

Formation of the solar system by instability

Evgeny Griv and Michael Gedalin

Department of Physics, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel
email: griv@bgu.ac.il

Abstract. The early gas and dust protosolar nebula of the solar composition is considered analytically. A simultaneous formation of the sun and all the planets around it ($\approx 5 \times 10^9$ yr ago) through a local gravitational Jeans-type instability of small-amplitude gravity perturbations in the nebula disk is suggested. It is shown that a collective process, forming the basis of the disk instability hypothesis, solves with surprising simplicity the two main problems of the dynamical characteristics of the system, which are associated with its observed spacing and orbital momentum distribution, namely, Bode's law on planet spacing and the concentration of angular momentum in the planets and mass in the sun. Besides, the analysis is found to imply the existence of new planets or other Kuiper-type belts of comets at mean distances from the sun of 87 AU, 151 AU, 261 AU, 452 AU, 781 AU (Mercury, Venus, ..., Asteroid belt, ..., Neptune, Kuiper belt, new planets or other Kuiper-type belts).

Keywords. Planetary systems, planetary formation, solar system, instabilities and waves

1. Introduction

The “standard” theory of the multistage accretionary formation of planets, or the so-called core accretion mechanism (Safronov 1972; Pollack *et al.* 1996) remained the most popular until recently, when it was criticized by Boss (2002, 2003) and others. The main problem is the timescale, which is longer than estimates of the lifetime of many planet-forming disks (Taylor 1992, §1.9.1; Feigelson & Montmerle 1999).

We suggest that the sun and all the planets around it were created simultaneously by disk instability. That is, at an early stage, the protosolar nebula refers to a fragment that separated from a molecular cloud. Planetary formation is thought to start with inelastically colliding gaseous and dust particles settling to the central plane of this rotating nebula to form a thin layer around the plane. During the early evolution of the disk it is believed that the dust particles coagulate into kilometer-sized rocky comets—“planetesimals.” On attaining a certain critical thickness (and, correspondingly, very low temperature) small in comparison with the radius of the system R , as a result of a local gravitational collapse the nebula disintegrated into the central body and a number of separate protoplanets.† Boss (2004) already demonstrated that convective cooling is able to cool the disk midplanes at the desired rate to produce bound clumps in marginally unstable disks. Following Boss *et al.* (2002), the hypothesis of disk instability envisions coagulation and settling of dust grains within the protoplanets to form ice and rock cores. A protoplanet accreted an additional amount of gas subsequently from the solar nebula after accumulating a solid core, followed by the loss of the light elements of the terrestrial planets through the thermal emission of the early sun. After its formation and

† Destabilizing self-gravity is much more “dangerous” in thin disks than in thick disks. If a rotating gaseous disk has a large vertical thickness owing to a high internal temperature, then it is stabilized against gravitational instabilities. Instabilities arise as the thickness of the disk is reduced (Shu 1970; Safronov 1980).

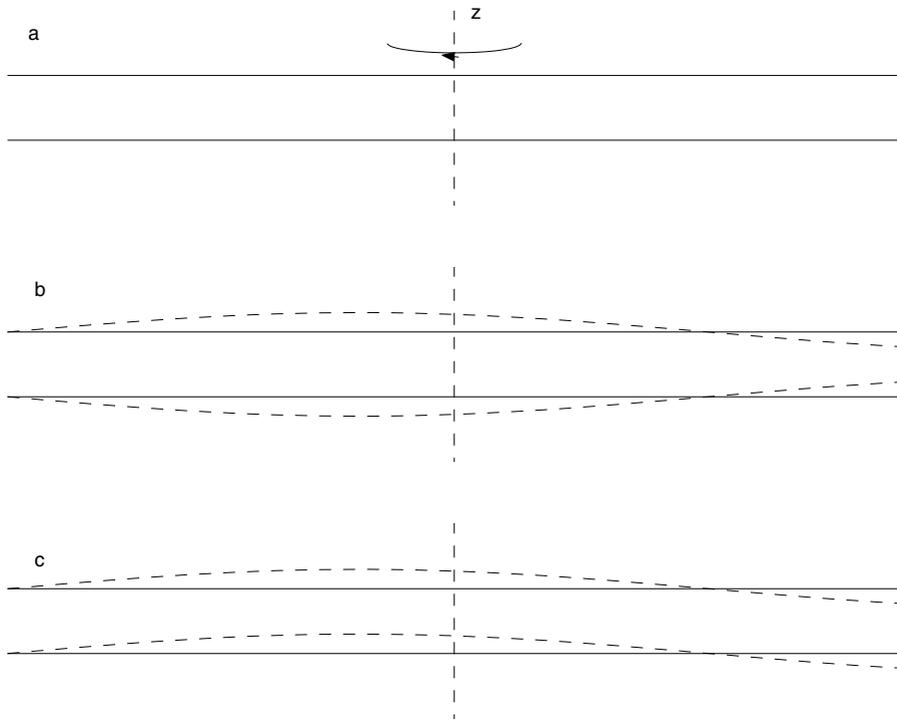


Figure 1. Sketch of perturbations of a three-dimensional protoplanetary disk. In (a) a section of the disk is shown edge-on. In (b) an even (Jeans-type) perturbation is shown (the dashed line). In (c) an odd (bending firehose-type; Kulsrud *et al.* 1971; Bertin & Casertano 1982; Griv & Chiueh 1998) perturbation is illustrated (the dashed line).

the loss of the light elements, the terrestrial planets had suffered a “late planetesimal bombardment” (Wetherill 1989).

