slices". The ringlet structure of SR, according to the Voyager I reports, may to some extent support this conclusion.

(2) *The diffusion coefficient* of grains or planetesimals in Kepler orbits which collide inelastically is not positive (as is generally taken for granted) but

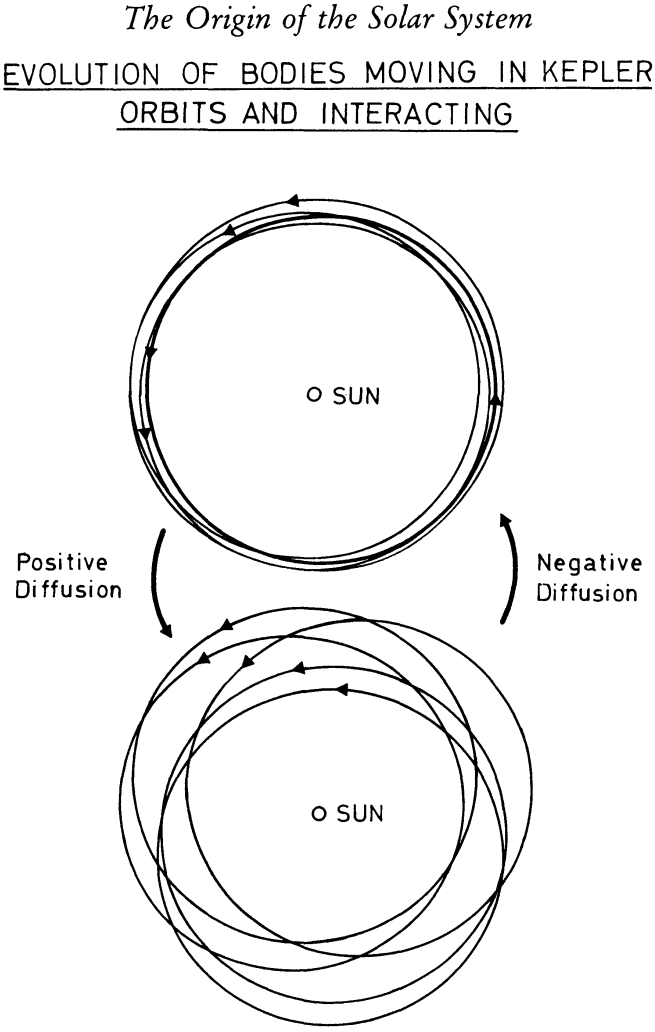


Fig. 4: Negative diffusion. If a number of grains or planetesimals orbit around a central body, as in the upper figure, collisions would change the orbits into the configuration shown in the lower figure if the diffusion were positive. In reality, collisions between the orbiting grains will tend to equalize their orbits, hence transforming the lower pattern into the upper (negative diffusion).

negative, as shown by Baxter and Thompson (1973, 1975). This is illustrated by Fig. 6. A positive diffusion coefficient would convert state A into B. But as already a qualitative discussion shows, inelastic collisions between the orbiting bodies will tend to smooth out the differences between their orbital parameter, so that the state B is transferred into A, which means a *negative* diffusion. The mathematical proof and the clarification of the circumstances under which this is true, are found in the papers by Baxter and Thompson (1973, 1975). Their

