CAN GIANT PLANETS FORM BY DIRECT GRAVITATIONAL INSTABILITY?

ROMAN R. RAFIKOV

Received 2004 June 21; accepted 2005 January 18; published 2005 January 31

ABSTRACT

Gravitational instability has been invoked as a possible mechanism of the giant planet production in protoplanetary disks. Here we critically revise its viability by noting that to form planets directly, it is not enough for protoplanetary disks to be gravitationally unstable. They must also be able to cool efficiently (on a timescale comparable to the local disk orbital period) to allow the formation of the bound clumps by fragmentation. A combination of the dynamical and thermal constraints puts very stringent lower limits on the properties of disks capable of fragmenting into the self-gravitating objects: for the gravitational instability to form giant planets at 10 AU in the disk cooled by the radiation transfer, the gas temperature must exceed $10^4$ K with a minimum disk mass of 0.7 $M_\odot$ and a luminosity of 40 $L_\odot$. Although these requirements are relaxed in the more distant parts of the disk, masses of the bound objects formed as a result of instability are too large even at 100 AU ($\sim 10 M_\odot$) to explain the characteristics of known extrasolar giant planets. Such protoplanetary disks (and planets formed in them) have very unusual observational properties, and this severely constrains the possibility of giant planet formation by direct gravitational instability.

Subject headings: planets and satellites: formation — solar system: formation

1. INTRODUCTION

Recent discoveries of Jupiter-like planets around solar-type stars have rejuvenated the interest in the issue of the origin of giant planets. The core instability scenario (Perri & Cameron 1974; Mizuno 1980) postulates that Jupiter-like planets acquire their massive gaseous atmospheres by unstable gas accretion onto the preexisting massive solid cores. For a rather long time, this avenue of planet formation did not seem compatible with the short observed lifetimes ($10^6-10^7$ yr) of protoplanetary disks because of the long time needed for the core accumulation. However, in their recent work, Rafikov (2004a) and Goldreich et al. (2004) found the core formation time to be quite short, considerably alleviating the timescale issue in the core instability scenario. Nevertheless, gravitational instability (GI) has been put forward as an alternative mechanism of giant planet formation (Cameron 1978; Boss 1998). In this model, a massive gaseous disk becomes gravitationally unstable and rapidly fragments into a number of self-gravitating bound structures, which further collapse to become giant planets. A number of recent hydrodynamical simulations (Boss 1998; Mayer et al. 2002) employing an isothermal equation of state (EOS) have confirmed this general picture. The goal of this study is to constrain this avenue of planet formation by putting special emphasis on the thermodynamical conditions required for disk fragmentation.

2. DYNAMICAL AND THERMAL CONSTRAINTS

In a Keplerian disk, GI can operate only when the gas surface density $\Sigma$, angular frequency $\Omega$, and sound speed $c_s = (kT/\mu)^{1/2}$ ($T$ is the gas temperature, $\mu$ is its molecular weight, and $k$ is a Boltzmann constant) satisfy the condition

$$Q \equiv \frac{\Omega c_s}{\pi G \Sigma} < Q_0$$

(Safronov 1960; Toomre 1964). Here $Q$ is the so-called Toomre $Q$; analytical arguments and results of numerical simulations suggest that $Q_0 \approx 1$. An equivalent way of formulating the condition for GI to work is

$$\rho \equiv \frac{\Omega^2}{\pi G Q_0} = 1.9 \times 10^{-7} \text{ g cm}^{-3} a_{\text{AU}}^{-3} Q_0^{-1}$$

where $\rho$ is a midplane gas density and $a_{\text{AU}} \equiv a/(1 \text{ AU})$.