The basic idea of planet formation through gravitational instability has been pioneered by Kuiper (1951), Urey (1958), and Cameron (1978) by considering the formation of giant planets in the solar nebula. Gravitational instability appears to be capable of forming giant planets with modest cores of ice and rock faster than the core accretion mechanism can (Boss 2002). Apparently, Cassen *et al.* (1981) were the first to simulate the planet formation in gravitationally unstable gaseous protostellar disks by N -body experiments. Recently, Mayer *et al.* (2002), Johnson & Gammie (2003), and Rice *et al.* (2003) used hydrodynamic simulations and N -body orbit integrations to study the long-term evolution of a fragmenting disk with realistic cooling. The advantages of the disk instability model are that (1) the instability process itself is quite fast, and could form planets in $\lesssim 10^4$ yr (Boss 2003), (2) in unstable, nonaxisymmetric disks differential rotation can simultaneously transfer angular momentum outward and mass inward through gravitational torques (Larson 1989; Taylor 1992, §2.9.2), and (3) such a model obviates the requirements for turbulent viscosity, frequently appealed to as a physical mechanism for outward transfer of angular momentum. This paper has precedents in earlier studies of gravity disturbances in galactic disks and Saturn’s ring disk (Lin & Shu 1966; Shu 1970; Griv *et al.* 2001, 2002, 2003; Griv & Gedalin 2003) and uses a largely previously published analysis (Griv *et al.* 1999; Griv & Gedalin 2004) of the linear growth of perturbations in a rotating disk to draw conclusions about the formation of the solar system.

2. Oscillation spectrum

The dynamics of the gaseous component in the presence of the collective self-gravitational field is considered. A Lagrangian description of the motion of a gas element under the influence of a perturbed gravity field is used, looking for time-dependent waves which propagate in a differentially rotating, spatially inhomogeneous, and two-dimensional disk. This approximation of an infinitesimally thin disk is a valid approximation if one considers perturbations with a radial wavelength that is greater h , the typical disk thickness (Shu 1970; Safronov 1980).

The time dependent surface density $\sigma(\vec{r}, t)$ is split up into a basic and a developing (perturbation) part, $\sigma = \sigma_0(r) + \sigma_1(\vec{r}, t)$ and $|\sigma_1/\sigma_0| \ll 1$, where r, φ, z are the cylindrical coordinates and the axis of the disk rotation is taken oriented along the z -axis. The gravitational potential of the disk $\aleph(\vec{r}, t)$ and the gaseous pressure $P(\vec{r}, t)$ are also of this form. These quantities σ, \aleph, P are then substituted into the equations of motion of a gas element, the continuity equation, the Poisson equation, and the second order terms of the order of $\sigma_1^2, \aleph_1^2, P_1^2$ may be neglected with respect to the first order terms. In our study we restrict the analysis to a treatment of “sausage-like” Jeans perturbations (Bertin & Casertano 1982) which are symmetric with respect to the $z = 0$ equatorial plane of the disk (which do not cause it to bend). See Fig. 1 for an explanation. These perturbations are associated with such phenomena as, for example, the appearance of the spiral structure of disk galaxies (Lin & Shu 1966; Lin *et al.* 1969; Shu 1970; Lin & Lau 1979; Griv *et al.* 1999). The resultant equations of motion are cyclic in the variables t and φ , and hence by applying the widely used local WKB method (Alexandrov *et al.* 1984) one may seek solutions in the form of normal modes by expanding

$$\sigma_1(\vec{r}, t) = \sum_{\vec{k}} \tilde{\sigma}_{\vec{k}} \exp(ik_r r + im\varphi - i\omega_{\vec{k}} t) + \text{c.c.}, \tag{2.1}$$

where $\tilde{\sigma}_{\vec{k}} = \text{const}$ is the real amplitude, $k_r(r)$ is the real radial wavenumber, m is the non-negative (integer) azimuthal mode number, $\omega_{\vec{k}} = \Re\omega_{\vec{k}} + i\Im\omega_{\vec{k}}$ is the complex frequency of excited waves, suffixes \mathbf{k} denote the \vec{k} th Fourier component, and “c.c.” means the complex conjugate. In the linear theory, one can select one of the Fourier harmonics: $\tilde{\sigma} \exp(ik_r r + im\varphi - i\omega t) + \text{c.c.}$. The solution in such a form represents a spiral plane wave with m arms or a ring ($m = 0$). The imaginary part of ω corresponds to a growth ($\Im\omega > 0$) or decay ($\Im\omega < 0$) of the components in time, $\sigma_1 \propto \exp(\Im\omega t)$, and the real part to a rotation with angular velocity $\Omega_p = \Re\omega/m$. Thus, when $\Im\omega > 0$, the medium transfers its energy to the growing wave and oscillation buildup occurs.

It is important to note that in the WKB method, the radial wavenumber is presumed to be of the form

$$k_r(r) = \mathcal{A}\Psi(r), \tag{2.2}$$

where \mathcal{A} is a large parameter and $\Psi(r)$ is a smooth, slowly varying function of the radial distance r , i.e., $d \ln k_r / d \ln r = O(1)$, and in the WKB approximation $|k_r| r \gg 1$.