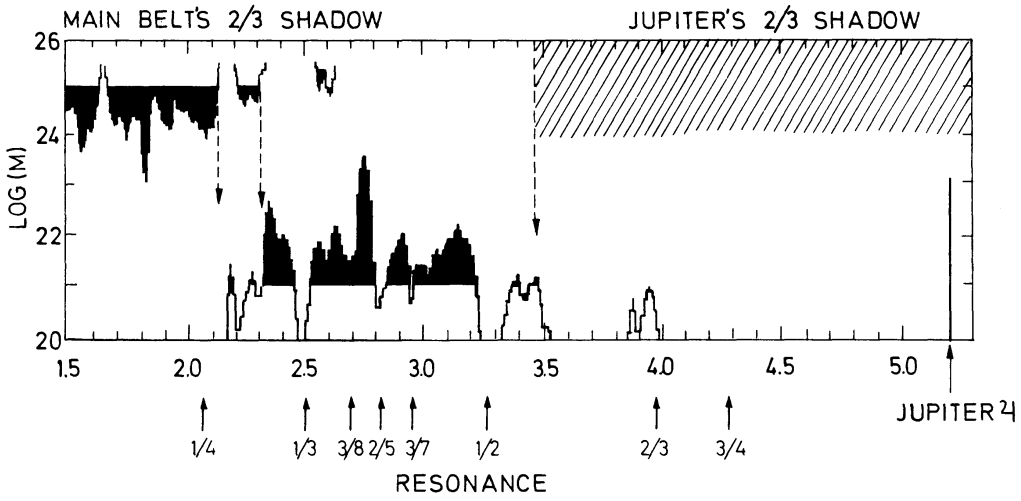


Fig. 5: Mass distribution in the asteroidal belt as a function of semi-major axis a. As the diagram is logarithmic, almost only the black areas count. Above is plotted the same diagram upside down and diminished by ⅔ (Jupiter included). The Kirkwood gaps are marked. If their distortion of the cosmogonic effects is eliminated we find that the belt has a high density region between 3.22 and 2.36, and low density region up to 3.50 and down to 2.20. The inner limit of the high density region is ⅔ of the beginning of the belt, and the inner limit of the whole belt is ⅔ of the beginning of the high density region. See further Fig. 9 and 10.

results may also explain *the large number of "ringlets"* observed by Voyager I. These may be analogous to the "jet streams" which seem to be essential for the transfer of the planetesimal state into the present planet and satellite state.

(3) Other arguments are given by Alfvén and Arrhenius (1975, p. 163 and 1976, p. 311).

5. Pre-Voyager results on cosmogonic shadow effects in SR and AR

Before the Voyager 1/Saturn encounter, the best observations of the Saturnian rings were those of Dollfus (1961, 1970) and Coupinot (1973). The state of the shadow effect in the light of these was summarized by Alfvén (1976). This paper also included a comparison with the AR.