Dynamical constraint (1) is a necessary condition for GI to set in. However, even when it is fulfilled, planet formation occurs only if the disk can actually fragment into bound self-gravitating objects. Recent studies (Gammie 2001; Rice et al. 2003) have shown fragmentation to be possible only if the disk cooling time $t_{\text{cool}}$ satisfies

$$\Omega t_{\text{cool}} < \xi,$$

where $\xi \sim 1$ is a parameter; numerical simulations using constant $t_{\text{cool}}$ suggest that $\xi \approx 3$ (Gammie 2001; Rice et al. 2003). The cooling time of the disk is estimated as the ratio of its thermal energy to the escaping radiative flux:

$$t_{\text{cool}} \approx \frac{\Sigma c_s^2}{\gamma - 1} \frac{f(\tau)}{2\sigma T^4}, \quad f(\tau) = \tau + \frac{1}{\tau},$$

where $\sigma$ is the Stephan-Boltzmann constant, $T$ is the midplane temperature, $\tau \approx k\Sigma/2$ is the disk optical depth ($k$ is the opacity), and $\gamma$ is the adiabatic index of gas. The function $f(\tau)$ describes the efficiency of cooling that is most effective when $\tau \approx 1$. It is important for us that in a disk that cools by radiative losses from the photosphere, $f(\tau)$ is above unity for any $\tau$ and any mechanism of energy transfer from the midplane to the photosphere, since the effective temperature of such a disk cannot exceed its midplane temperature. A specific form of $f(\tau)$ in equation (4) assumes radiative energy transfer within the disk (for simplicity, we do not differentiate between Rosseland and Planck mean opacities).

Dynamical constraint (1) sets an upper limit on $c_s$. On the other hand, expressing $T$ in equation (4) through $c_s$, we find that equation (3) sets a lower limit on the sound speed (in the
optically thick regime, this is true whenever opacity $\kappa$ does not grow with temperature faster than $T^2$; see § 3). A combination of these conditions leads to the following constraint on $c_s$, necessary for planet formation by GI:

$$
\left[ \frac{f(\tau)}{\tilde{\xi}} \frac{\Omega}{\sigma} \left( \frac{k}{\mu} \right) \right]^{4/5} \lesssim c_s \lesssim \pi Q_0 \frac{G \Sigma}{\Omega},
$$

(5)

where $\tilde{\xi} \equiv 2 \tilde{\xi} (\gamma - 1) \approx 1$ (we also allow $\tilde{\xi}$ to absorb any additional numerical factors appearing in a more careful calculation of $\Sigma_{\text{inf}}$). Only when condition (5) is fulfilled can the disk be gravitationally unstable and its cooling be fast enough for fragmentation to produce bound gaseous clumps, which later collapse to become planets. A similar argument has been advanced by Levin (2003) in application to self-gravitating disks of active galactic nuclei.

At a specific location in the protoplanetary disk, the condition (5) is satisfied only if

$$
\Sigma \geq \Sigma_{\text{min}} = \Sigma_{\text{inf}} [f(\tau)]^{4/5},
$$

(6)

where

$$
\Sigma_{\text{inf}} \equiv \tilde{\Omega} \tilde{\xi} (\pi Q_0 \Omega) -^{4/5} \frac{1}{\tilde{\xi} Q_0 \Omega} \left( \frac{k}{\mu} \right)^{4/5}
\approx 6.6 \times 10^3 \text{ cm}^{-2} a_{\text{dH}}^{2/5} (Q_0 Q_\text{dH} \tilde{\xi})^{-1/5} \left( \frac{M_a}{M_\odot} \right)^{3/10}.
$$

(7)

Here $\tilde{\mu}/m_H$ is the mean molecular weight relative to the atomic hydrogen mass $m_H$ and $M_a$ is the mass of the central star ($M_\odot$ is the solar mass). For molecular gas of solar composition, $\Sigma_{\text{min}} \approx 3.4 \times 10^5$ cm$^{-2}$ at 1 AU for $f(\tau)/(Q_0 \tilde{\xi}) = 1$; for atomic hydrogen, $\Sigma_{\text{min}} \approx 5.6 \times 10^5$ cm$^{-2}$ at 1 AU. In both cases, $\Sigma$ (1 AU) exceed that in the minimum-mass solar nebula (Hayashi 1981) by $\approx 10^5$, implying that giant planet formation by GI is possible only in a very massive disk.