Paralleling the analysis leading to Eq. (34) in Griv *et al.* (1999), we obtain that

$$\sigma_1 = \frac{\sigma_0 \beth}{\omega_*^2 - \kappa^2} \left(k_r^2 + \frac{4\Omega^2 - \kappa^2 + \omega_*^2 m^2}{\omega_*^2} \frac{m^2}{r^2} + \frac{2\Omega}{\omega_*} \frac{m}{rL} \right) + \text{c.c.}, \tag{2.3}$$

where $\sigma_1(t \rightarrow -\infty) = 0$, so by considering only growing perturbations we neglected the effects of the initial conditions, $\omega_* = \omega - m\Omega$ is the Doppler-shifted (in a rotating frame) wavefrequency, $\Omega(r)$ is the angular velocity of rotation at the distance r from the center, $\kappa \approx \Omega$ is the epicyclic frequency, $\beth = \aleph_1 + c^2 \sigma_1 / \sigma_0$, and $c = (\partial P / \partial \sigma)^{1/2}$ is the sound

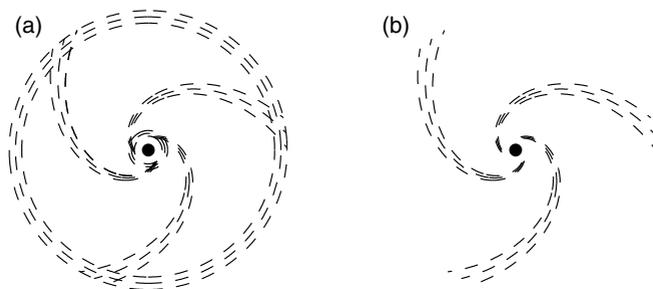


Figure 2. A schematic model of a Jeans-unstable disk with $m = 3$ spiral arms; (a) the Safronov–Toomre unstable disk ($c < c_T$) and (b) the Safronov–Toomre stable disk ($c \geq c_T$ but $c < (2\Omega/\kappa)c_T$ and in the protosolar nebula $2\Omega/\kappa \approx 2$).

speed. In Eq. (2.3) only the most important low-frequency ($|\omega_*^2| < \kappa^2$) perturbations developing in the $z = 0$ plane between the inner and outer Lindblad resonances are considered (Lin & Shu 1966; Shu 1970; Griv *et al.* 1999, 2002).

Equating the perturbed density σ_1 (Eq. (2.3)) to the perturbed density given by the asymptotic ($k_r^2 \gg m^2/r^2$) solution of the Poisson equation $\sigma_1 = -|k|N_1/2\pi G + \text{c.c.}$ (e.g., Griv *et al.* 1999, 2003), the solution of the generalized Lin–Shu-type dispersion relation is easily obtained

$$\omega_{*1,2} \approx \pm p|\omega_J| - 2\pi G\sigma_0 \frac{\Omega}{\omega_J^2} \frac{m}{r|k|L}, \quad (2.4)$$

where $p = 1$ for gravity-stable perturbations with $\omega_*^2 \approx \omega_J^2 > 0$, $p = i$ for gravity-unstable perturbations with $\omega_*^2 \approx \omega_J^2 < 0$, $|L| = |\partial \ln \sigma_0 / \partial r|^{-1}$ is the radial scale of spatial inhomogeneity, $|kL| \gg 1$, the term involving L^{-1} is the small correction,

$$\omega_J^2 = \kappa^2 - 2\pi G\sigma_0(k_*^2/|k|) + k_*^2 c^2 \quad (2.5)$$

is the squared Jeans frequency, $k = \sqrt{k_r^2 + m^2/r^2}$ is the total wavenumber, $k_*^2 = k^2 \{1 + [(2\Omega/\kappa)^2 - 1] \sin^2 \psi\}$ is the squared effective wavenumber, and $\psi = \arctan(m/rk_r)$ is the perturbation pitch angle. Equation (2.4) determines the spectrum of oscillations. This equation differs from the standard Lin–Shu expression (Lin & Shu 1966; Lin *et al.* 1969; Shu 1970) by the appearance of the total k and effective k_* wavenumbers, which originate from the consideration of the nonaxisymmetrical modes $\propto \psi$, and by the factor $\propto L^{-1}$, which originates from the consideration of the effects of spatial inhomogeneity (see Morozov 1980, 1981 and Griv *et al.* 2002, 2003 for a discussion).