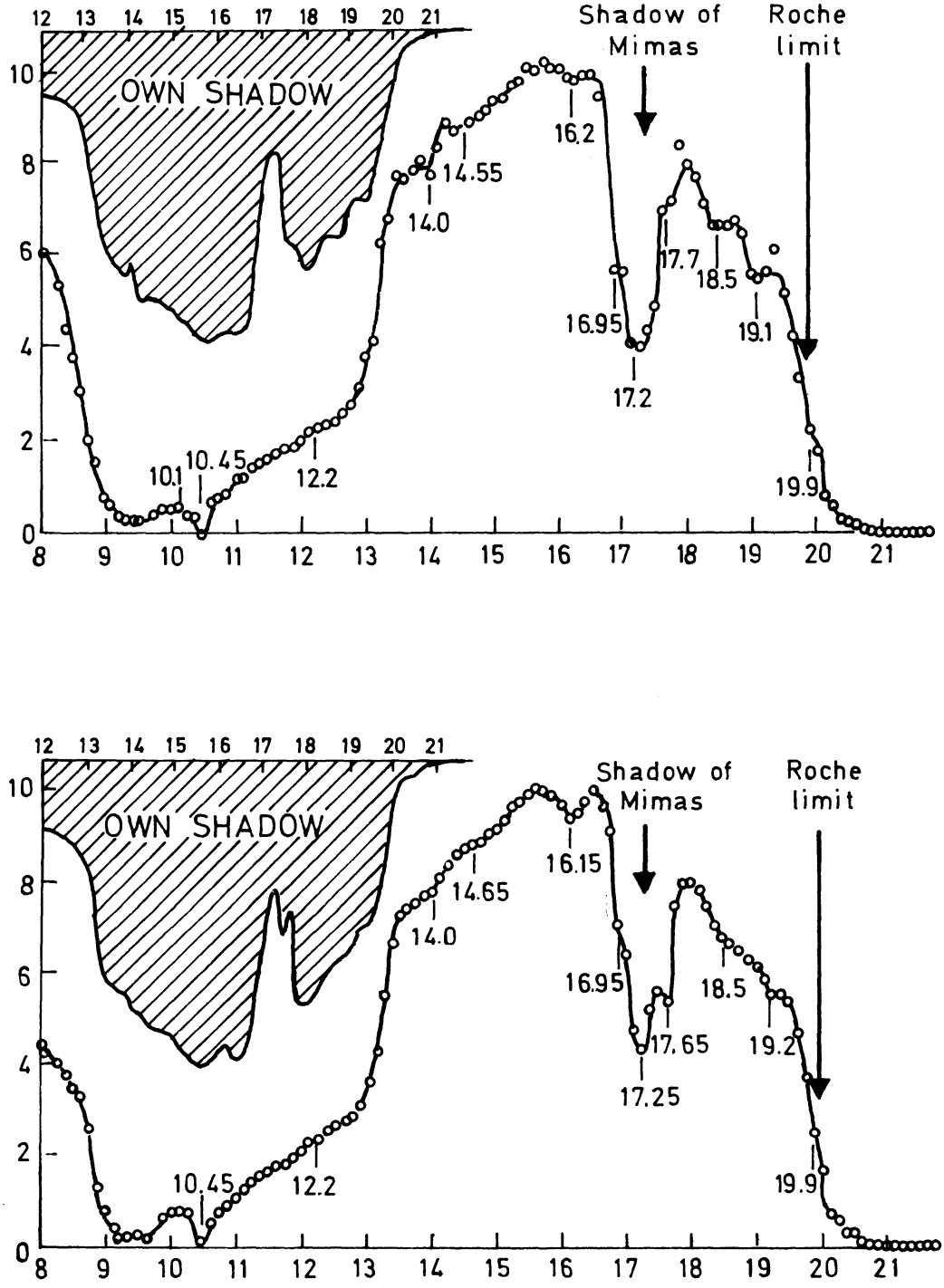


*Fig. 6: Cosmogonic shadow effects explaining Coupinot's photometric diagrams of the Saturnian rings.*The same treatment as in Fig. 5 shows that Cassini's division is the cosmogonic shadow of Mimas, and the fall in intensity marking the border between the B and the C rings is the shadow of the Roche limit. See further Fig. 8.

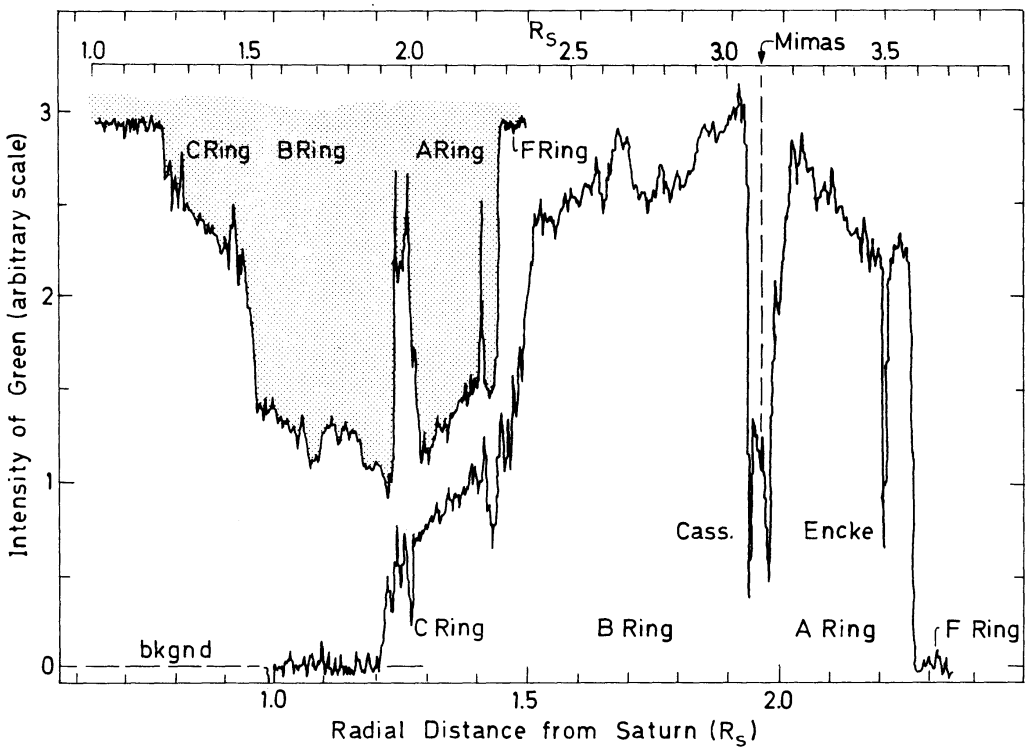


Fig. 7: Voyager 1/Saturn results. The same diagram as in Fig. 6, but with the Voyager 1/Saturn data. The fall-down ratio is chosen to be 0.63. The results of the comparison are given in Table 1.

The result of the latter is shown in Fig. 5, which gives the *mass* density of the asteroidal belt as a function of a. Since the cosmogonic theory primarily refers to the *mass* distribution, it is essential to use this instead of the usual plot of the *number* as a function of a. The mass is calculated from the absolute magnitude of each asteroid, assuming the density and albedo to be constant. As only the order of magnitude of the mass is essential, this approximation is legitimate.

The comparison with the expected cosmogonic shadow effect is made by introducing the same diagram (including the position of Jupiter) upside down and diminished in the ratio 2 : 3. In studying the diagram, it should be observed that the y-axis is logarithmic. Hence, to a first approximation, only the black areas count. The mass of the white areas represent less than 1 % of the mass.