According to condition (5), whenever equation (7) is fulfilled, $c_s$ is also bounded from below:

$$
c_s \geq c_{s,\text{min}} = \tilde{\Omega}^{2/5} \left[ \frac{f(\tau)}{\tilde{\xi} Q_0 \Omega} \left( \frac{k}{\mu} \right)^{4/5} \right]^{1/5}
\approx 6.9 \text{ km s}^{-1} a_{\text{dH}}^{-3/5} \left[ \frac{f(\tau)}{Q_0 Q_\text{dH} \tilde{\xi}} \right]^{1/5} \left( \frac{M_a}{M_\odot} \right)^{1/5}.
$$

(8)

As a result, the midplane temperature has to satisfy

$$
T \geq T_{\text{min}} = T_{\text{inf}} [f(\tau)]^{2/5},
$$

(9)

where

$$
T_{\text{inf}} \equiv \tilde{\Omega}^{4/5} (\tilde{\xi} Q_0 \Omega)^{2/5} \left( \frac{k}{\mu} \right)^{3/5}
\approx 5800 \text{ K} a_{\text{dH}}^{6/5} \tilde{\xi}^{3/5} \left( Q_0 \tilde{\xi} \right)^{-2/5} \left( \frac{M_a}{M_\odot} \right)^{2/5}.
$$

(10)

Using equations (6)–(10), one can set lower limits on disk luminosity and mass that can be compared with observations.

The smallest lower limit on the disk mass can be set if planets are produced only at a single location $a$ from the central star. The scale length of the fastest-growing perturbation in the marginally gravitationally unstable disk ($Q \approx 1$) is $\lambda \approx 2\pi h$, where $h = c_s/\Omega$ is the vertical disk scale height. To form planets by GI, it is sufficient that only an annulus of the disk with the radial width $\lambda$ has the surface density and temperature exceeding $\Sigma_{\text{min}}$ and $T_{\text{min}}$. Using equation (8), we estimate

$$
\frac{h}{a} > \frac{1}{\sqrt{\pi} \sigma} \frac{\left[ f(\tau) \left( \frac{k}{\mu} \right) \right]^{3/5}}{\tilde{\xi} Q_0 \tilde{\xi} \Omega} \approx 0.23 a_{\text{dH}}^{1/10} \left[ \frac{f(\tau)}{\mu} \right]^{3/5} \left( \frac{M_a}{M_\odot} \right)^{3/10}.
$$

(11)

This and all subsequent numerical estimates assume $Q_0 = 1$ and $\tilde{\xi} = 1$. Because of the high midplane temperature, the disk is rather thick, and GI is likely to have a global character. The mass of a “minimum planet-forming annulus” $M_a$ centered on $a$ and having width $\lambda$ is

$$
M_a \geq 0.8 M_\odot a_{\text{dH}}^{-8/5} \left[ \frac{f(\tau)}{\mu} \right]^{3/5} \left( \frac{M_a}{M_\odot} \right)^{2/5}.
$$

(12)

This limit on disk mass depends very weakly on the semimajor axis of the annulus. $M_a$ amounts to a sizable fraction of $M_a$ [even for a molecular annulus at 1 AU, $M_a \approx 0.2 M_\odot$ for $f(\tau) = 1$]. The total disk mass should be even higher (easily by a factor of several) because $M_a$ accounts only for the mass of an unstable and fragmenting annulus. Besides, in the real nebula, such an annulus can be either very optically thick or thin, which translates into a large value of $f(\tau)$, additionally increasing the limit on $M_a$ (see below). Typical masses of observed protoplanetary disks vary between $10^{-3}$ and 0.1 $M_\odot$ (Kitamura et al. 2002), but these masses of gas are typically extended over $\approx 100$ AU (not over just a narrow annulus).

The luminosity of a unit surface area of the disk is

$$
\frac{dL}{dS} = \frac{2 a_{\text{dH}}^2}{f(\tau)} > 2 a_{\text{dH}}^{16/5} \left[ \frac{a_{\text{dH}}}{Q_0} \right]^{1/5} \left[ \frac{f(\tau)}{\mu} \right]^{3/5} \left( \frac{M_a}{M_\odot} \right)^{2/5}.
$$

(13)

where we made use of equation (10). Since $dL/dS \propto a^{-24/5}$, most of the energy is radiated at the inner edge of the gravitationally unstable region of the disk. The total luminosity of the annulus with an inner cutoff at $a$ is

$$
L = 1.6 \times 10^4 L_a a_{\text{dH}}^{14/5} \left[ \frac{f(\tau)}{\mu} \right]^{3/5} \left( \frac{M_a}{M_\odot} \right)^{8/5}.
$$

(14)

This estimate holds even for the full disk (rather than an annulus), which is a direct consequence of the very steep dependence of $dL/dS$ on the inner edge radius.

