

Planet formation and migration

To cite this article: John C B Papaloizou and Caroline Terquem 2006 *Rep. Prog. Phys.* **69** 119

View the [article online](#) for updates and enhancements.

You may also like

- [TEN NEW AND UPDATED MULTIPLANET SYSTEMS AND A SURVEY OF EXOPLANETARY SYSTEMS](#)
J. T. Wright, S. Upadhyay, G. W. Marcy et al.
- [A Comparison of the Composition of Planets in Single-planet and Multiplanet Systems Orbiting M dwarfs](#)
Romy Rodríguez Martínez, David V. Martin, B. Scott Gaudi et al.
- [Ultra-short-period Planets in K2. III. Neighbors are Common with 13 New Multiplanet Systems and 10 Newly Validated Planets in Campaigns 0–8 and 10](#)
Elisabeth R. Adams, Brian Jackson, Samantha Johnson et al.

Planet formation and migration

John C B Papaloizou^{1,2} and Caroline Terquem^{3,4,5,6}

¹ Astronomy Unit, School of Mathematical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, UK

² Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK

³ Institut d'Astrophysique de Paris, UMR7095 CNRS, Université Pierre & Marie Curie–Paris 6, 98bis boulevard Arago, 75014 Paris, France

⁴ Université Denis Diderot–Paris 7, 2 Place Jussieu, 75251 Paris Cedex 5, France

⁵ Institut Universitaire de France

E-mail: terquem@iap.fr

Received 9 June 2005, in final form 3 October 2005

Published 9 November 2005

Online at stacks.iop.org/RoPP/69/119

Abstract

We review the observations of extrasolar planets, ongoing developments in theories of planet formation, orbital migration and the evolution of multiplanet systems.

(Some figures in this article are in colour only in the electronic version)

⁶ Author to whom any correspondence should be addressed.

Contents

	Page
1. Introduction	122
2. Properties of exoplanet systems	122
2.1. Detection methods	123
2.1.1. Radial velocity	123
2.1.2. Transits	124
2.1.3. Astrometry	124
2.1.4. Microlensing	125
2.1.5. Direct imaging	126
2.2. Statistical properties of observed extrasolar planets	126
2.2.1. Mass and separation distributions	126
2.2.2. Eccentricity distribution	128
2.2.3. Metallicity of stars harbouring planets	129
2.2.4. Planets in multiple stellar systems	130
2.2.5. Free floating planets	130
3. Theories of planet formation	130
3.1. Terrestrial planet and planetary core formation	130
3.1.1. Growth from dust to centimetre per metre sized particles	131
3.1.2. Growth from centimetre per metre sized particles to 0.1/1 km sized planetesimals	132
3.1.3. Accretion of a planetary core from planetesimals	135
3.2. Formation of giant planets: the core accretion model	144
3.2.1. Equations governing an envelope at equilibrium	145
3.2.2. Models enveloped by the disc with luminosity derived from planetesimal accretion	147
3.2.3. Models enveloped by the disc with no planetesimal accretion	149
3.2.4. Models with free boundary accreting from the protostellar disc	149
3.3. Formation of giant planets: gravitational instabilities	153
4. Disc–planet interactions and orbital migration	155
4.1. Angular momentum transfer due to scattering and type II migration	156
4.1.1. The impulse approximation applied to a protoplanet in a gap	156
4.1.2. Angular momentum exchange rate with the disc	157
4.1.3. Gap formation	157
4.1.4. Type II migration	159
4.1.5. Stopping type II migration	160
4.2. Linear torques and type I migration	161
4.2.1. Timescales	161
4.2.2. Type I migration in a turbulent disc	164
4.3. Coorbital torques and type III migration	166
5. Multiplanet systems and their interactions	169
5.1. Orbital relaxation in young planetary systems	169

5.2. Effect of an outer relaxing distribution of protoplanets on a low-mass planet formed in an inner disc	170
6. Concluding remarks	171
Acknowledgments	173
Appendix. Disc-protoplanet torques and angular momentum conservation	173
A.1. Conservation of angular momentum	174
A.2. A migrating protoplanet	175
A.3. Conservation of angular momentum	175
References	177

1. Introduction

At the time of the writing of this review, the tenth anniversary of the discovery of the first extrasolar planet around a solar-type star is being celebrated. When this detection was announced, it came as a shock, not because an extrasolar planet was at last found around a star like our Sun—nobody really had doubts that such a detection would occur one day or another—but because of the very short period of this planet, 51 Pegasi. It orbits its parent star in only 4 days! It is 100 times closer to the star than Jupiter is from the Sun. The temperatures at these distances from the star are so high that the material needed to assemble the core of such planets cannot be found in a solid phase there. The implication is that the planet formed further out and moved in. It did not take long for the theorists to propose an explanation: tidal interactions between the planet and the disc in which it formed led to orbital migration. The theory of such processes had already been developed, about 15 years before the discovery of 51 Pegasi, and orbital migration of planets had been predicted. The real surprise, then, was that nobody ever predicted that planets might be found on very short period orbits.

As of today, about 150 planets have been detected. Their orbital characteristics suggest that planet/disc tidal interactions is very important in determining the characteristics of planetary systems. This review is divided in four sections. The first section is devoted to the observations of extrasolar planets. We first present the different methods used to detect planets, and review the statistical properties of the systems observed so far. In the second section, we review the theories of planet formation. We describe the different stages that lead to the build-up of a terrestrial planet or a planetary core (grain sedimentation, obtention of planetesimals, runaway accumulation), and the process by which an envelope of gas may be captured to produce a giant planet. We also discuss giant planet formation by gravitational instabilities. The third section is devoted to disc/planet interactions and the different types of orbital migration. Finally, in the fourth section, we review some results about multiplanet systems and their interactions.

2. Properties of exoplanet systems

The first object outside our solar system with a mass in the range of planetary masses was detected in 1992 around the millisecond pulsar PSR1257 + 12a (Wolszczan and Frail 1992). Three planets of 0.02, 4.3 and 3.9 Earth masses orbiting at 0.19 AU, 0.36 AU and 0.46 AU (astronomical units) from the pulsar, respectively, were actually found. Another object of about 100 Earth masses with a separation of 40 AU was reported more recently (Joshi and Rasio 1997, Wolszczan 1996). In 1993, Backer *et al* announced the detection of a 2.5 Jupiter mass (M_J) planet orbiting the millisecond pulsar 1620-26 at a distance of 23 AU. This pulsar is in a binary system with a white dwarf companion (Thorsett *et al* 1999) and a separation of about 40 AU. It is not clear whether these planets formed before the explosion of the parent supernova or after, either from the material which was ejected during the explosion or, in the case of the pulsar 1620-26, in the disc of the material transferred from the white dwarf.

The first extrasolar planet around a sun-like star (51 Pegasi b) was reported in 1995 by Mayor and Queloz. It was detected from Haute-Provence Observatory and immediately confirmed by Marcy and Butler (1995), who observed it from Lick Observatory (see the review by Marcy and Butler (1998) for more details about the detection of the first eight extrasolar planets). Since then there has been an uninterrupted flow of detections. As of today (16 March 2005), 152 planets in 134 planetary systems (including 14 multiple planet systems) have been reported around solar type stars. The lightest object has a mass of $0.042 M_J$. We will focus here on the objects with masses smaller than $13 M_J$, as heavier objects are usually considered

to be brown dwarfs. This review is concerned with extrasolar planets around solar type stars, so we will not discuss further planets around pulsars.

2.1. Detection methods

2.1.1. Radial velocity. So far, almost all the planets have been detected by this technique, which consists in measuring the Doppler shift due to the motion of the star around the centre of mass of the system. The velocity of the star projected along the line of sight (or *radial velocity*) is, to within a constant, $v = K f(t)$, where the function $f(t)$ varies periodically with time t around the orbit and has zero mean and

$$K = \left(\frac{2\pi G}{T} \right)^{1/3} \frac{M_p \sin i}{(M_p + M_\star)^{2/3}} \frac{1}{\sqrt{1 - e^2}}. \quad (1)$$

Here G is the constant of gravitation, T is the orbital period, M_\star is the mass of the star, M_p is the planet mass, i is the angle of the line of sight with respect to the perpendicular to the orbital plane and e is the eccentricity of the orbit. The observation of v as a function of time gives T directly. If the orbit is circular, $f(t)$ is a sinusoid with unit amplitude. For non-circular orbits the amplitude can reach $1 + e$ and the departure from a sinusoid allows the determination of the eccentricity e . Given T and e , one can deduce $M_p \sin i$ from the measurement of K (M_\star being known). Note that since $M_p \ll M_\star$, the semi-major axis a can be calculated from T .

Since i is unknown, only the lower limit $M_p \sin i$ for the planet mass, or *projected mass*, can be obtained. This was of course an important matter of debate when the very first planet was detected. However, for a population of about 150 objects, the ratio of the real mass to the projected mass is on average on the order of unity. The expression for K shows that only relatively massive planets with short periods can be detected with the radial velocity method. Using $M_p \ll M_\star$, we can rewrite K and T in the form

$$K \text{ (m s}^{-1}\text{)} = 28.4 \left(\frac{T}{1 \text{ yr}} \right)^{-1/3} \left(\frac{M_p \sin i}{M_J} \right) \left(\frac{M_\star}{M_\odot} \right)^{-2/3} \frac{1}{\sqrt{1 - e^2}}, \quad (2)$$

$$T \text{ (yr)} = \left(\frac{a}{1 \text{ AU}} \right)^{3/2} \left(\frac{M_\star}{M_\odot} \right)^{-1/2}. \quad (3)$$

The detection limit of the instruments that are currently being used is about 1 m s^{-1} . In principle, Jupiter mass planets could be detected rather far from their parent star. However, the detection is confirmed only if the planet is monitored for at least one orbital period. Given that surveys began about 10 years ago, only planets with semi-major axes less than about 4–5 AU can be observed. Note that a Jupiter mass planet with $a = 5 \text{ AU}$ around a solar mass star induces a velocity $v \simeq 11 \text{ m s}^{-1}$.