It is seen that in Jupiter's "cosmogonic shadow"—i.e., outside ⅔ of Jupiter's orbit, which means a = 3.5—there is very little mass. Only the Hilda group at

TABLE 1: SHADOW EFFECT IN SR

	a	Γ
Mimas	3.10	0.63
Cassini centre	1.96	
F-Ring	2.33	0.63
Roche	2.26	0.65
Border B-C	1.48	
B outer	1.93	0.64
C inner	1.23	
Theoretical value		0.63 – 0.65

a = 3.9 – 4.0 represents any appreciable mass. (The Hilda group may be connected with the 3 : 2 resonance, and hence not directly relevant to our problem. See further Fig. 11.)

The Kirkwood gaps which are resonance phenomena, dominate the fine structure of the belt. Indeed, the 2 : 1 and 3 : 1 resonances are very deep, and the 5 : 2 and 7 : 3 resonances are also conspicuous. The 2 : 1 resonance is the most pronounced depression, and may be the reason why the mass density in the belt does not reach its full value outside the gap. The most massive part of the belt begins at a = 3.20, immediately inside the 2 : 1 resonance, and continues until a = 2.36. Inside this limit, the average mass density (smoothing out the Kirkwood gaps) is smaller by about one order of magnitude. A sharp inner limit is located at 2.20. There are a number of small asteroids inside this limit, but their total mass is negligible. (The same material will be presented in a somewhat different and, in some respects, clearer form, in Fig. 9.)

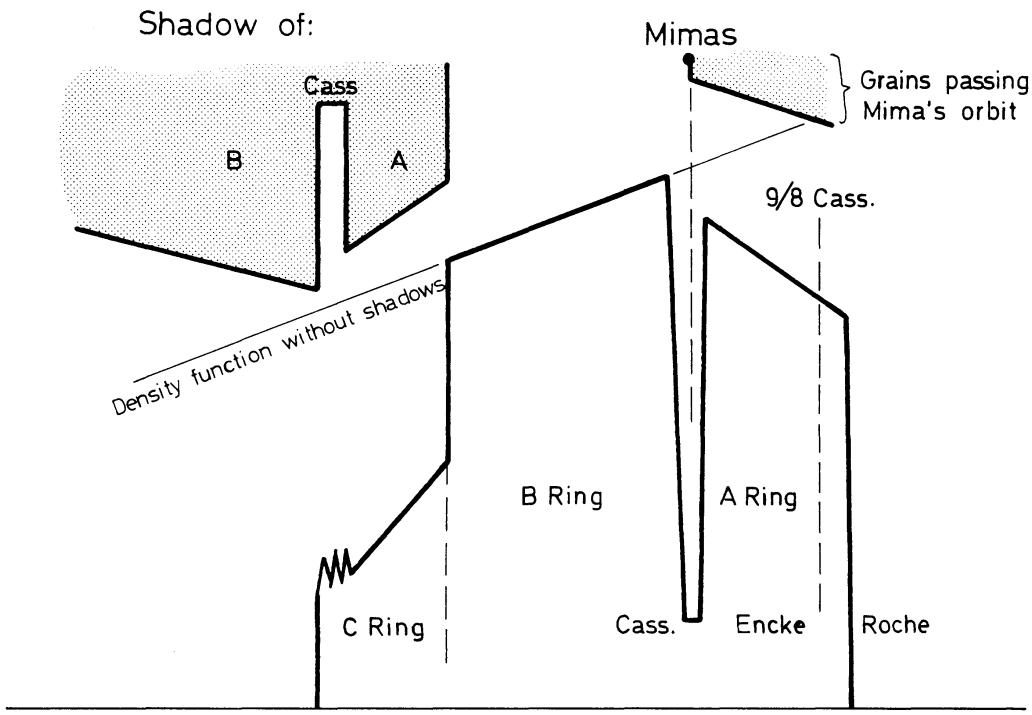


Fig. 8: Idealized picture of the Voyager 1/Saturn results. In order to concentrate the attention on the macro-structure, the micro-structure is smoothed out. Mimas produces the Cassini division, the Roche limit produces the fall in intensity between the B and C rings, and the outer edge of the B ring produces the inner limit to the C ring. If the centre of Cassini is multiplied by 9 : 8, we reach the position of the Encke division.

A similar diagram for the SR using two of Coupinot's photometric diagrams is seen in Fig. 6. The shadow of Mimas gives the Cassini division and the beginning of the "own" shadow of the SR explains the rather sharp fall in luminosity, marking the limit between the B and the C rings. The C ring data were not considered to be reliable.