We now look at what these limits imply for disk properties at different locations in the protoplanetary nebula.

3. APPLICATION TO PROTOPLANETARY DISKS

Constraints on disk surface density and temperature (see eqs. [6]–[10]) still depend on the opacity behavior. However, bearing in mind that $f(\tau) > 1$ for any $\tau$, one immediately sees that $\Sigma_{\text{inf}}$ and $T_{\text{inf}}$ (from infimum—the greatest lower bound of a set) represent absolute opacity-independent lower limits on $\Sigma$ and $T$. Thus, fragmentation of a gravitationally unstable disk (necessary for giant planet formation) cannot occur unless at least $\Sigma > \Sigma_{\text{inf}}$ and $T > T_{\text{inf}}$. These lower limits are very robust as they do not depend on a specific mechanism of energy transfer in the disk (by radiation or convection).

In practice, using even the most basic properties of energy transport in the disk, one can formulate more stringent con-
straints. In this study, we will assume that energy is transferred by radiative diffusion. Then, using the opacity dependence $\kappa(\rho, T)$ on temperature $T$ and gas density $\rho$ (which is fixed by the condition [2]), we can set $\Sigma = \Sigma_{\text{min}} = T = T_{\text{min}}$ (minimum requirements for planet formation by GI), substitute $\tau = \kappa(\rho, T)\Sigma/2$ into equations (6) and (9), and solve for $\Sigma_{\text{min}}$ and $T_{\text{min}}$. As we mentioned before, if $\kappa$ increases with $T$ faster than $T^2$, then equation (3) does not constrain $T$ from below. Using opacities from Bell & Lin (1994), we find this to be the case for $1700 \lesssim T \lesssim 10^4$ K, in the region where molecular or H-scattering opacities dominate. However, whenever a solution for $\Sigma_{\text{min}}$ and $T_{\text{min}}$ in this temperature range exists, there is always another, lower temperature solution in the region where the dependence of $\kappa$ on $T$ is less steep than $T^2$. This last extreme solution does limit from below the disk temperature necessary for fragmentation, and we will use it in our discussion.

3.1. Limits on the Disk Properties at 1 AU

First we consider the possibility of planet formation by GI in the terrestrial planet region, at $a = 1$ AU. Constraints obtained in § 2 imply that planetary genesis at this location requires quite extreme disk properties: $T > T_{\text{inf}} = 5.2 \times 10^4$ K ($\kappa = 1.2$ in eq. [10]) since gas cannot be molecular at such temperature, $\Sigma > \Sigma_{\text{inf}} = 5.7 \times 10^3$ g cm$^{-2}$, and $L > 10^2 L_\odot$ (all assuming $\zeta = 1$ and $Q_0 = 1$). The disk mass must exceed $M_s \approx 0.6 M_\odot$. Apparently, even these conservative bounds strongly disagree with the observed properties of protoplanetary disks.

These limits are improved by noticing that a disk with so high a $\Sigma$ is very optically thick, meaning that $\Sigma_{\text{min}}$ and $T_{\text{min}}$ set better constraints than $\Sigma_{\text{inf}}$ and $T_{\text{inf}}$. Using opacities from Bell & Lin (1994), we find that at 1 AU, the disk has to be extremely hot, and the opacity has to be due to electron scattering. The temperature would exceed 10$^6$ K, and such a disk would not be bound to the star. This firmly rules out the possibility of planet formation by GI at $\approx 1$ AU.

3.2. Limits on the Disk Properties at 10 AU

At 10 AU, in the region of giant planets, the temperature is probably low enough for the gas to be molecular. Using $\kappa = 2.3$, we find $\Sigma_{\text{inf}} = 2.7 \times 10^2$ g cm$^{-2}$ and $T_{\text{inf}} = 220$ K. The disk luminosity has to exceed only 3.4 $L_\odot$, and the minimum-mass annulus has to contain at least $M_s = 0.13 M_\odot$ of gas. These limits are more reasonable than at 1 AU, although the mass is still high and the disk is too hot compared to the observed systems, which typically contain less than 0.1 $M_\odot$ of gas and have $T \lesssim 10^4$ K at 10 AU (e.g., Kitamura et al. 2002).