From Eq. (2.5), the disk is Jeans-unstable ($\omega_J^2 < 0$) to both axisymmetric (radial) and nonaxisymmetric (spiral) perturbations if $c < c_T$, where $c_T = \pi G\sigma_0/\kappa$ is the Safronov–Toomre (Safronov 1960; Toomre 1964) critical sound speed to suppress the instability of axisymmetric $m = 0$ perturbations. Nonaxisymmetric (m , or $\psi \neq 0$) instabilities in a differentially rotating disk is more difficult to stabilize; stability is achieved only for sufficiently large sound speed $c \approx (2\Omega/\kappa)c_T \approx 2c_T$ (Morozov 1980, 1981; Griv *et al.* 1999, 2002). Thus, if the disk is thin, $c \ll r\Omega$, and dynamically cold, $c < c_T$ (or Toomre’s stability parameter $Q \equiv c/c_T < 1$, respectively), then such a model will be gravitationally unstable, and it should almost instantaneously (see below for a time estimate) taken on the form of a cartwheel, that is, a structure of spirals and rings (Fig. 2a). One concludes that Toomre’s Q -parameter that is < 1 suggests that the disk is likely subject to radial and azimuthal gravitational instabilities and might therefore be clumpy. The instability is driven by a strong nonresonant interaction of the gravity fluctuations (e.g., those

produced by a spontaneous perturbation or a satellite system) with the bulk of the particle population: in Eq. (2.3), $\omega_* - l\kappa \neq 0$, where $l = 0, \pm 1$.

The growth rate of the instability is relatively high, $\Im\omega_* \approx \sqrt{2\pi G\sigma_0(k_*^2/|k|)} \sim \Omega$, that is, the instability develops rapidly on a dynamical time scale (on a time of 3–4 disk rotations, or $\lesssim 10^4$ yr in the solar nebula). An important feature of the instability under consideration is the fact that it is almost aperiodic ($|\Re\omega_*/\Im\omega_*| \ll 1$). From Eq. (2.5), the growth rate of the instability has a maximum at the wavelength $\lambda \approx 4c^2/G\sigma_0$. At the boundary of instability ($c \approx c_T$, or $Q \approx 1$, respectively), $\lambda \approx \lambda_{\text{crit}} = 4\pi^2 G\sigma_0/\kappa^2 \sim (2-4)\pi h$; thus $\lambda_{\text{crit}} \gg h$. It means that of all harmonics of initial gravity perturbation, one perturbation with $\lambda_{\text{crit}} \approx 3\pi h$, with the associated number of spiral arms m_{crit} , and with the pitch angle ψ_{crit} will be formed asymptotically. For the parameters of the solar nebula ($R \sim 1000$ AU, $\kappa = 2\pi/T_{\text{orb}} \sim 10^{-10}$ s $^{-1}$, and the total mass of the disk $M_d \sim M_\odot$), one obtains the typical mass of a protoplanet $M_c \sim \lambda_{\text{crit}}^2 \sigma_0 \sim 10^{-3} M_\odot \sim 300 M_\oplus$. The latter is coincident in order of magnitude with the masses of giant planets. (The detection of a number of extrasolar planets with minimum masses ranging from 0.5 to $4 M_J$, where $M_J = 318 M_\oplus$ = the mass of Jupiter, has removed much of the concern that giant planets might be rare in our Galaxy; Boss 2002.)

The larger part of the initial mass of protoplanets of the Earth's group was probably blown away due to intensive thermal emission of the early sun. Such a point of view is not unnatural since the planets of the Earth's type consist mainly of elements with a high melting temperature and are almost lacking light elements. By adding to the present masses of the terrestrial planets the amount of light gases which is necessary to restore the chemical composition of giant planets, one obtains masses larger by a factor of several hundreds, coincident with the masses of giant planets.

3. Spacing of the planets

There exists the empirical Titius–Bode (TB) rule which gives the mean orbital distances of the planets and which can be written in the Blagg–Richardson formulation:

$$r_n = r_0 A^n, \quad (3.1)$$

where r_n is the distance of the n th planet from the Sun (in AU), $n = 1$ for Mercury, 2 for Venus, ..., and 9 for Neptune, $A \approx 1.73$ is the mean ratio between two consecutive planetary distances, and $r_0 \approx 0.21$. (The pure geometric sequence in the Blagg–Richardson form is better than the original TB form $r_n = 0.4 + 0.3 \times 2^n$, where $n = -\infty, 0, 1, \dots, 5$ and r_n are the distances (in AU), respectively, to Mercury, Venus, ..., Neptune. In particular, the skip in index from minus infinity to zero in the older form is quite unnatural. The mean Asteroid belt distance is considered as a regular planetary distance, $r_5 \approx 2.8$ AU. Unlike Polyachenko (Polyachenko & Fridman 1972), we do not treat Pluto as a planet but as a member of a “resonant” population of the Kuiper belt objects. The recently discovered Kuiper belt constitutes one of the few fossil records of the formation of planets and planetesimals in the early solar system.) Also, one cannot overlook the fact that many of the regularities which are found in the planetary system are also to be seen in the regular satellite systems of Jupiter, Saturn, and Uranus, e.g., the spacing of the regular satellites is a variation of the TB rule (Fig. 3). This suggests that the same cosmogonic process must have been responsible for the origin of both types of systems. Lynch (2003) has argued that it is not possible to conclude unequivocally that laws of TB type are, or are not, significant. Therefore, the possibility of a physical explanation for the observed distributions remains open. Using the Lynch method, Neslušan (2004) has continued the discussion on the statistical significance of agreement between planetary