For an Earth mass planet with $a = 1 \text{ AU}$ around a solar mass star, $v \simeq 0.1 \text{ m s}^{-1}$. It is not clear whether such a precision could be reached. Such a signal would have to be extracted from the ‘noise’ at the stellar surface, i.e. random Doppler shifts due to atmospheric oscillations, stellar spots or long period effects caused by magnetic cycles. Note that the distance to the star is not a limiting factor as long as enough photons can be collected.

Several long-term radial velocity surveys for planets are ongoing in different countries (for a list see Perryman and Hainault (2005)). They all use ground-based telescopes. New detections are regularly announced. So far all the planets found with this technique are within about 100 pc (parsecs) from the Earth.

2.1.2. Transits. The first transit of an extrasolar planet (HD 209458 b) was detected in September 1999 (Charbonneau *et al* 2000). The planet, which is on an orbit with $a = 0.045$ AU, had already been observed by the radial velocity method. The detection of the transit, in addition to proving beyond doubt that the Doppler shift was indeed caused by a planet, enabled the planet mass ($0.63 M_J$) to be determined, as $\sin i$ has to be close to 1 for a transit to be seen. It also enabled the planet radius (1.3 Jupiter radius) to be measured, as the relative variation of the stellar flux during the transit is

$$\frac{\Delta F_\star}{F_\star} = \left(\frac{R_p}{R_\star} \right)^2, \quad (4)$$

where R_p is the planet radius and R_\star is the stellar radius. For Jupiter orbiting the Sun, $\Delta F_\star/F_\star = 1\%$, whereas for the Earth it is 0.01% . From the ground, the accuracy is about 0.1% and is limited by the variable extinction of the atmosphere. From space, the accuracy reaches $10^{-4}\%$ so that terrestrial planets could be detected without any problem. However, the probability of detecting a transit, R_\star/a , is rather small. For $R_\star = 1 R_\odot$, this probability is about 10% for $a = 0.05$ AU, i.e. for the planets on the tightest orbits known to date. Therefore, a large number of stars have to be observed for a significant number of transits to be detected.

Note that the mean mass density of HD 209458 b can be calculated from its mass and radius. It is found to be about 400 kg m^{-3} , significantly smaller than the mean density of Saturn (700 kg m^{-3}), the least dense of the planets of our solar system. This confirms that this planet is made mainly of gas. The atmosphere of the planet was further detected by HST, which observed additional sodium absorption due to the light passing through the planet atmosphere as it transited across the star (Charbonneau *et al* 2002). More recently, atomic hydrogen, oxygen and carbon have also been detected in the extended envelope around the planet (Vidal-Madjar *et al* 2003, 2004), suggesting that the atmosphere is evaporating.

The first planet found by the transit method, TrES-1, was detected by the STARE (STellar Astrophysics and Research on Exoplanets) program (Alonso *et al* 2004) and confirmed 8 days later by an amateur astronomer in Belgium. Since then 5 more planets have been found by this method by OGLE (optical gravitational lensing experiment). All these planets have lately been confirmed by radial velocity measurements. Their separation from the central star ranges from 0.0225 to 0.047 AU. The planets found by OGLE are about 1500 pc away, much further away than the planets found by the radial velocity method.

Several ground-based surveys for planet transits are ongoing. A few space missions dedicated at least in part to the planet search are also being developed. The first experiment, which will be launched in 2006, is COROT (for CONvection ROTation and planetary Transits), a mini-satellite developed by the French National Space Agency (CNES). Its precision is 0.03% , so that it will detect only planets with a radius larger than twice the Earth radius. It will monitor about $60\,000$ stars during 2.5 years. The next experiment is KEPLER, which will be launched in 2007 by NASA. Its precision is $10^{-3}\%$ and it will monitor 10^5 stars during 4 years. It is expected to find at least 50 terrestrial planets.

2.1.3. Astrometry. The radial velocity method described above uses the motion of the star along the line of sight. It is also possible to measure the motion of the star *perpendicular* to the line of sight and to derive the characteristics of the planet and its orbit from it. If the planet is on a circular orbit, the angular motion of the star is

$$\alpha \text{ (arcsec)} = \frac{M_p}{M_\star} \frac{a/1 \text{ AU}}{d/1 \text{ pc}}, \quad (5)$$

where d is the distance to the star, measured in parsecs. Since d and M_\star are known and a is given by the periodicity of the star motion, M_p can be derived from the measure of α . The

motion of the Sun caused by Jupiter and seen from a distance of 10 pc is $500 \mu\text{as}$, whereas that due to the Earth, and seen from the same distance, is $0.3 \mu\text{as}$.

From the ground, the interferometers provide the best astrometric accuracy. It is about 1 mas for the VLTI and KECK in the near IR and should reach $10 \mu\text{as}$ in the near future over narrow fields of view. Surveys for planets with these interferometers will begin in the very near future.

For space observations, the accuracy will be on the order of $1 \mu\text{as}$. Therefore, since the stars that are being observed are at least at a distance of a few parsecs, it will not be possible to detect terrestrial planets in habitable zones with this method, which is better suited for the observation of giant planets. Note that astrometric measurements are easier for longer period systems. This method and the radial velocity method are thus complementary.

To date, only one space mission, Hipparcos (which was launched by ESA in 1989 and operated until 1993), was dedicated to astrometric measurements. Although it was not meant to detect planets, it did observe stars around which planets were later observed. Its measurements were therefore used to constrain the mass of some of the planets discovered by the radial velocity method.

Two space experiments are currently being developed. SIM (Space Interferometry Mission) is scheduled for launch by NASA in 2010. Its accuracy will be $1 \mu\text{as}$. It will not observe a very large number of stars but will focus in particular on the multiple systems already detected from the ground to get a precise measurement of the orbital characteristics. GAIA, which is scheduled for launch by ESA in 2011, will have an accuracy of $2\text{--}10 \mu\text{as}$ and will observe $\sim 10^5$ stars within a distance of about 200 pc from the Sun during 5 years. Simulations indicate that GAIA could detect between 10 000 and 50 000 giant planets.

2.1.4. Microlensing. This technique relies on a background star to provide a source of light. Then, foreground stars act as gravitational lenses when they pass in front of the background star and cause the source star to suddenly brighten by as much as 1000 times. This typically lasts for a few weeks or months. While this is the normal pattern of a microlensing event, things are substantially different when the lensing (foreground) star has a smaller companion. If a planet is positioned close enough to the lensing star, it will indeed perturb the light curve by adding two successive spikes of brightness, superimposed upon the regular pattern of the microlensing event (Mao and Paczynski 1991, Gould and Loeb 1992, Bennett and Rhie 1996). Microlensing is most sensitive to planets at a separation $\sim 1\text{--}5$ AU from the lens star. In addition, effects on light curves are large even for planets of less than an Earth mass. This technique is therefore well suited for finding terrestrial planets in habitable zones. An inconvenient aspect of the microlensing technique is that the events do not repeat.

Several collaborations, from around the world, are presently searching for planets with the microlensing technique. Two of them (OGLE and MOA) are survey programs which monitor large number of stars. The two other projects (PLANET and MicroFUN) pick up on the alerts that are issued by the survey collaborations when a microlensing event is detected and do a follow-up monitoring of the event.

One planet (OGLE235–MOA53) has been found in 2004 simultaneously by the MOA and OGLE collaborations (Bond *et al* 2004). Its mass and semi-major axis are about $2 M_J$ and 3 AU, respectively. It is located at a distance of 5.2 kpc from Earth, which is much too far away for the radial velocity method to be used.

So far, the microlensing programs have mainly put limits on the percentage of stars harbouring planets within a given range of separations. Analysis of 5 years of PLANET photometry indicates that less than 33% of the M dwarfs in the Galactic bulge can have

Jupiter-mass companions within a distance between 1.5 and 4 AU. The upper limit is raised to 45% if the range of separations considered is 1–7 AU (Gaudi *et al* 2002).

A space-based gravitational microlensing program (GEST, for Galactic Exoplanet Survey Telescope) is currently being studied by NASA. GEST, which would be able to detect planets with masses as low as that of Mars, would survey more than 100 million stars. It would be expected to find about 100 Earth mass planets.

2.1.5. Direct imaging. The direct detections of the transiting planets HD 209458 b (Deming *et al* 2005) and TrES-1 (Charbonneau *et al* 2005) were reported simultaneously in March 2005. These were the very first observations of photons emitted by extrasolar planets. The detections were made by the Spitzer Space Telescope by observing the decrement in infrared flux during the secondary eclipses, when the planets passed behind their parent star. These observations enable the brightness temperature of the planets at a given wavelength to be calculated and, assuming some model for the atmosphere, a blackbody temperature to be estimated.