6. The Voyager 1/Saturn results

Fig. 7 shows the photometric curve of the measurements in green, kindly placed at my disposal by Dr. J. Cuzzi. The diagram also shows the cosmogonic shadow, as earlier obtained by reducing the size by a factor $\Gamma \approx \frac{2}{3}$ and turning it

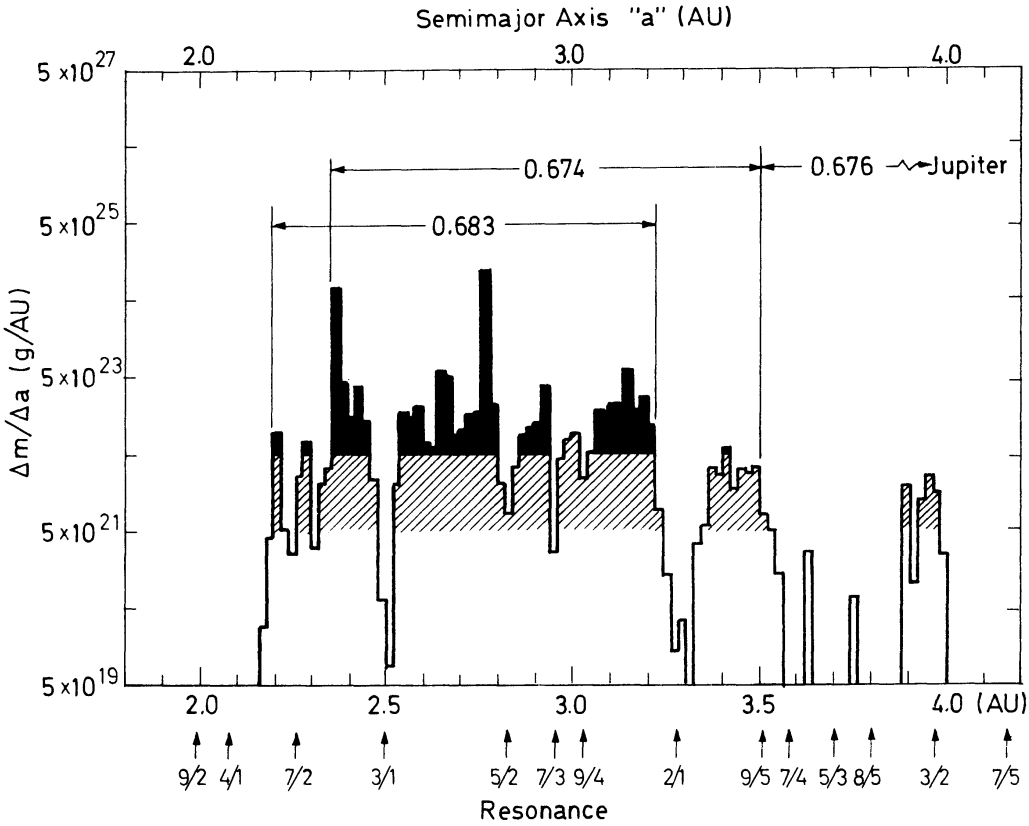
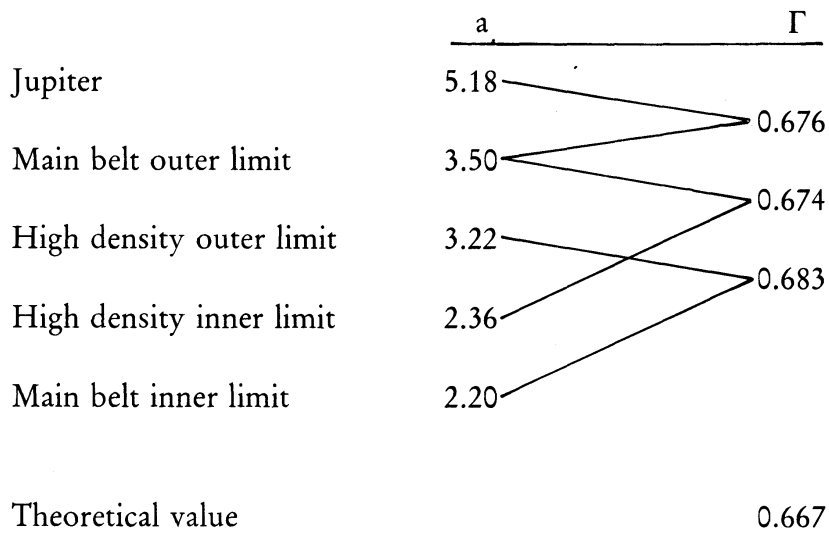


Fig. 9: Shadow effect in the AR. Essentially the same data as in Fig. 6, but represented in such a way as to make the numerical comparison as accurate as possible. Note also that this diagram is logarithmic, so that to a first approximation only the black areas should be considered. The shaded areas represent one order of magnitude smaller mass density, and the white areas can be almost neglected. Results are summarized in Table 2.

upside down. Also, the orbital distance of Mimas is included. The exact value of Γ is somewhat uncertain but as stated in section 3, it should be in the narrow range $0.63 < \Gamma < 0.65$. In order to place the shadow of Mimas exactly at the centre of the Cassini division, the value of 0.63 has been chosen.