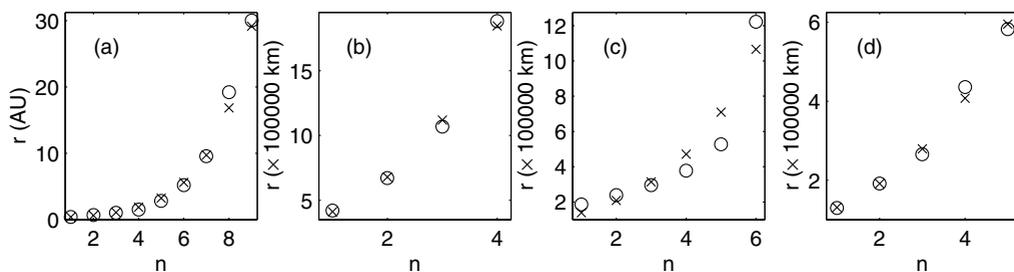


Figure 3. Relation between distances of planets (satellites) from the Sun (giant planets) r and their numbers n . Data observed are represented by circles: (a) the solar system, $r_0 = 0.21$ and $A = 1.73$, (b) the satellite system of Jupiter, $r_0 = 249.679$ and $A = 1.649$, (c) the satellite system of Saturn, $r_0 = 92.416$ and $A = 1.503$, (d) the satellite system of Uranus, $r_0 = 89.737$ and $A = 1.46$. The crosses represent the TB rule, Eq. (3.1). The organization of both types of systems does not contradict to the TB rule.

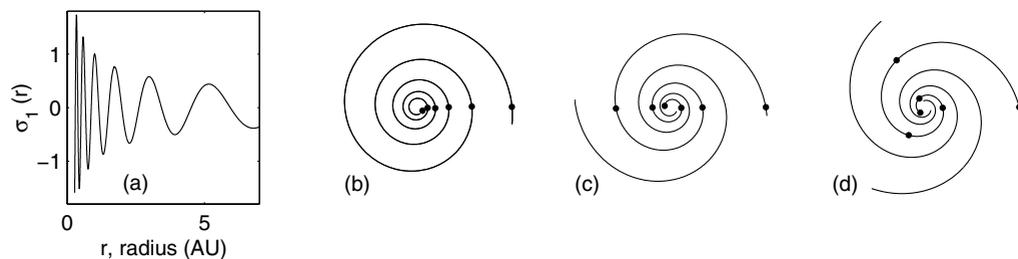


Figure 4. (a) Dependence of the perturbed surface density of the protoplanetary disk $\sigma_1(r)$ (arbitrary units) on the radius r , Eq. (3.3). The maxima of the perturbed density coincide with locations of the planets. (b) Spiral density waves with $m = 1$ arm in the (r, φ) -plane, (c) density waves with $m = 2$ arms, and (d) density waves with $m = 3$ arms. The filled circles represent the maxima of the perturbed density (protoplanets) of Jeans waves, which are unstable to both axisymmetric and nonaxisymmetric perturbations.

distributions and a power law. Interestingly, the mean orbital distance to the recently discovered Kuiper belt objects, $r \approx 46$ AU (Luu & Jewitt 2002), is in fair agreement with that given by the TB rule for the solar system’s 10th planet, $r_{10} \approx 50$ AU. Liboff (2004) has predicted the existence of a planet at a mean radius from the sun of ≈ 51 AU.

Equation (3.1) can be rewritten:

$$(2\pi / \ln 1.73) \ln(r_n / 0.21) = 2\pi n. \tag{3.2}$$

Next, the surface density of the disk may be represented in the form of the sum of the basic equilibrium surface density $\sigma_0(r)$ and the perturbed surface density

$$\sigma_1(r) = \tilde{\sigma}(r)e^{3\omega t} \cos[11.46 \ln(r/0.21) + m\varphi], \tag{3.3}$$

where $\tilde{\sigma}(r)$ is the amplitude varying slowly with radius, and $[11.46 \ln(r/0.21) + m\varphi]$ represents the phase varying rapidly with r , that is,

$$|k_r|r \equiv 11.46 |(d/dr) \ln(r_n/0.21)| r \gg 1.$$

Equation (3.2) and the condition $\tilde{\sigma}(r) > 0$ on the initial phase imply that the maximum values of the perturbed density in Eq. (3.3) coincide with the positions of all the planets (Fig. 4a).

Interestingly, and this is the central part of our theory, the TB rule (Eq. (3.2)) *satisfies* the conditions of the WKB wave with the effective TB radial wavenumber (Eq. (3.3))

$$k_{\text{eff}} = 11.46 r^{-1}, \tag{3.4}$$

$d \ln k_{\text{eff}} / d \ln r = O(1)$, and $k_{\text{eff}} r \gg 1$ (cf. Eq. (2.2)). Note that Polyachenko (Polyachenko & Fridman 1972) has considered this analogy in his investigation of the possibility of the explanation of the law of planetary distances by the gravitational instability, but evidently without success. In particular, Polyachenko studied only axisymmetric perturbations, which do not carry angular momentum (see §4 below). Because $k_{\text{eff}} r \gg 1$, the short-wavelength ($\lambda_{\text{crit}}/R \ll 1$) WKB approximation used in the theory does not fail.