In general, direct imaging of extrasolar planets will enable long period objects to be detected and therefore will complement radial velocity surveys. Very low-mass non-transiting objects can currently be detected only if they are isolated (e.g. free floating planets). Bound planets are much more difficult to observe because of the very high star to planet flux ratio. The contrast is smaller in the IR, which is therefore the most favourable range of wavelengths for planet imaging. The light coming from the star can in addition be blocked by using a coronagraph or nulling techniques on interferometers. To limit the halo of the star so that it is not brighter than the planet, both turbulence and defects of the instrument have to be corrected with unprecedented precision, so a high-performance adaptive optics system is needed. The latest developments in adaptive optics (the first light from NAOS, a system installed by ESO on the VLT, was received in 2001) suggest that a system for direct detection of extrasolar planets may be available on 8–10 m ground-based telescopes within 5 years.

Several ground-based projects are currently being developed or under study. Most of them rely on interferometers to image planets. There are also several space based projects, among which are Darwin (studied by ESA) and Terrestrial Planet Finder (studied by NASA). These space interferometers would be able to detect Earth-like planets in habitable zones and to perform spectroscopic observations of their atmosphere. Given the ambitious nature of both projects, NASA and ESA may collaborate on the final mission.

2.2. Statistical properties of observed extrasolar planets

So far, about 5% of the stars monitored in radial velocity surveys have been found to harbour at least one planet. This is a lower limit of course, as low-mass and/or long-period objects cannot be detected with the current techniques.

2.2.1. Mass and separation distributions. The histograms of the projected mass $M_p \sin i$ and semi-major axis a of the known extrasolar planets are shown in figure 1.

The lightest object has a mass of $0.042 M_J$ or 13.3 Earth masses, which is very close to the mass of Uranus. Although the radial velocity surveys are biased towards massive planets, 37% of the planets detected so far have a mass smaller than $1 M_J$ and only 13% have a mass larger than $5 M_J$. *Within a few AU from the central star, there is a deficit of massive planets.*

The distribution of a shows that there is a significant number of planets (12%) with a between 0.035 and 0.05 AU, i.e. an orbital period between 2.5 and 4 days (the so-called

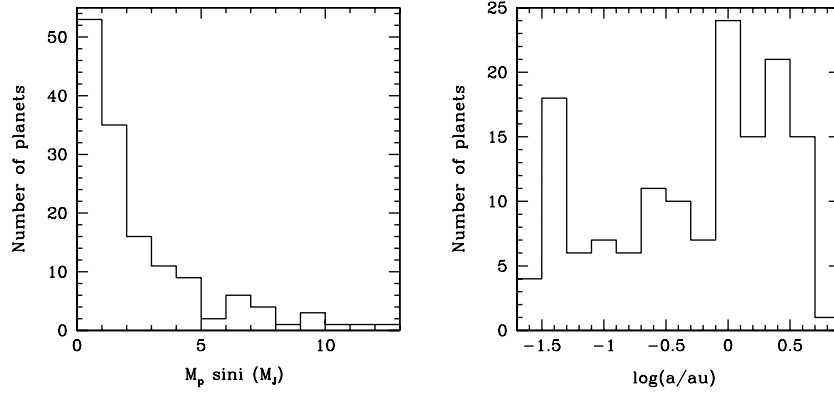


Figure 1. Distribution of projected mass $M_p \sin i$ (in M_J , *left panel*) and semi-major axis a (in AU, logarithmic scale, *right panel*) for the known extrasolar planets.

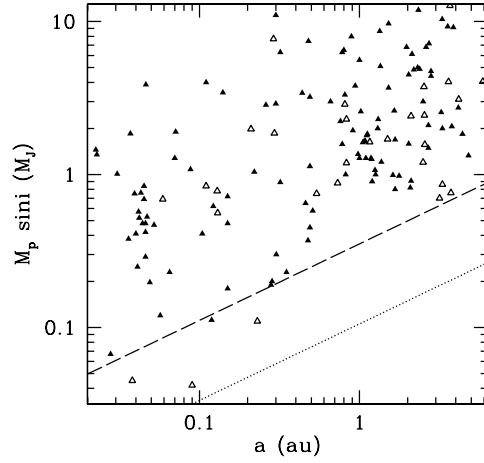


Figure 2. Projected mass $M_p \sin i$ (in M_J) versus semi-major axis a (in AU) on logarithmic scales for the known extrasolar planets. Planets in multiple systems are indicated by open symbols. The dotted and dashed lines represent the detection limit for planets on circular orbits and an accuracy of 3 m s^{-1} and 10 m s^{-1} , respectively (equation (1) with $e = 0$).

‘hot Jupiters’). Only 4 objects have a smaller period, which suggests that *there is a pile-up at a rather well-defined separation ($\sim 0.4 \text{ AU}$) or period ($\sim 3 \text{ days}$)*. Only very few objects with a mass larger than that of Uranus seem to be able to ‘leak’ through what looks like a barrier. Since observations are biased towards short-period planets, as noted by, e.g. Marcy *et al* (2005), *there is a clear rise of the distribution with separation*. Beyond $\sim 0.4 \text{ AU}$, the data appears to be subject to \sqrt{N} noise with dependence on binning, an approximate smooth representation for the number of planets $N(a)$ with semi-major axis less than a in AU is $dN/d(\log a) = 50 \log a + 95$. According to this, the number expected within an interval of 0.2 in $\log a$ is ~ 9 or 19 at $\log a = -1$ or 0 , respectively.

Figure 2 shows the projected mass $M_p \sin i$ versus semi-major axis a . Planets in multiple systems are indicated by open symbols. *There is a deficit of sub-Jupiter mass planets at separations larger than $\sim 0.5 \text{ AU}$* . More precisely, there is no planet with $M_p \sin i < 0.7 M_J$ at $a > 0.5 \text{ AU}$. Since about 10% of the planets have a mass lower than $0.7 M_J$, the absence

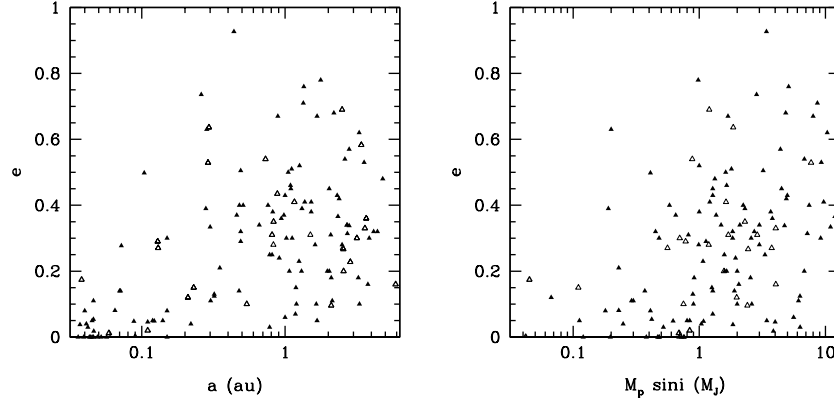


Figure 3. Eccentricity e versus semi-major axis a (in AU) on a logarithmic scale (*left panel*) and projected mass $M_p \sin i$ (in M_J) on a logarithmic scale (*right panel*) for the known extrasolar planets. Planets in multiple systems are indicated by open symbols.

of such planets at separations larger than 0.5 AU, where 57% of the planets are, is statistically significant. The dotted and dashed lines in figure 2 indicate the detection limit for planets on circular orbits and an accuracy of 3 m s^{-1} and 10 m s^{-1} , respectively. It is not clear that the deficit of low-mass planets at larger a is the result of observational biases (see also Udry *et al* (2003)). *There is also a deficit of high-mass planets at separations smaller than $\sim 0.3 \text{ AU}$.* More precisely, there is no planet with $M_p \sin i > 4 M_J$ at $a < 0.29 \text{ AU}$. As observations favour massive planets with short-periods, this deficit is not due to an observational bias. It has further been pointed out that planets more massive than $2 M_J$ at $a < 4 \text{ AU}$ are actually members of binary star systems (Zucker and Mazeh 2002).

2.2.2. Eccentricity distribution. Figure 3 shows the eccentricity e versus semi-major axis a and projected mass $M_p \sin i$. Planets with $a < 0.05 \text{ AU}$ have almost circular orbits, which is consistent with tidal circularization. At larger separations, *the eccentricity varies from 0 up to 0.93*. This is in sharp contrast with the planets of our solar system, which are on nearly circular orbits. The upper limit of 0.93 may be special, as it may be due to the secular perturbation of the planet HD 80606 b by a stellar companion on an inclined orbit (Wu and Murray 2003). Nonetheless, eccentricities in the range 0–0.8 are common. Note that higher eccentricities seem to be obtained for more distant planets. The eccentricity distribution of planets in systems of multiple planets is similar to that of planets in single systems.

Figure 3 shows that higher eccentricities are obtained for more massive planets. This is consistent with the trend noted above of higher eccentricities at larger separations together with the fact that more massive planets are at larger separations. But note that the observed distributions could be produced either by physical processes tending to produce higher eccentricities for higher masses or higher eccentricities at larger orbital separations or a combination of the two. A suggested process of the former type is eccentricity generation through disc planet interaction (Artymowicz 1992), and one of the latter type is orbital circularization close to the central star. Here again, the eccentricity distribution of planets in systems of multiple planets is similar to that of planets in single systems.