However, using opacities from Bell & Lin (1994) and $\rho$ as given by condition (2), we find that real disks (i.e., not cooling at the maximum efficiency) are considerably more extreme. There are several possible solutions for $T_{\text{min}}$ and $\Sigma_{\text{min}}$ at 10 AU corresponding to different opacity regimes. The least extreme is a “cold” solution with $\tau \approx 60$, $T_{\text{min}} \approx 1100$ K, $\Sigma_{\text{min}} \approx 6 \times 10^2$ g cm$^{-2}$, $L \approx 40 L_\odot$, and $M_s \approx 0.7 M_\odot$ corresponding to molecular gas at the temperature of grain evaporation (other solutions have $T_{\text{min}} \approx 7 \times 10^4$ K). This disk is too massive and hot to satisfy current observational constraints. Thus, it is unlikely that GI can allow disk fragmentation and subsequent giant planet formation even at 10 AU.

3.3. Limits on the Disk Properties at 100 AU

We also look at the possibility of giant planet formation by GI in the distant regions of the protoplanetary nebula, at $a = 100$ AU. For the molecular disk at this location, we find $\Sigma_{\text{inf}} = 20$ g cm$^{-2}$, $T_{\text{inf}} = 14$ K, $L \approx 5 \times 10^{-3} L_\odot$, and $M_s \approx 0.08 M_\odot$. These properties change only a little when the inefficiency of disk cooling is properly accounted for: using $\kappa \approx 0.1(7/10)^{-1}$ K cm$^{-2}$ g$^{-1}$ in agreement with observations of protoplanetary nebulae (Beckwith et al. 1990; Kitamura et al. 2002), we find that to support planet formation by GI, the disk has to possess at least $\tau \approx 2$ (marginally optically thick), $T_{\text{min}} \approx 20$ K, $\Sigma_{\text{min}} \approx 25$ g cm$^{-2}$, $L \approx 10^{-3} L_\odot$, and $M_s \approx 0.1 M_\odot$ at 100 AU. Although $M_s$ is very near the upper end of the observed distribution of the protoplanetary disk masses, these parameters seem to be acceptable from an observational point of view.

3.4. Fragment Masses

Another important constraint casting doubt on the possibility of planet formation by GI even at 100 AU comes from comparing observed masses of extrasolar giant planets with the typical masses of fragments into which the disk breaks up when $t_\text{cool} < \xi \Omega^{-1}$. Since the length scale of the most unstable mode is $\lambda \approx 2\pi h$, the minimum fragment mass is

$$M_f \approx \Sigma_{\text{min}} \lambda^2 \approx 0.15 M_\odot a_{\text{AU}}^{-3/10} \left(\frac{f(\tau)}{M_\odot}\right)^{1/5} \left(\frac{M_\odot}{M_f}\right)^{1/10}. \quad (15)$$

At 100 AU, a molecular disk meeting the requirements outlined in § 3.3 would fragment into self-gravitating clumps with the mass of roughly $(5M_\odot)^{1/5}$, where $M_f$ is the Jupiter mass; at smaller semimajor axes, $M_f$ would be higher. Even for $f(\tau) = 1$, this mass is larger than the masses of most extrasolar giant planets detected to date (Marcy et al. 2003). A more realistic cooling efficiency corresponding to the optical depth of $\tau \approx 2$ at 100 AU leads to $M_f \approx 9M_f$, landing the minimum fragment mass not too far from the brown dwarf regime.

4. DISCUSSION

The novel analytical constraints presented in §§ 2 and 3, when confronted with observations of protoplanetary disks, severely undermine the possibility of giant planet formation by GI. We have shown that disks capable of producing giant planets by GI at $\approx 1$ AU cannot exist on dynamical grounds—to cool efficiently, they must be too hot to be bound to the central star. This rules out the possibility of an in situ formation of close-in extrasolar giant planets by GI. Rafikov (2004b) has also presented arguments against an in situ formation of such “hot Jupiters” via the core instability. Thus, the most natural way to explain the existence of the close-in giant planets is to accept that they formed elsewhere under more favorable conditions and then migrated to their current locations.