Thus, if the space dependence of the perturbed surface density of the protoplanetary disk has the form of Eq. (3.3) with $\Im\omega > 0$, the maxima of both radially and azimuthally unstable gravity perturbations are located in places of the solar system’s planets (Figs 4b, c, d). Let us define conditions under which the density maxima are localized on planetary orbits. If the disk is inhomogeneous with respect to equilibrium parameters, the radial wavelength of a perturbation with a maximum growth rate λ_{crit} will be a function of the radius r . From the above, the radial wavelength $\lambda_{\text{crit}} \approx 4\pi^2 G \sigma_0 / \kappa^2$, corresponding to the minimum on the dispersion curve (2.4) (see also Griv *et al.* 2002, their Fig. 1). On the other hand, the radial wavelength is $\lambda_{\text{eff}} = 2\pi / k_{\text{eff}}$. Comparing λ_{crit} with λ_{eff} , we see that in the case where the disk density is dependent on radius according to the law

$$\sigma_0(r) \approx 0.014 G^{-1} \kappa^2 r, \tag{3.5}$$

the maxima of time-increasing, both radially and azimuthally Jeans-unstable density perturbations are arranged in it according the TB rule. Interestingly, both optical and near-infrared observations of pre-main-sequence stars of intermediate mass have revealed the structure of spirals and rings, and thus presumably the Jeans instability of radial and spiral perturbations, in the circumstellar disks (Fukagawa *et al.* 2004).

One concludes, therefore, that if the surface density of a protoplanetary disk falls according to the law (3.5), the increasing maxima of density perturbations of a Safronov–Toomre unstable disk ($c < c_T$, or $Q < 1$, respectively) are located between the Lindblad resonances in places of the planets. We believe to have obtained a theoretical interpretation of the TB rule: the distance between planets is the wavelength of the most Jeans-unstable perturbations λ_{crit} at the given point of the protoplanetary disk.

4. The transfer of angular momentum

We next turn to the question of how to account for the concentration of angular momentum in the planets and of mass in the sun. The collective torque exerted by the gravity perturbations on the disk is $\Gamma = - \int \int d^2r (\vec{r} \times \vec{\nabla} \mathcal{N}_1) \sigma_1$ or

$$\Gamma = - \int_{r_1}^{r_2} r dr \int_0^{2\pi} \sigma_1(r, \varphi') \frac{\partial \mathcal{N}_1(r, \varphi')}{\partial \varphi'} d\varphi'. \tag{4.1}$$

The points r_1 and r_2 in which $\omega_* \pm \kappa = 0$ are called the points of inner and outer Lindblad resonances. They play an important role in the theory: the solution of spiral type (2.1) rapidly oscillating in the radial direction lies between r_1 and r_2 . Outside the resonances, $r < r_1$ and $r > r_2$, the solution decreases exponentially. A special analysis of the solution near spatially limited corotation ($\omega_* = 0$) and Lindblad ($\omega_* \pm \kappa = 0$) resonances is required. Resonances of a higher order, $\omega_* \pm l\kappa = 0$ and $l = 2, 3, \dots$, are dynamically of less importance (Shu 1970). The present analysis is restricted to consideration of only

the main part of a disk between the inner and outer Lindblad resonances. Lynden-Bell & Kalnajs (1972), Goldreich & Tremaine (1980), Meyer-Vernet & Sicardy (1987), and Griv *et al.* (2000) have investigated the wave-particle resonances.

Using Eq. (2.3), in terms of the Fourier components defined in Eq. (2.1), $\Gamma = \sum_{m=1}^{\infty} \Gamma_m$, from Eq. (4.1) one finds ($|\omega_*^2| < \kappa^2$)

$$\Gamma_m \approx -\frac{8\pi m^2}{\Omega \Im \omega_*} \aleph_1 \aleph_1^* \int_{r_1}^{r_2} dr \frac{\partial \sigma_0}{\partial r} \approx -8\pi \frac{m^2 \sigma_0}{\Omega \Im \omega_*} |\tilde{\aleph}|^2 e^{2\Im \omega_* t} \quad \text{if } \Im \omega_* > 0, \quad (4.2)$$

or $\Gamma_m = 0$ if $\Im \omega_* \leq 0$. In Eq. (4.2), \aleph_1^* is the complex conjugate potential, and the values of \aleph_1 , \aleph_1^* , σ_0 , Ω are evaluated at $r = r_1$. Four physical conclusions can be deduced from Eq. (4.2). (a) The distribution of the angular momentum of a disk will change under the action of only the nonaxisymmetric forces $\propto m$. The latter is obvious: axially symmetric motions of a system, studied by Polyachenko, produce no gravitational couplings between the inner parts and the outer parts. (b) The distribution of the angular momentum will change with time only under the action of growing, that is, Jeans-unstable perturbations ($\Im \omega_* > 0$). In the opposite limiting case, $\Im \omega_* \rightarrow 0$, absorption and emission of angular momentum are confined only to resonate particles (Lynden-Bell & Kalnajs 1972). (c) Unstable perturbations can transfer angular momentum only in an inhomogeneous disk ($\partial \sigma_0 / \partial r \neq 0$). And (d) $\Gamma_m < 0$: the spiral perturbations remove angular momentum from the disk. This takes place in the main part of the disk between the Lindblad resonances where spiral density waves are self-excited by means of a nonresonant wave-“fluid” interaction. This in turn cannot be done for all masses because the total orbital momentum must remain constant. As a result, the bulk of angular momentum is transferred outward whereas the bulk of mass is correspondingly transported inward; a relatively small group of outer particles with radii $r > r_2$ moves outward, taking almost all of the angular momentum. We speculate that a large portion of the initial mass of the nebula was transported toward the sun about 5×10^9 yr ago by the gravitational torque.†