Halbwachs *et al* (2005) suggest that the differences between the eccentricity distributions of planets and stellar binaries are an indication that these two classes of objects have different formation mechanisms.

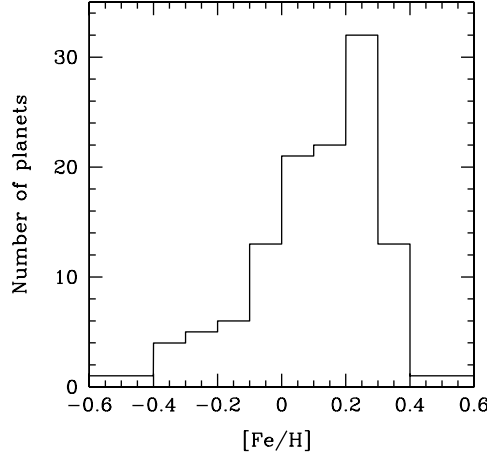


Figure 4. Distribution of metallicity $[\text{Fe}/\text{H}]$ for the stars harbouring planets.

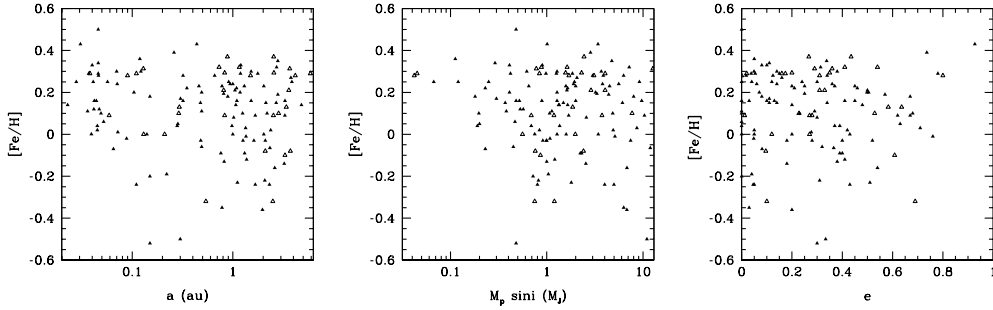


Figure 5. Stellar metallicity $[\text{Fe}/\text{H}]$ versus semi-major axis a (in AU) in logarithmic scale (*left panel*), projected mass $M_p \sin i$ (in M_J) in logarithmic scale (*middle panel*) and eccentricity e (*right panel*) for the known extrasolar planets. Planets in multiple systems are indicated by open symbols.

2.2.3. Metallicity of stars harbouring planets. Figure 4 shows the metallicity distribution of the stars around which planets have been found. The metallicity of most of the field dwarfs with no known companion lies in the range -0.3 – 0.2 (Santos *et al* 2001). The planet host stars are therefore metal rich, and the planetary frequency is clearly increasing with metallicity. This metallicity excess appears to have a ‘primordial’ origin (Pinsonneault *et al* 2001, Santos *et al* 2001, 2002, Sadakane *et al* 2002), rather than being due to the infall of planetary material onto the star which might result in observed heavy element pollution for stars in the mass range 1 – $1.3 M_\odot$ (e.g. Sandquist *et al* (1998, 2002)). The favoured explanation for the high metallicity of planet host stars is therefore that planets form more easily in a metal rich environment.

Note that no short-period planets were found in the metal poor globular cluster 47 Tucanae in which about 34 000 stars were monitored by HST. Simulations had shown that about 17 hot Jupiters should have been detected by photometric transit signals if the frequency of these objects were the same in the solar neighbourhood and in 47 Tucanae (Gilliland *et al* 2000).

Figure 5 shows the stellar metallicity versus a , $M_p \sin i$ and e . Sozzetti (2004) has found that there is a weak correlation (at the 2 to 3σ level) between the stellar metallicity and the orbital period of the system. This correlation becomes stronger when only single stars with one detected planet are considered.

2.2.4. Planets in multiple stellar systems. As noted above, the most massive short-period planets are all members of stellar binaries (Zucker and Mazeh 2002). It also appears that planets orbiting a component of a multiple star system tend to have a very low eccentricity when their period is shorter than about 40 days, i.e. when a is smaller than about 0.2 AU (Eggenberger *et al* 2004).

2.2.5. Free floating planets. Recently, a population of about 30 very young (a few million years) free-floating objects with planetary masses (from a few to $13 M_J$) has been found in the Sigma Orionis open cluster (Zapatero Osorio *et al* 2000) and in the Trapezium (Lucas and Roche 2000). The lightest of these objects is S Ori 70 (Zapatero Osorio *et al* 2002). Its mass is derived from evolutionary models and is estimated to be $3 M_J$. However, Burgasser *et al* (2004) claim that this object may actually not be a member of the σ Orionis cluster, in which case its mass may be much higher.

Béjar *et al* (2001) have pointed out that if the distribution of stellar and substellar masses in σ Orionis is representative of the Galactic disc, older and much lower luminosity free-floating objects with masses down to a few M_J should be abundant in the solar vicinity, with a density similar to M-type stars.

The discovery of these free-floating objects has brought some confusion in the definition of planets, as they are not orbiting a star and they probably have not formed through the standard core-accretion scenario (see below).

3. Theories of planet formation

Terrestrial planets are believed to be formed via solid body accretion of kilometre-sized objects, which themselves are produced as a result of the sedimentation and collisional growth of dust grains in the protoplanetary disc (see Lissauer (1993) and references therein). This theory comes from the idea already proposed in the 19th century that the Earth and the other terrestrial planets were formed from meteoritic material. At the beginning of the 20th century, this idea was supported by the geologist T C Chamberlin who called ‘planetesimals’ the solid objects that had accumulated to form the planets. But it is only in the 1960s that the theory was developed in a quantitative way by Safronov (1969), who calculated in details the different stages of terrestrial planet formation. This pioneering work, originally written in Russian, was published in English in 1972 under the title ‘Evolution of the Protoplanetary Cloud’. Since then, the theory has been further developed, in particular, by Wetherill and collaborators who have used intensive numerical simulations to study planetesimal accumulation.

As far as giant planets are concerned, two theories have been proposed. According to the first theory, giant planets are formed through the collapse and fragmentation of protostellar discs. This scenario is related to that proposed by Laplace in 1796. In its modern version, it has been developed by Kuiper in the 1950s and by Cameron in the 1960s and 1970s. It is still being studied today. The second theory, proposed by Cameron in 1973, is called the ‘core accretion model’. A solid core is first assembled in the same way as terrestrial planets. Once the core becomes massive enough to gravitationally bind the gas in which it is embedded (typically at around a tenth of an Earth mass (see section 3.2.2), a gaseous envelope begins to form around the core.

3.1. Terrestrial planet and planetary core formation

In the formation of terrestrial planets or planetary cores, several stages can be considered. First, micron sized dust grains that are present when the disc forms a sediment towards the

disc midplane. These grains are subject to both the drag force exerted by the gas and the gravitational force from the central star. Accumulation of grains during this first stage leads to centimetre per metre sized particles. It is still not clear how these objects further grow to become 0.1–1 km sized planetesimals. It is likely that growth occurs through collisions and accumulation. The planetesimals are massive enough such that mutual gravitational interactions have to be taken into account in computing their evolution. Further collisions and accumulation lead to terrestrial planets or the core of giant planets. These different processes are described in detail below.

3.1.1. Growth from dust to centimetre per metre sized particles. When the disc forms out of the collapse of a molecular cloud, it contains dust grains the size of which is that of the interstellar grains, i.e. 0.1–1 μm . Under the action of the gravitational force exerted by the central star and the drag force exerted by the gas, when there is no turbulence present, these grains tend to undergo damped oscillations around the disc midplane, towards which they eventually sediment. Simultaneously, they drift towards the disc centre. This is because the coupling with the gas forces them to rotate at a slightly sub-Keplerian velocity. Since they are not supported by pressure forces, they consequently cannot stay on a fixed orbit.

For typical disc parameters, the gas mean free path around the disc midplane is $\sim 1\text{ m}$. This is very large compared with the size of the grains we are interested in here. In this regime, the drag force exerted by the gas on the grains is $\sim -\pi r_d^2 \rho c_s (\mathbf{v} - \mathbf{v}_d)$ (Epstein law, e.g. Weidenschilling (1977)), where r_d is the radius of the grains, which are assumed to be spherical, c_s is the thermal velocity of the gas (i.e. the sound speed), ρ is the mass density of gas in the disc, \mathbf{v}_d is the velocity of the grains and \mathbf{v} that of the gas. This formula is valid provided $|\mathbf{v} - \mathbf{v}_d| \ll c_s$, which is the case here as can be checked *a posteriori*. We define the time:

$$\tau_e \equiv \frac{m_d}{\pi r_d^2 \rho c_s},$$

with m_d being the mass of a dust grain. This is the timescale it takes for the velocity of a grain subject only to the drag force to decrease by a factor e . For $r_d = 1\text{ }\mu\text{m}$, $\rho \sim 10^{-10}\text{ g cm}^{-3}$, $c_s \sim 10^5\text{ cm s}^{-1}$ (which corresponds to a temperature of about 700 K), which are typical conditions at about 1 AU from the central star, and assuming the mass density of the grains to be 3 g cm^{-3} (which is comparable to the mass density of the Earth's crust and that of the asteroids), we get $\tau_e \sim 10\text{ s}$.