Planet formation by GI is also extremely unlikely within several tens of AU: disks with required properties must be so hot ($T \approx 10^4$ K), luminous (several tens of solar luminosity), and massive ($\approx 1 M_\odot$) that they would clearly stand out in a sample of observed protoplanetary nebulae. Beyond $\approx 10^2$ AU, the observed disk properties would enable planets to form by GI, although the required disk masses still reside at the very top of their observed distribution. However, it is still very unlikely that known extrasolar giant planets could have been produced there by GI. The problem is not only in the potential difficulty of migrating planets from $\approx 10^2$ AU all the way in to several AU, but also in forming them with the right mass. As our estimate (eq. [15]) shows, objects produced by GI even at $10^2$ AU are too massive ($\approx 10M_\odot$) to explain the observed mass distribution of extrasolar planets.
Our study emphasizes the importance of the proper treatment of the disk thermodynamics (especially its cooling) for determining the possibility of planet formation by GI. Virtually all simulations that were able to demonstrate the fragmentation of gravitationally unstable disks and the collapse of the resulting dense objects have used the isothermal EOS (e.g., Mayer et al. 2002; but see Boss 2004). This, however, is equivalent to artificially setting \( t_{\text{cool}} = 0 \), which relaxes one of the requirements for planet production and is misleading. Not surprisingly, calculations following thermodynamics in greater detail typically do not exhibit the fragmentation of gravitationally unstable disks that are not capable of cooling efficiently (Pickett et al. 1998; Gammie 2001; Rice et al. 2003). Thus, isothermal simulations should not be used to study planet formation in real protoplanetary disks.

Future infrared and millimeter observations will show whether or not the protoplanetary disks with extreme properties conducive to giant planet formation by GI really exist.

I am grateful to Yuri Levin, who has been working on similar subjects, and Lynne Hillenbrand for illuminating discussions. The thoughtful comments by Charles Gammie were deeply appreciated. The author is a Frank and Peggy Taplin Member at the Institute for Advanced Study; he is also supported by the W. M. Keck Foundation and NSF grant PHY-0070928.

REFERENCES

Beckwith, S. V. W., Sargent, A. I., Chini, R. S., & Guesten, R. 1990, AJ, 99, 924
Bell, K. R., & Lin, D. N. C. 1994, ApJ, 427, 987
Boss, A. P. 1998, ApJ, 503, 923
———, 2004, ApJ, 610, 456
Cameron, A. G. W. 1978, Moon Planets, 18, 5
Gammie, C. F. 2001, ApJ, 553, 174
Goldreich, P., Lithwick, Y., & Sari, R. 2004, ApJ, 614, 497
Hayashi, C. 1981, Prog. Theor. Phys. Suppl., 70, 35
Kitamura, Y., Momose, M., Yokogawa, S., Kawabe, R., Tamura, M., & Ida, S. 2002, ApJ, 581, 357
Levin, Yu. 2003, preprint (astro-ph/0307084)
Marcy, G. W., Butler, R. P., Fischer, D. A., & Vogt, S. S. 2003, in ASP Conf. Ser. 294, Scientific Frontiers in Research on Extrasolar Planets, ed. D. Deming & S. Seager (San Francisco: ASP), 1

Mayer, L., Quinn, T., Wadsley, J., & Stadel, J. 2002, Science, 298, 1756
Mizuno, H. 1980, Prog. Theor. Phys., 64, 544
Perri, F., & Cameron, A. G. W. 1974, Icarus, 22, 416
Pickett, B., Cassen, P., Durisen, R. H., & Link, R. 1998, ApJ, 504, 468
Rafikov, R. R. 2004a, AJ, 128, 1348
———, 2004b, ApJ, submitted (astro-ph/0405507)
Rice, W. K. M., Armitage, P. J., Bate, M. R., & Bonnell, I. A. 2003, MNRAS, 339, 1025
Safronov, V. S. 1960, Ann. d’Astrophys., 23, 979
Toomre, A. 1964, ApJ, 139, 1217