Finally, let us evaluate the gravitational torque for a realistic model of the protoplanetary disk. In accordance with the theory developed above, the fastest growing spiral mode with $m \gtrsim 1$, $k_* = k_{\text{crit}}$, and $\Im \omega_* \sim \Omega$ is considered. Taking into account that $8\pi m^2 \aleph_1 \aleph_1^* \sim \aleph_0^2$ (an astrophysicist might well consider a perturbation with \aleph_1 / \aleph_0 of 1/10 or even 1/3 to be quite small) and $\aleph_0 \sim r^2 \Omega^2$, where \aleph_0 is the basic potential, from Eq. (4.2) one obtains $|\Gamma| \sim \sigma_0 r^4 \Omega^2$. The angular momentum of the disk $\mathcal{L} \sim \sigma_0 r^4 \Omega$. Then the characteristic time of the angular momentum redistribution is $t \sim \mathcal{L} / |\Gamma| \sim \Omega^{-1}$. Thus, already in the first three to four disk revolutions, in, say, about 10^4 yr, the gas-dust protoplanetary disk sees almost all of its angular momentum transferred outward and mass transported inward. With this efficient transport of angular momentum the sun would be able to grow in mass. We conclude that the Jeans instability studied here can give rise to torques that can help to clear the nebula on timescales of $10^5 - 10^6$ yr, in accord with astronomical requirements (Taylor 1992, §2.9.2). Besides, the analysis is found to imply the existence of new *planets* or other Kuiper-type *belts* of comets at mean distances from the sun of $r_{11} \approx 87$ AU, $r_{12} \approx 151$ AU, $r_{13} \approx 261$ AU, $r_{14} \approx 452$ AU, $r_{15} \approx 781$ AU (Mercury, Venus, ..., Asteroid belt, ..., Neptune, Kuiper belt, new planets or other Kuiper-type belts).

† Lynden-Bell & Kalnajs (1972, p. 6) have proved that the gravitational torques can only communicate angular momentum outward if the spirals trail. However, in contrast to the present study, Lynden-Bell & Kalnajs considered a model perturbation propagating in a gravitationally stable disk, $\Im \omega = 0$.

5. Discussion

With the paradigm of the standard accretion theory in crisis, the theory of “disk instability,” proposed back in the 1950s, is regaining popularity. It postulates the disintegration of a gas-and-dust protosolar nebula under the influence of local gravitational instabilities into massive fragments, which then collapse into planets (and the sun). In this work, we considered the rather neglected in the theory and numerical simulations Safronov–Toomre unstable disks of gas and dust with Toomre’s Q -values less than unity. Clearly, more detailed studies, in particular, simulations which include a realistic treatment of the system under study, are needed to definitely distinguish between different mechanisms of the solar system formation. In turn, observations may provide an indication on whether the Q -value for protostellar disks is indeed comparable to or less than unity. (Interestingly, observations have already indicated that the outer regions of accretion disks in both active galactic nuclei and young stellar objects are close to gravitational instability, e.g., Johnson & Gammie 2003).

Even though the analysis presented here shows that there is a dominant nonaxisymmetric Fourier mode of maximum instability with the wavelength λ_{crit} , the number of spiral arms m_{crit} , and the pitch angle ψ_{crit} in the protosolar nebula, at the present we cannot explain these quantities in the *local* WKB version of our theory. In this work we investigate the collective instabilities of the self-gravitating protoplanetary disk on the following fundamental assumption, i.e., the local WKB analysis. This assumption may have essential defects for the wave phenomena and instabilities of self-consistent systems though the physical mechanisms of the instabilities are well clarified (Alexandrov *et al.* 1984). Therefore we have to investigate the effects of nonlocality in the next step (Alexandrov *et al.* 1984, p. 249). In the local WKB approximation it is assumed that the wave vector and the wavefrequency vary continuously. We will show in the next paper of the series by utilizing the more accurate nonlocal WKB approximation that in fact the characteristic oscillation frequencies of an inhomogeneous disk must be “quantized,” i.e., must pass through a discrete series of values. According to the WKB method the spectrum of frequencies ω is determined by the quasiclassical rules of Bohr–Sommerfeld quantization:

$$\int_{r_1}^{r_2} k_r(r') dr' = \left(n + \frac{1}{2} \right) \pi, \quad (5.1)$$

where r_1, r_2 are the “reflection” points (say, the inner and outer Lindblad resonances) and $n = 0, 1, 2, \dots$. Equation (5.1) implies that ω is independent of r ; that is, the spiral-wave front is not distorted by differential rotation. The pattern is composed of one or more modes, and discrete modes are stationary in a rotating frame and hence do not wind up. It seems likely that such an approach will allow us to determine the critical wavelength λ_{crit} , the critical number of spiral arms m_{crit} , and the critical pitch angle ψ_{crit} . The weakly inhomogeneous approximation ($|kL| \gg 1$) used throughout this paper has the meaning that the discrete spectrum will differ little from a continuous spectrum, and in the zero approximation may be regarded as continuous. Further, the nonlocal theory provides a formal basis for the main idea that formed the basis for the local description, namely, that of short-wavelength approximation $|k_r|r \gg 1$.