For the gas, we write $\mathbf{v} = (v_r, v_\phi, v_z)$ in cylindrical coordinates and, for the dust, $\mathbf{v}_d = (v_{r,d}, v_{\phi,d}, v_{z,d})$. We assume the motion of the grains to be almost Keplerian with a small drift velocity, i.e. $v_{\phi,d} \sim v_K$ and $|v_{r,d}| \ll |v_{\phi,d}|$, where v_K is the Keplerian velocity. In this situation in a steady state, the dust to a first approximation comoves with the gas so that $v_{\phi,d} = v_\phi$. This means that accelerations due to radial pressure gradients acting on the gas have to be balanced by the accelerations due to drag acting on the dust which occur because of the radial drift velocity which will be given by

$$\frac{v_{r,d}}{\tau_e} = \frac{1}{r} (v_\phi^2 - v_K^2) = \frac{1}{\rho} \frac{\partial P}{\partial r}.$$

As expected, if the pressure increases inwards, $v_{r,d} < 0$. Also $v_{r,d} \rightarrow 0$ as $r_d \rightarrow 0$, which means that very small grains are ‘carried’ by the gas. With $|\partial P / \partial r| \sim P/r$, $c_s^2 \sim P/\rho$ and $H/r \sim c_s/v_K$, where H is the disc semi-thickness, we get $v_{r,d} \sim -c_s(H/r)\tau_e\Omega_K$, where Ω_K is the Keplerian angular velocity. With the values of the parameters used above, this gives $v_{r,d}$ on the order of a few 10^{-2} cm s^{-1} at $r = 1\text{ AU}$.

If the grains were not subject to the gas drag force, they would oscillate around the disc midplane with the frequency Ω_K . Since $2\pi/\Omega_K \gg \tau_e$, the oscillations are suppressed and the grains sediment towards the disc midplane. As for a highly damped harmonic oscillator, the characteristic timescale for sedimentation is

$$\tau_s = \frac{1}{\Omega_K^2 \tau_e} = \frac{3\rho c_s}{4\Omega_K^2 \rho_d r_d}.$$

Here ρ_d is the mass density of the grains. For the values of the parameters used above, $\tau_s \simeq 10^5$ years at 1 AU. This is small compared with the planet formation timescale, as will be seen below. We have supposed that m_d and r_d are constant, so that τ_s given above is actually a maximum. Indeed, if the grains grow through collisions as they fall toward the midplane, the drag force due to the gas decreases and the sedimentation is accelerated. Note that during τ_s , the grains drift radially over a distance $|v_{r,d}|\tau_s$, which is on the order of 10^{-2} AU at 1 AU. This is very small compared with the disc semi-thickness (~ 0.1 AU), which justifies the neglect of $v_{r,d}$ in the discussion of the vertical motion. However, grains grow through collisions during their descent. Let us suppose that they ‘absorb’ all other grains on their way to the midplane. When a grain moves a distance $|dz|$, its mass then increases by $dm_d = \pi r_d^2 |dz| \rho_s$, where ρ_s is the mass density of solids (dust) in the disc. This corresponds to an increase of its radius of $dr_d = \rho_s |dz| / (4\rho_d)$. Therefore, for the values of the parameters used above, a grain initially at the altitude H would have a size \sim millimetre when it settles towards the midplane at a few AU. Its radial drift velocity then cannot be neglected.

Numerical simulations in laminar discs (Weidenschilling 1980, Nakagawa *et al* 1981) show that grains in the process of settling towards the midplane undergo a significant radial drift which promotes collisions. This enables them to grow further up to 0.1–1 m, and sedimentation is then found to occur in a few 10^3 years (sedimentation being faster for larger grains, as $\tau_s \propto 1/r_d$). Note however that these simulations assume that the grains stick perfectly to each other, so that they agglomerate when they collide.

Even though the growth of the grains during their sedimentation is not yet fully modelled by numerical simulations, it is generally accepted that this process leads to centimetre per metre sized particles around the (laminar) disc midplane in less than 10^4 years.

Calculations including turbulence have been performed by Weidenschilling (1984). He found that the coagulation rate is initially much greater than in a non-turbulent disc. Aggregates quickly reach sizes ~ 0.1 –1 cm, but erosion and breakup in collisions prevent growth of larger bodies. These aggregates are too small to settle on the plane of the disc in the presence of turbulence.

3.1.2. Growth from centimetre per metre sized particles to 0.1/1 km sized planetesimals. Before examining the processes that could lead to further growth of the solid particles, let us discuss the timescale on which such processes may have to operate.

Maximum radial drift. When the gas mean free path l is no longer large compared with the size of the grains r_d , the Epstein law cannot be used to calculate the drag force between the gas and the grains. In general, the drag force acting on a single grain is given by $-f(v_d - v)$, where f is a coefficient that depends on the characteristics of the grains and the gas. In steady state, the horizontal equations of motion for the grains can therefore be written as follows:

$$\frac{2v_\varphi \Delta v_\varphi}{r} + \frac{1}{\rho} \frac{\partial P}{\partial r} = \frac{f}{m_d} v_{r,d}, \quad (6)$$

$$\left(\frac{v_\varphi}{r} + \frac{\partial v_\varphi}{\partial r} \right) v_{r,d} = -\frac{f}{m_d} \Delta v_\varphi, \quad (7)$$

where $\Delta v_\varphi = v_{\varphi,d} - v_\varphi$. Note that

$$v_\varphi^2 = \frac{r}{\rho} \frac{\partial P}{\partial r} + v_K^2. \quad (8)$$

In equation (7) we have used $|v_{r,d}| \ll |v_{\varphi,d}|$ and $|\Delta v_\varphi| \ll v_\varphi$. We can further eliminate f/m_d to get a quadratic equation for Δv_φ . It is straightforward to show that this equation has a solution only if

$$\left(\frac{1}{\rho} \frac{\partial P}{\partial r} \right)^2 - 8 \left[\frac{2}{\rho r} \frac{\partial P}{\partial r} + \frac{2}{r} \frac{\partial}{\partial r} \left(\frac{1}{\rho r} \frac{\partial P}{\partial r} \right) + \frac{\Omega_K^2}{2} \right] v_{r,d}^2 \geq 0. \quad (9)$$

The first two terms in bracket are negligible compared with the last one since $c_s/(r\Omega_K) \sim H/r \ll 1$. This condition can therefore be written $|v_{r,d}| \leq |v_\varphi - v_K| \sim c_s^2/(r\Omega_K)$, where again we have used $v_\varphi/v_K \simeq 1$. At 1 AU, this gives a maximum value of $|v_{r,d}|$ on the order of 10^4 cm s^{-1} , independent of r_d . From section 3.1.1, $|v_{r,d}| \sim 10^{-2} \text{ cm s}^{-1}$ when $r_d \sim 1 \mu\text{m} \ll l$, i.e. when the Epstein law for drag still holds, much smaller than the maximum speed derived here. As $|v_{r,d}| \propto r_d$, the maximum speed is reached when r_d is at least as large as $l \sim 1 \text{ m}$. The inflow time corresponding to the maximum speed is very short, being less than 100 years at 1 AU.

Radial drift when $r_d \gg l$. In the regime $r_d \gg l$, the drag force acting on a single grain is $\simeq -C_D \pi r_d^2 \rho |v_d - v|(v_d - v)$, where C_D is the drag coefficient, being a number of order unity (Stokes law, e.g. Weidenschilling (1977)). In steady state, the radial equation of motion for the solid particles leads to $v_{\varphi,d} = v_K$ if we assume $|v_{r,d}| \ll |v_{\varphi,d}|$. This means that the angular velocity of large particles around the star is not affected by the gas. The azimuthal equation of motion (7), after using equation (8), then gives

$$v_{r,d} = \frac{C_D r_d^2}{2m_d \rho \Omega_K^3} \left(\frac{\partial P}{\partial r} \right) \left| \frac{\partial P}{\partial r} \right|. \quad (10)$$

Important results follow from the dependence of the migration speed on the pressure gradient. In particular, if the pressure increases inwards, inward drift occurs. Similarly, an outwardly increasing pressure results in outward drift. Thus, particles will accumulate at pressure maxima and be depleted at pressure minima. We comment that this is a general feature of a non-magnetic gas and dust mixture that arises because the forces acting on the two components are the same at pressure extrema so that they can comove there without drag. To estimate the typical drift speed magnitude in the general case, we set $\partial P/\partial r = nP/r$. Then equation (10) gives

$$v_{r,d} = c_s \frac{3C_D n |n| \Sigma}{16\pi \rho_d r_d} \frac{c_s^2}{r^2 \Omega_K^2}, \quad (11)$$

where Σ is the surface density of the gas disc.

For $C_D = 0.2$, $|n| = 3$, $\rho_d = 1 \text{ g cm}^{-3}$, $r_d = 10 \text{ m}$ and $\Sigma = 10^3 \text{ g cm}^{-2}$, we get $|v_{r,d}| = 0.1 c_s [c_s/(r\Omega_K)]^2$. The radial flow timescale is thus $(5/\pi)(r\Omega_K/c_s)^3 P_{\text{orb}}$, with P_{orb} being the local orbital period. For $c_s/(r\Omega_K) = 0.07$, this is characteristically ~ 5000 orbits, being more than an order of magnitude longer than the timescale corresponding to the maximum drift speed. Therefore, the maximum speed is reached when $r_d \sim l \sim 1 \text{ m}$.