The diagram shows that the rapid decrease in intensity marking the border between the B and C rings coincides with the shadow of the Roche limit, just as in Coupinot's diagrams). But the more reliable data from the C ring *give a new identification of a shadow effect:* the inner limit of the C ring, where the intensity drops to zero, is located where we expect the shadow of the outer limit of the B ring. This is logical, because it is not until we have passed the A ring and meet the B ring that the opacity of the ring system gets its full value.

TABLE 2: SHADOW EFFECT IN AR



The detailed results of the comparison are given in Table 1. The value of Γ is 0.63 for Cassini-Mimas, and a slightly higher value is obtained for the outer B–inner C ratio. The B-C border is not very well defined. The luminosity drops from a = 1.51 to a = 1.46, with the steepest drop at a = 1.49. If this is compared with the rapid increase at the Roche limit at a = 2.28, we obtain Γ = 0.65. A comparison with the F ring at 2.33 gives Γ = 0.63. It is possible that the lack of sharpness at the transition is due to the combination of the Roche and F ring shadows.

Fig. 8 shows an idealized picture of the observational curve with all the small-scale features smoothed out for concentrating the attention on the large-scale features. It is obvious that cosmogonic shadow effects can account for the main structure with the exception of the Encke division. A tentative theory of this as due to an adiabatic fall-down has recently been given (Alfvén, 1981 c).

7. Comparison with the asteroidal ring

The Voyager 1/Saturn results shall now be compared with AR density distribution. Fig. 9 is essentially the same diagram as Fig. 7, but the shadow diagram is removed and the material is represented in such a way as to make a

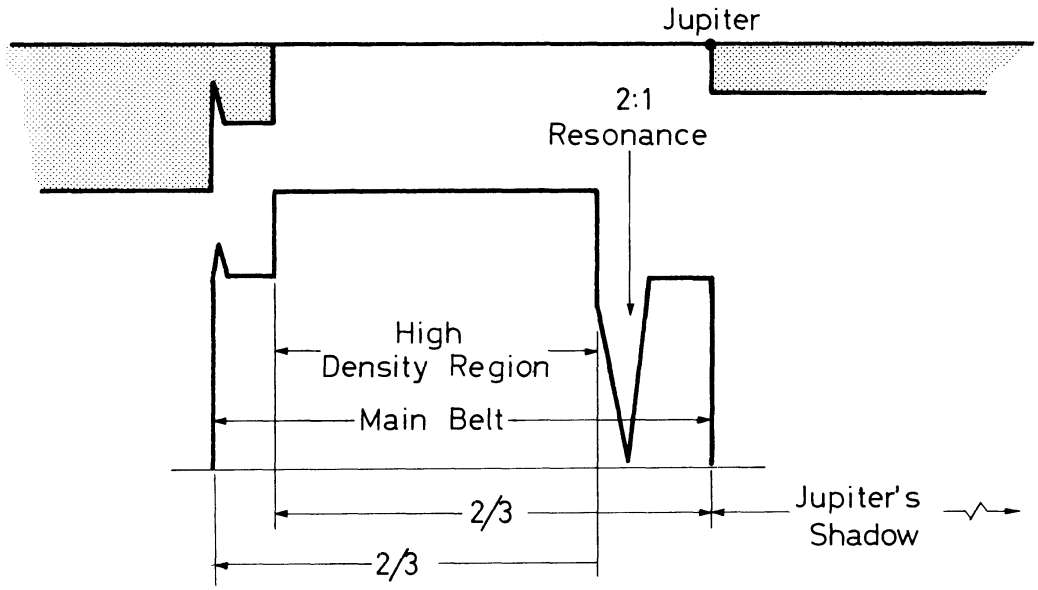


Fig. 10: Idealized picture of the AR with the Kirkwood gaps smoothed out.

quantitative evaluation of Γ as accurate as possible. The black areas represent a mass density of about one order of magnitude larger than the shaded areas, and in the white areas the mass is negligible.

We can divide the AR into three parts:

(1) The high density AR, essentially *black*.