Acknowledgements

We thank David Eichler, Yury Lyubarsky, Valery Polyachenko, Frank Shu, Raphael Steinitz, and Chi Yuan for valuable discussions. This work was supported in part by the Israel Science Foundation and the Israeli Ministry of Immigrant Absorption.

References

- Alexandrov, A.F., Bogdankevich, L.S., & Rukhadze, A.A. 1984, *Principles of Plasma Electrodynamics* (Berlin: Springer-Verlag)
- Bertin, G., & Casertano, S. 1982, *Astron. Astrophys.* 106, 274
- Boss, A.P. 2002, *Astrophys. J.* 576, 462
- Boss, A.P. 2003, *Astrophys. J.* 599, 577
- Boss, A.P. 2004, *Astrophys. J.* 610, 456
- Boss, A.P., Wetherill, G.W., & Haghighipour, N. 2002, *Icarus* 156, 291
- Cameron, A.G.W. 1978, *Moon Planets* 18, 5
- Cassen, P.M., Smith, B.F., Miller, R.H., & Reynolds, R.T. 1981, *Icarus* 48, 377
- Feigelson, E.D., & Montmerle, T. 1999, *ARA&A* 37, 363
- Fukagawa, M., Hayashi, M., Tamura, M., *et al.* 2004, *Astrophys. J.* 605, L53
- Goldreich, P., & Tremaine, S. 1980, *Astrophys. J.* 241, 435
- Griv, E., & Chiueh, T. 1998, *Astrophys. J.* 503, 186
- Griv, E., & Gedalin, M. 2003, *Planet. Space Sci.* 51, 899
- Griv, E., & Gedalin, M. 2004, *Astron. J.* 128, 1965
- Griv, E., Gedalin, M., & Eichler, D. 2001, *Astrophys. J.* (Letters) 555, L29
- Griv, E., Gedalin, M., Eichler, D., & Yuan, C. 2000, *Phys. Rev. Lett.* 84, 4280
- Griv, E., Gedalin, M., & Yuan, C. 2002, *Astron. Astrophys.* 383, 338
- Griv, E., Gedalin, M., & Yuan, C. 2003, *Mon. Not. R. Astron. Soc.* 342, 1102
- Griv, E., Yuan, C., & Gedalin, M. 1999, *Mon. Not. R. Astron. Soc.* 307, 1
- Johnson, B.M., & Gammie, C.F. 2003, *Astrophys. J.* 597, 131
- Kuiper, G.P. 1951, *Proc. Natl. Acad. Sci.* 37, 1
- Kulsrud, R.M., Mark, J.W.-K., & Caruso, A. 1971, *Astrophys. Space Sci.* 14, 52
- Larson, R.B. 1989, in: H.A. Weaver & L. Danly (eds.), *The Formation and Evolution of Planetary Systems*, Cambridge: Cambridge Univ. Press, p. 31
- Liboff, R.L. 2004, *Astron. J.* 126, 3132
- Lin, C.C., & Lau, Y.Y. 1979, *SIAM. Stud. Appl. Math.* 60, 97
- Lin, C.C., & Shu, F.H. 1966, *Proc. Natl. Acad. Sci.* 55, 229
- Lin, C.C., Yuan, C., & Shu, F.H. 1969, *Astrophys. J.* 155, 721
- Luu, J.X., & Jewitt, D.C. 2002, *ARA&A* 40, 63
- Lynch, P. 2003, *Mon. Not. R. Astron. Soc.* 341, 1174
- Lynden-Bell, D., & Kalnajs, A.J. 1972, *Mon. Not. R. Astron. Soc.* 157, 1
- Mayer, L., Quinn, T., Wadsley, J., & Stadel, J. 2002, *Science* 298, 1756
- Meyer-Vernet, N., & Sicardy, B. 1987, *Icarus* 69, 157
- Morozov, A.G. 1980, *Soviet Astron.* 24, 391
- Morozov, A.G. 1981, *Soviet Astron.* 25, 421
- Neslušan, L. 2004, *Mon. Not. R. Astron. Soc.* 351, 133
- Pollack, J.B., Hubickyj, O., Bodenheimer, P., *et al.* 1996, *Icarus* 124, 62
- Polyachenko, V.L., & Fridman, A.M. 1972, *Soviet Astron.* 16, 123
- Rice, W.K.M., Armitage, P.J., Bonnell, I.A., *et al.* 2003, *Mon. Not. R. Astron. Soc.* 346, L36
- Safronov, V.S. 1960, *Ann. d'Astrophys.* 23, 979
- Safronov, V.S. 1972, *Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets* (Jerusalem: Israel Program for Scientific Translations)
- Safronov, V.S. 1980, in: D. Lal (ed.), *Early Solar System Processes* (Amsterdam: North-Holland), p. 73
- Shu, F.H. 1970, *Astrophys. J.* 60, 99
- Taylor, S.R. 1992, *Solar System Evolution*. Cambridge Univ. Press
- Toomre, A. 1964, *Astrophys. J.* 139, 1217
- Urey, H.C. 1958, *Proc. Chem. Soc. London*, 67 (March)
- Wetherill, G.W. 1989, in: H.A. Weaver & L. Danly (eds.), *The Formation and Evolution of Planetary Systems*, Cambridge Univ. Press, p. 1