Although there is a dependence on the local state variables, the drift timescale for metre sized bodies at 1 AU is potentially as short as ~ 100 years. Therefore, either the process by which sub-metre sized bodies grow to become planetesimals has to be very fast or possibly special flow features, like pressure extrema or vortices, play an important role (see below).

Note however that the calculation presented above is for a particle which is isolated. In a laminar disc in which a dust layer forms around the midplane, the drift velocity cannot

be calculated by assuming that the gas velocity is given by equation (8). In the dust layer, the dust density is probably large enough that the gas is carried around at a velocity closer to the Keplerian value. The relative velocity of the solid particles with respect to the gas is therefore smaller than what we have assumed. It is actually the *collective drag* on the surface of the layer which is responsible for orbital decay. The lifetime of the dust layer against orbital decay at 1 AU is estimated to be $\sim 10^3$ years (Goldreich and Ward 1973) so that the problem still remains.

The process by which 0.1/1 km sized bodies form in the dust layer in a laminar disc is still unclear. Safronov (1969) and Goldreich and Ward (1973) independently pointed out that the dust may concentrate sufficiently in the layer to undergo gravitational instability. This occurs when the Toomre parameter $Q = \Omega c_d / (\pi G \Sigma_d)$ (with c_d and Σ_d being the velocity dispersion and surface mass density, respectively, of the dust) approaches unity. In a typical disc, this requires the mass density of dust to be $\sim 10^{-7} \text{ g cm}^{-3}$. Fragmentation of this layer then produces planetesimals the size of which is on the order of 1 km. The advantage of this scenario is that it operates very fast, on a timescale on the order of the orbital timescale.

However, it was pointed out by Weidenschilling (1980) that if the dust density reaches such high values, then it dominates the disc in the midplane, i.e. it carries the gas at a Keplerian velocity around the star. As mentioned above, because there is no pressure associated with the dust, it rotates at a different rate than the gas around the layer. In these conditions, the velocity shear between the dust layer and the gas around it gives rise to Kelvin–Helmholtz instabilities that result in turbulence, even in a disc which is laminar to begin with. This leads to the mixing of the dust so that the density in the layer is decreased and gravitational instabilities are prevented. This result was confirmed by detailed calculations by Cuzzi *et al* (1993) who showed that a rapid accretion of planetesimals by gravitationally unstable fragmentation on an orbital timescale is unlikely to occur until objects have been already accreted by some other process to the mass of the largest known meteorite samples, if at all. Note that if the disc is turbulent to begin with, settling is inhibited for particles below a certain size $\sim 1 \text{ cm}$, as already mentioned above.

The instability scenario has recently been revisited. Sekiya (1998) showed that gravitational instabilities can occur if the dust to gas ratio is much larger than the cosmic value. Youdin and Shu (2002) have discussed mechanisms for increasing this ratio. In particular, they showed that gas drag in a laminar disc causes global redistribution and concentration of small solids as they drift inwards. This is because the drift velocity of these particles decreases as they move in to smaller radii. Youdin and Shu (2002) conclude that the dust to gas ratio can become high enough for gravitational instabilities to develop in less than a few million years. However, Weidenschilling (2003) pointed out that their arguments are based on motions obtained for independent particles and claimed that collective motion due to turbulent stress on the particle layer acts to inhibit the concentration of particles and may prevent gravitational instability.

Recently, Goodman and Pindor (2000) discovered a secular instability that may lead to the rapid formation of planetesimals in the dust layer. This instability is not driven by self-gravity but by drag. If the drag exerted by the (turbulent) gas on the surface layer is collective, then the drift speed of the dust layer varies inversely with its surface density. A slight increase of the mass of some annulus would therefore result in a decrease of its drift velocity, leading to the accumulation of the dust inflowing from adjacent larger radii. The instability, which grows exponentially in a frame moving at the unperturbed drift speed, leads to the formation of overdense rings on an orbital timescale. These rings further fragment to form planetesimals of size $\sim 10 \text{ km}$ at 1 AU. However, turbulent stresses acting on the particle layer may act to inhibit concentration of particles in this case also.

If none of the mechanisms described above operate, one has to rely on sticky particle collisions (coagulation) to obtain planetesimals. Cuzzi *et al* (1993) argue that 10–100 km sized objects can be formed through this process in the layer in about a million years (also see below). Note that coagulation and growth can be made more efficient if particles can be trapped in vortices or overdense regions of the disc. Planetesimal formation in vortices has been studied by, e.g. Barge and Sommeria (1995), Tanga *et al* (1996) and Klahr and Henning (1997) who found that growth was indeed much faster. But note that Cuzzi *et al* (2001) make the point that dust density enhancement may only be significant for a narrow particle size range. Furthermore, it is not clear that vortices can live long enough in discs for any dust densification process to be efficient. More recently, trapping of particles in the spiral arms of a gravitationally unstable disc has been examined. Numerical simulations of 1–10 m sized objects in marginally stable, self-gravitating discs by Rice *et al* (2004) show that the drag force causes the solid objects to drift towards the peaks of the spiral arms where the density and pressure are highest, in accord with the behaviour predicted using equation (10) (self-gravity was not taken into account in this equation, but it would not be expected to modify the situation significantly as it acts equally on dust and gas). It is further speculated that the density enhancements may be sufficient for the growth of planetesimals through direct gravitational collapse. Similarly, Durisen *et al* (2005) suggest that the dense rings that appear in their simulations of gravitationally unstable discs are conducive to accelerated growth of planetesimals.

However, there remains an issue as to the lifetime of these features in an unstable self-gravitating disc. In future years, more may be learnt about the feasibility of the processes discussed here through more extensive three-dimensional numerical simulations.

3.1.3. Accretion of a planetary core from planetesimals. Once planetesimals of radius ~ 10 km have formed, they proceed to evolve by processes of velocity dispersion growth due to mutual gravitational interaction, gas drag, disc planet interaction and collisional accretion of mass which results in the accumulation of massive solid cores.

It is believed that the process starts with some planetesimals entering a regime of runaway growth and forming a population of embryo cores. Once these are massive enough to significantly increase the velocity dispersion equilibrium of the smaller planetesimals, runaway growth ceases and the system enters an oligarchic growth phase in which cores separated by 15 or so Hill radii accrete at comparable rates. Throughout these phases, gravitational focusing plays an important part in making the accretion rates rapid enough such that cores can enter the Earth mass range within the lifetime of the disc on a scale of 5 AU.

After this, the cores can become isolated and processes like gap formation may come into play to inhibit further accretion. The mobility of the cores as a result of disc protoplanet interaction in the presence of turbulence may be important. Note that, if the disc is laminar, rapid inward type I migration due to disc protoplanet interaction poses a serious threat to core survival. Thus the final accumulation of cores up to the $5\text{--}15 M_{\oplus}$ range is still poorly understood. Nonetheless, if favourable circumstances exist (see discussion and references in sections 4.2.1 and 4.2.2), this may be possible within the lifetime of the gas disc and thus giant planets can form.

A population of cores may be left behind to undergo gas free accumulation when the gas disc has disappeared. The formation time of the Earth through this process is expected to be $\sim 2 \times 10^8$ yr. In the outer solar system, gravitational scattering of the cores can produce escapers leaving a residue to accumulate the remaining small planetesimals and form Uranus and Neptune in a few billion years.

We now go on to discuss these processes in more detail. In doing so, we follow many authors and adopt a so-called standard, or minimum mass, solar nebula model. This has a gas

surface density profile $\Sigma \propto r^{-3/2}$ and 2 Jupiter masses within 5.2 AU (Hayashi 1981). This is an estimate of the spread-out minimum mass required to form the planets in the solar system. We further adopt a solid or condensate surface density of 1% of the gas surface density. The characteristic lifetime of the gas disc is between one and ten million years and giant planets have to be formed on this timescale in the solid core followed by gas accretion scenario.

The gas disc is also believed to undergo outward angular momentum transport allowing accretion onto the central star in the manner of an accretion disc. Associated with this is an effective kinematic viscosity, ν , the origin of which is most likely associated with turbulence driven by the magnetorotational instability (Balbus and Hawley 1991, 1998). In the standard model, we adopt $\nu/(r^2\Omega) = \alpha(H/r)^2\Omega = 10^{-5}$, with α being the well-known Shakura and Sunyaev (1973) dimensionless viscosity parameter.

Gravitational focusing and runaway accretion. We first show that *direct accumulation* of planetesimals cannot lead to the formation of Earth sized objects within the disc lifetime. Consider the simplest possible accumulation scenario where a core of mass m_α grows through the accretion of sticky solid particles of mass m_p which have number density n_p , and thus mass density $\rho_p = n_p m_p$, through direct impacts. We suppose that the accreting particles are decoupled from any gas present, so that their local root mean square velocity dispersion, v_p , is related to their estimated vertical semi-thickness, H_p , through $H_p = v_p/\Omega$ and their surface density is $\Sigma_p = 2\rho_p H_p$.