(2) The inner and outer rings of the main belt, essentially *shaded* regions.

(3) The Hilda band, which is *marginally shaded*.

As limits of these regions, we take the points where the rise or fall are steepest, as shown in Figure 9. Figure 10 shows an idealized picture with the Kirkwood gaps smoothed out. The values of Γ, as defined in section 2, are given in Table 2.

As mentioned earlier, these values should theoretically be 2 : 3 = 0.667, not corrected for shadow load (because this is taken up by the sub-visual asteroids which are likely to have a larger total cross-section than the visual asteroids).

Hence, the agreement between the observational and theoretical values is very good (discrepancies of only one or two per cent).

8. Resonances

Fig. 11 gives a comparison between the SR and AR, normalized to the orbital distances of Jupiter and Mimas, and with some of the main resonances with these bodies indicated.

The relation between resonances and shadow effects is most clearly shown in the AR. The Kirkwood gaps are very clearly marked and hence, are easily distinguished from the shadow pattern. The very strong 2 : 1 resonance may be the cause of the low density of the outer ring, but this is far from certain. We do not yet have a clear reason for the low density of the main belt outside the 2 : 1 resonance. The 3 : 1 resonance is clearly shown and seems not to affect the average density which is essentially the same outside as inside this resonance. The 4 : 1 resonance falls far below the inner limit (of the *mass* distribution, but there are numerous very small asteroids below this limit).

The resonances in the SR are less easy to discuss. Earlier, the whole Cassini division has been attributed to the 2 : 1 resonance with Mimas, but this seems difficult to accept (Alfvén and Arrhenius, 1975, p. 64, 66, 160–163; 1976, p. 308–311). However, it seems likely that the inner edge of the Cassini division is sharpened by the 2 : 1 resonance. The 3 : 1 and 4 : 1 resonances fall close to the B-C border and the inner limit of the C ring. It seems unlikely that resonances can explain the macro-structure, but they are probably very important for the micro-structure. Further detailed study is necessary in order to clarify what effects are due to cosmogonic shadow, to resonances, and to other effects.

9. Discussion: Remarks on the model

We have seen that the 1942 model (somewhat developed in the 1954 model) can account for the macro-structure of the SR as found by Voyager 1 in 1980, with an accuracy of a few per cent. The agreement in the AR is perhaps even better. However, this does not necessarily mean that this old model, which has not been developed at all for a quarter of a century, is of a kind which can claim to be acceptable today. Indeed, our knowledge of cosmic plasma physics has changed so drastically during this time that this would be unreasonable.

Hence the theoretical problem we are facing is to work out a model of a modern type, which is similar to the old model in the respect that a ⅔ fall-down ratio can