Neglecting any effects due to gravitational focusing of the colliding masses, m_α increases at a rate given by

$$\frac{dm_\alpha}{dt} = \pi a^2 \rho_p v_p, \quad (12)$$

where a is the core radius. Thus the mass accumulation timescale is given by

$$\frac{1}{t_{\text{acc}}} = \frac{dm_\alpha}{dt} \frac{1}{m_\alpha} = \frac{3\Sigma_p\Omega}{8a\rho_\alpha}, \quad (13)$$

where the density of the accumulating core is ρ_α . This gives

$$t_{\text{acc}} = \frac{8a\rho_\alpha}{6\pi\Sigma_p} \left(\frac{r}{1\text{ AU}}\right)^{3/2} \text{ yr}. \quad (14)$$

Adopting $\rho_\alpha = 1\text{ g cm}^{-3}$ and $\Sigma_p/\Sigma = 0.01$, we find

$$t_{\text{acc}} = 40 \left(\frac{\Sigma}{1\text{ g cm}^{-2}}\right)^{-1} \frac{a}{1\text{ cm}} \left(\frac{r}{1\text{ AU}}\right)^{3/2} \text{ yr}. \quad (15)$$

This indicates that, for disc models like the minimum mass solar nebula model with $\Sigma \sim 200\text{ g cm}^{-2}$ at 5 AU, straightforward accumulation processes operating with efficient sticking could conceivably form objects of up to 10–100 km in size at 5 AU, within the lifetime of protoplanetary discs of a few million years. However, to produce objects of significantly larger size, an increase of the accretion cross section resulting from gravitational focusing, which can occur at low relative impact velocities, needs to be invoked.

To take into account *gravitational focusing*, we have to include the gravitational attraction of the core on the colliding particles. This increases the impact parameter for a collision from the radius a to $a_{\text{eff}} = a[1 + 2Gm_\alpha/(av_p^2)]^{1/2}$. The effect of this gravitational focusing is to increase the effective cross section for impacting collisions accordingly. Thus equation (13) is modified to become

$$\frac{1}{t_{\text{acc}}} = \frac{dm_\alpha}{dt} \frac{1}{m_\alpha} = \frac{3\Sigma_p\Omega}{8a\rho_\alpha} \left(1 + \frac{2Gm_\alpha}{v_p^2 a}\right). \quad (16)$$

The enhancement factor given by the term in brackets on the right-hand side of equation (16) is large when v_p is very much smaller than the escape velocity from the core. For one Earth mass with a mean density of unity at 5 AU immersed in a planetesimal swarm with $v_p = 0.01r\Omega$, the enhancement factor is close to 4250.

Thus, under these conditions, forming cores in the Earth mass range in discs with properties similar to those of the minimum mass solar nebula becomes a possibility. Whether the relative impact speeds are at an appropriate level depends on the velocity dispersion equilibrium attained by the distribution of cores and impacting particles or planetesimals. This we shall consider below.

At this point, we comment that when gravitational focusing dominates, a phenomenon known as *runaway accretion* is possible (e.g. Safronov 1969, Wetherill and Stewart 1989). This occurs when the accretion rate increases rapidly enough with mass such that, if a particular core has a mass slightly exceeding that of other cores, it will grow increasingly faster and so runaway from its neighbours in mass.

For example, consider the situation when v_p is fixed. Then, given $a \propto m_\alpha^{1/3}$, from equation (16) we obtain $dm_\alpha/dt \propto m_\alpha^k$, with $k = 4/3$. When, as in this case, $k > 1$, $m_\alpha \rightarrow \infty$ in a finite time, such that most of the evolution time is spent with m_α in the neighbourhood of the smallest mass followed by rapid growth to large values. Hence the term ‘runaway’. In this situation, ultimately a single mass dominates. More generally, runaway accretion requires $v_p \propto m_\alpha^{k'}$ with $k' < 1/6$. However, note that accumulation with strong gravitational focusing may occur without runaway. Then most time is not necessarily spent with m_α close to the smallest mass.

Determination of the root mean square velocity dispersion v_p for the impacting planetesimals requires consideration of the balance between growth due to gravitational scattering and damping due to interaction with the gas which we now consider.

Analysis of gravitational scattering. When the planetesimals which form the building blocks from which the core must accrete have a single characteristic mass $m_\alpha \sim 10^{18}$ g, we require there to be $\sim 10^{10}$ such objects that can be accumulated to form a body in the Earth mass range. Such a large number is best treated by statistical means. A system of many bodies interacting under their mutual gravitation can be described by the Fokker–Planck equation which we write in the form

$$\frac{Df_\alpha}{Dt} = \Gamma_{\text{coll}}(f_\alpha) + \Gamma_{\text{gas}}(f_\alpha). \quad (17)$$

Here f_α denotes the phase space number density of bodies with mass m_α . The operator giving rise to evolution due to mutual interactions among the planetesimals which lead to gravitational scattering is Γ_{coll} . The operator causing evolution due to interaction with the gaseous disc is Γ_{gas} . The latter combines the effects of gas drag, eccentricity and inclination damping as well as orbital migration. The derivative operator is taken following a particle orbit under the gravitational force due to the central mass. Thus

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \mathbf{v} \cdot \frac{\partial}{\partial \mathbf{r}} - \nabla \Phi \cdot \frac{\partial}{\partial \mathbf{v}}, \quad (18)$$

where the central potential at a distance r from the central mass M_\star is $\Phi = -GM_\star/r$. The particle position and velocity vectors measured with respect to an arbitrary coordinate system are denoted by \mathbf{r} and \mathbf{v} , respectively.

We consider the situation when the planetesimals have a dispersion velocity that is very small compared with the orbital velocity. In this case, the position of a particular planetesimal does not deviate much from a particular radius on orbital timescales, implying that a local

description of the interactions is possible. Such a description can be given in the context of a local shearing box (Goldreich and Lynden-Bell 1965).

Local shearing sheet approximation. Distribution function for the planetesimal swarm.

In the shearing box, we consider a uniformly rotating local Cartesian coordinate system for which the origin, located at some point of interest in circular orbit with radius/semi-major axis r , corotates with the Keplerian angular velocity Ω . The x -axis points radially outwards, the y -axis points in the azimuthal direction in the direction of rotation while the z -axis points in the vertical direction. A linear expansion for the combined acceleration due to gravity and the centrifugal force is used such that

$$-\nabla(\Phi - \frac{1}{2}\Omega^2 r^2) = (3\Omega^2 x, 0, -\Omega^2 z). \quad (19)$$

Then for an axisymmetric disc with no dependence on y :

$$\frac{Df_\alpha}{Dt} \equiv \frac{\partial f_\alpha}{\partial t} + v_x \frac{\partial f_\alpha}{\partial x} + v_z \frac{\partial f_\alpha}{\partial z} - 2\Omega v_x \frac{\partial f_\alpha}{\partial v_y} + (3\Omega x + 2v_y)\Omega \frac{\partial f_\alpha}{\partial v_x} - \Omega^2 z \frac{\partial f_\alpha}{\partial v_z}. \quad (20)$$

The Liouville equation $Df_\alpha/Dt = 0$ does not have a solution corresponding to an isotropic Gaussian but does have one for which f_α takes the form of an anisotropic Gaussian given by

$$f_\alpha = C_\alpha \exp\left(-\frac{v_x^2}{2\sigma_x^2} - \frac{u_y^2}{2\sigma_y^2} - \frac{v_z^2 + \Omega^2 z^2}{2\sigma_z^2}\right). \quad (21)$$

Here the root mean square velocity dispersions $(\sigma_x, \sigma_y, \sigma_z)$ are constant and such that $\sigma_y = \sigma_x/2$. There is no constraint on σ_z . The velocity $u_y = v_y + 3\Omega x/2$ is measured relative to the local circular velocity, $v_y = -3\Omega x/2$, and we shall adopt local relative velocity vectors $\mathbf{v} = (v_x, u_y, v_z)$. The constant C_α is related to the spatial number density in the midplane, n_α , through

$$C_\alpha = \frac{n_\alpha}{(2\pi)^{3/2}\sigma_x\sigma_y\sigma_z}. \quad (22)$$

Evolution of the cores and planetesimal swarm. In general, a growing core is expected to accrete from a planetesimal swarm with a distribution of masses that itself evolves under gravitational scattering and accretion. However, following Wetherill and Stewart (1989) and Ida and Makino (1993), we simplify matters by assuming that there are only two populations present. One corresponds to the growing cores and the other to the planetesimal swarm from which they accrete. The latter distribution consists of a large number of objects with a small fixed mass, m_α . The former consists of a much smaller number of cores of larger mass, m_β , which increases as the system evolves. We adopt $m_\alpha = 10^{18}$ g. This is the mass that can be accumulated within the disc lifetime without the need for gravitational focusing as indicated above. This is also the mass that emerges from the assumption that planetesimals are able to form through the gravitational instability in a dust layer (Goldreich and Ward 1973).

The evolution of each component may be described by a Gaussian such as that given by equation (21) with velocity dispersions and number density that slowly evolve under the action of collisions.

System evolution through gravitational scattering. Encounters between planetesimals that occur without direct physical impacts tend to convert kinetic energy from shear into that of random motions, in the same way as the action of viscosity converts energy from shear into thermal motions in a gaseous disc. For the planetesimal swarm, effects due to evolution resulting from gravitational scattering occur on a timescale that is much longer than orbital.

To model this, we use the form of Γ_{coll} given by Binney and Tremaine (1987). This neglects rotation about the central mass, which should be a reasonable approximation as long

as the timescale associated with a binary encounter is short compared with Ω^{-1} . This in turn requires that $\sigma_x/\Omega > r_H$, where $r_H = [m_\gamma/(3M_\star)]^{1/3}r$ is the Hill radius appropriate to the largest characteristic mass, m_γ , involved in the encounter.