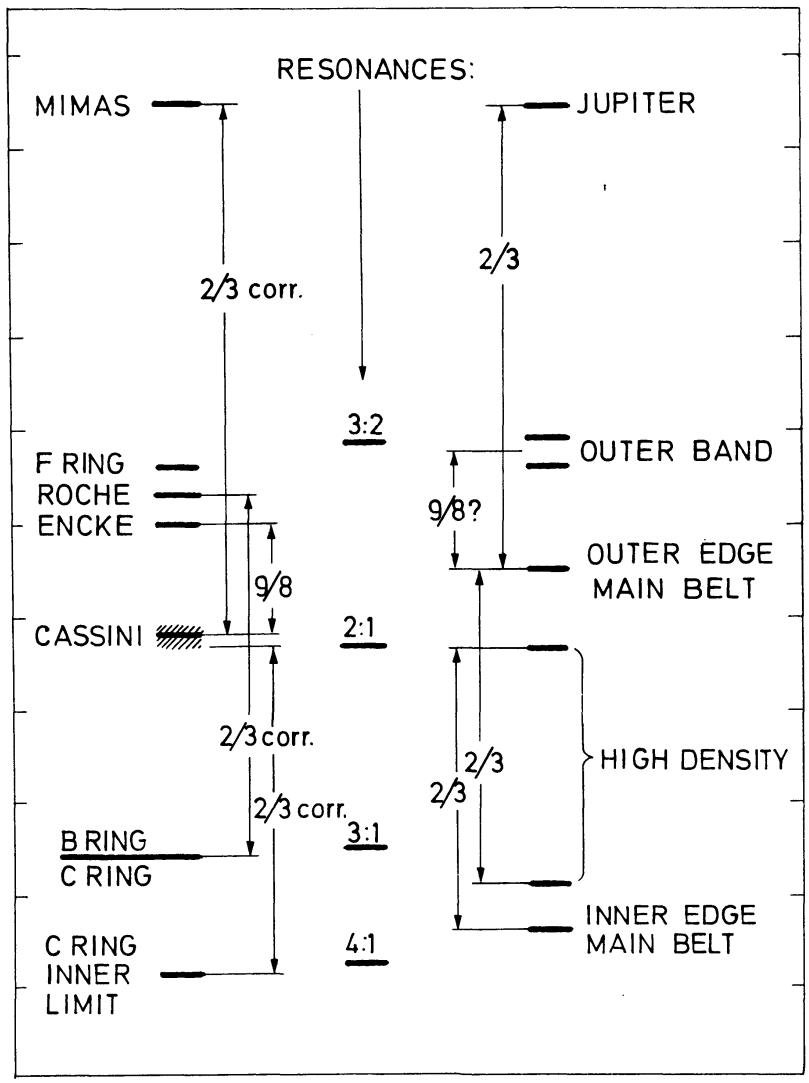


Fig. 11: Comparison between the Saturnian Ring and the Asteroid Ring. Summary of the results of this paper, with the resonances plotted in the middle. In the AR the resonances are very pronounced (Kirkwood gaps) but there is no obvious connection with the smoothed out macro-structure, except that the high density main belt begins close below the 2 : 1 resonance. Whether the outer band (the Hilda's) are due to a 3 : 2 resonance or are 9 : 8 of the outer edge of the belt (or both!) is an open question.

Because of the very small mass of Mimas, the macro-structure of SR is probably not produced by resonances at all. However the 3 : 1 and 4 : 1 resonances fall close to the B-C limit and inner limit of C ring, and the resonance effects should be studied more closely. The 2 : 1 resonance probably sharpens the inner limit of Cassini.

be derived from it. This seems quite possible to do. Indeed, the ⅔ fall-down ratio is a rather general characteristic of the transition from a magnetized plasma state to a non-plasma state under the condition that the magnetic field is a dipole field. (As Bonnevier has shown, rather large deviations from a dipole field are allowed without changing the fall-down ratio appreciably).

If an acceptable model is found we could calculate within what limits the essential parameters (like density, degree of ionization, magnetic field, etc.) must be. Then the observational fall-down ratios will allow us to conclude that at the time when the SR and AR were formed, the parameters were within these limits in these regions. This may give us an important criterion for checking the viability of different cosmogonic theories.

Addendum

Since this was written, a first comprehensive report of the Voyager 1/Saturn results have been published (*Science*, **212,** p. 159, 1981). The numerical values on which the above conclusions were based, differ by only a few per cent. There is also rather much new information which indicates that the same fall-down ratio may be found in a number of places. The detailed analysis of all the obtained data should be reserved to a future publication.

References

Alfvén, H., 1942, On the cosmogony of the solar system, *Stockholms Observatoriums Annaler* **14,** No. (2): 3; (5): 3.

Alfvén, H., 1954, *On the Origin of the Solar System.* Oxford: Clarendon Press.

Alfvén, H., 1976, The Saturnian rings, *Astrophys. & Spa. Sci.*, **43,** 97.

Alfvén, H., 1981a, Origin of Solar System. Introductory lecture at the COSPAR Budapest Meeting. Pub. in *Adv. Space Res.*, **1,** 5.

Alfvén, H., 1981b, *Cosmic Plasma.* Dordrecht, Holland: D. Reidel Publishing Co.

Alfvén, H., 1981c, The Voyager 1/Saturn Encounter and the Cosmogonic Shadow Effect, TRITA-EPP-81-01, Royal Institute of Technology, Dept. of Plasma Physics, Stockholm, Sweden (to be published in *The Moon and the Planets*).

Alfvén, H. and Arrhenius, G., 1975 *Structure and Evolutionary History of the Solar System.* Dordrecht, Holland: D. Reidel Publishing Co.

Alfvén, H. and Arrhenius, G., 1976, *Evolution of the Solar System.* Washington, D. C.: NASA SP-345.

Baxter, D. and Thompson, W., 1971, Jetstream formation through inelastic collisions, in *Physical Studies of Minor Planets,* NASA SP-267, T. Gehrels, ed., 319, Washington, D. C.

Baxter, D. and Thompson, W., 1973, Elastic and inelastic scattering in orbital clustering, *Astrophys. J.,* **183,** 323.

Coupinot, G., 1973, *Icarus,* **19,** 212.

Dollfus, A., 1961, Visual and photographic studies of planets, in *The Solar System* by G. P. Kuiper and B. M. Middlehurst, Vol, 3., p. 568, Chicago.

Dollfus, A., 1970, *Icarus,* **11,** 101.