Following Binney and Tremaine (1987), we write using the summation convention for repeated indices:

$$\Gamma_{\text{coll}}(f_\alpha) = -\frac{\partial}{\partial v_i}(A_i f_\alpha) + \frac{1}{2}\frac{\partial}{\partial v_i}\left(D_{ij}\frac{\partial f_\alpha}{\partial v_j}\right). \quad (23)$$

We consider two kinds of interaction. The first, which is important during the early stages of the core mass build-up, arises when objects in the distribution with small mass, m_α , interact with each other. Then

$$A_i = 4\pi G^2 \ln(\Lambda_\alpha) m_\alpha^2 \frac{\partial h}{\partial v_i} \quad (24)$$

and

$$D_{ij} = 4\pi G^2 \ln(\Lambda_\alpha) m_\alpha^2 \frac{\partial^2 g}{\partial v_i \partial v_j}, \quad (25)$$

with

$$g = \int f_\alpha(\mathbf{v}') |\mathbf{v} - \mathbf{v}'| d^3 \mathbf{v}' \quad \text{and} \quad h = \int \frac{f_\alpha(\mathbf{v}')}{|\mathbf{v} - \mathbf{v}'|} d^3 \mathbf{v}'. \quad (26)$$

We take $\Lambda_\alpha = 3\sigma_x^2 H_\alpha / (4Gm_\alpha)$, giving the ratio of maximum to minimum impact parameters as disc semi-thickness to impact parameter for a typical deflection (e.g. Binney and Tremaine (1987)).

The second type of interaction that we consider, which is important during the later stages of core mass build-up, is between an object from the first distribution with small mass, m_α , and an object from the second distribution with large mass, m_β , the number density of these being n_β . Because of their high inertia, these can be assumed to be in circular orbit and have zero velocity dispersion. In this limit:

$$A_i = 0 \quad (27)$$

and

$$D_{ij} = 4\pi G^2 n_\beta \ln(\Lambda_\beta) m_\beta^2 (\delta_{ij} v^2 - v_i v_j) / v^3, \quad (28)$$

where $v = |\mathbf{v}|$. Here $\Lambda_\beta = 3\sigma_x^2 H_\alpha / (4Gm_\beta)$.

Growth of the velocity dispersion. The effect of gravitational scattering is to cause the velocity dispersion to increase. This is most easily seen by formulating the Boltzmann \mathcal{H} theorem for the problems we are considering. This states that for a single mass species interacting either with itself or another in fixed circular orbits with no velocity dispersion,

$$\mathcal{H} \equiv - \int f_\alpha \ln(f_\alpha) d^3 \mathbf{v} \quad (29)$$

increases monotonically with time. \mathcal{H} , which can be related to the entropy, can remain constant only for an isotropic Gaussian which cannot be attained here because of the form of particle orbits in the central potential (see equation (21)). Using the distribution function given by equation (21), and assuming that n_α and the ratio of velocity dispersion components are independent of time (see Papaloizou and Larwood (2000)), we obtain at the midplane ($z = 0$)

$$\frac{d\mathcal{H}}{dt} = \frac{3n_\alpha}{\sigma_x} \frac{d\sigma_x}{dt}. \quad (30)$$

Thus $d\mathcal{H}/dt > 0$ implies that the velocity dispersion must increase with time.

We first evaluate $d\mathcal{H}/dt$ for a single mass species interacting with itself. Using the expression (29) of \mathcal{H} and $\Gamma_{\text{coll}}(f_\alpha)$ given by equations (23)–(25), we obtain

$$\begin{aligned} \frac{d\mathcal{H}}{dt} = & -4\pi G^2 m_\alpha^2 \ln(\Lambda_\alpha) \int \frac{1}{|\mathbf{v} - \mathbf{v}'|} \frac{\partial f_\alpha(\mathbf{v})}{\partial v_i} \frac{\partial f_\alpha(\mathbf{v}')}{\partial v'_i} d^3\mathbf{v} d^3\mathbf{v}' \\ & + 2\pi G^2 m_\alpha^2 \ln(\Lambda_\alpha) \int \frac{\partial f_\alpha(\mathbf{v})}{\partial v_i} \frac{\partial f_\alpha(\mathbf{v})}{\partial v_j} \frac{f_\alpha(\mathbf{v}')}{f_\alpha(\mathbf{v})} \frac{\partial^2 |\mathbf{v} - \mathbf{v}'|}{\partial v_i \partial v_j} d^3\mathbf{v} d^3\mathbf{v}'. \end{aligned} \quad (31)$$

When $\sigma_y = \sigma_z$, the integral may be evaluated analytically for an anisotropic Gaussian specified by equation (21). The relaxation time defined through $t_R = \sigma_x / (d\sigma_x/dt)$ can then be calculated from equation (30) in the disc midplane:

$$\frac{1}{t_R} = \frac{8\sqrt{\pi} G^2 m_\alpha^2 \ln(\Lambda_\alpha) n_\alpha}{\sigma_x^3} \left[\frac{\sqrt{3}}{4} \ln \left(\frac{2 + \sqrt{3}}{2 - \sqrt{3}} \right) - 1 \right]. \quad (32)$$

Here we have used the relation $\sigma_y = \sigma_x/2$ which is valid in the Keplerian case.

In terms of the semi-thickness of the planetesimal disc $H_\alpha = \sigma_z/\Omega$ and the planetesimal surface density $\Sigma_\alpha = \sqrt{2\pi} H_\alpha m_\alpha n_\alpha$, we have

$$\Lambda_\alpha = \left(\frac{3M_\star}{m_\alpha} \right) \left(\frac{H_\alpha}{r} \right)^3 \quad (33)$$

and

$$\frac{1}{t_R} = 0.03 \ln(\Lambda_\alpha) \frac{M_D m_\alpha}{M_\star^2} \left(\frac{r}{H_\alpha} \right)^4 \Omega, \quad (34)$$

where $M_D = \pi \Sigma_\alpha r^2$ gives an estimate of the mass within the protoplanetary disc comprising the planetesimal swarm that is contained within r .

A number of effects may act to oppose the growth of the velocity dispersion induced by gravitational scattering and allow a quasi-steady state distribution to be reached. These include (see below for additional details) gas drag, tidal interactions with the disc and physical collisions between planetesimals which, if very inelastic, will prevent the dispersion velocity from significantly exceeding the planetesimal escape velocity.

To estimate likely relaxation times, we adopt $M_\star = 1 M_\odot$, $m_\alpha = 10^{18}$ g together with $H_\alpha/r = 3 \times 10^{-4}$, which corresponds to a dispersion velocity comparable to the escape velocity, and $M_D = 10 M_\oplus$. This gives $t_R \sim 5 \times 10^5$ yr at $r = 1$ AU.

But note that the relaxation time is very sensitive to the velocity dispersion, making the processes that determine it significant for determining in turn the timescale for core accumulation.

We now evaluate $d\mathcal{H}/dt$ in the case when the gravitational interactions of the planetesimal swarm, labelled by α , with the system of cores with large masses in circular orbits with zero velocity dispersion, labelled by β , become more important than interactions between its own members. Using equations (23), (27) and (28), conjointly with the expression (29) of \mathcal{H} , we find

$$\frac{d\mathcal{H}}{dt} = 2\pi G^2 m_\beta^2 n_\beta \ln(\Lambda_\beta) \int \frac{\partial \ln f_\alpha(\mathbf{v})}{\partial v_i} \frac{\partial f_\alpha(\mathbf{v})}{\partial v_j} \frac{\delta_{ij} v^2 - v_i v_j}{v^3} d^3\mathbf{v}. \quad (35)$$

In this case, with identical assumptions about the ratios of the velocity dispersion components as in the self-interaction case, the integral is readily performed to give the relaxation time in the midplane through

$$\frac{1}{t_{R,\text{mid}}} = \frac{16\sqrt{\pi} G^2 m_\beta^2 \ln(\Lambda_\beta) n_\beta}{\sqrt{2}\sigma_x^3} \left[\frac{\sqrt{3}}{2} \ln(2 + \sqrt{3}) - 1 \right]. \quad (36)$$

Here $\Lambda_\beta = (3M_\star H_\alpha^3)/(m_\beta r^3)$. When, $m_\beta \gg m_\alpha$, as here, the disc semi-thickness associated with the heavy distribution, $H_\beta = \Sigma_\beta/(\sqrt{2\pi}m_\beta n_\beta)$, is much smaller than that associated with the light distribution. Accordingly, we take an appropriate average of the relaxation rate by multiplying it by the ratio $3H_\beta/\sqrt{2}H_\alpha$, this being a measure of the proportion of time spent by a member of the light distribution interacting with the heavy distribution. Thus we obtain the final mean relaxation rate:

$$\frac{1}{t_R} = \frac{24G^2 m_\beta \ln(\Lambda_\beta) \Sigma_\beta}{\sqrt{2} H_\alpha \sigma_x^3} \left[\frac{\sqrt{3}}{2} \ln(2 + \sqrt{3}) - 1 \right]. \quad (37)$$

Comparison of scattering by small and large masses. As an ensemble of planetesimals evolves, the direction of evolution is expected to be such that, because of runaway accretion, the distribution becomes dominated by an increasing number of larger masses. Prior to this, when the smaller masses dominate the dynamical relaxation, because their mass does not change, and their velocity dispersion equilibrium is determined by a balance between dynamical relaxation and damping through the interaction with the gas disc, the relaxation rate is independent of the magnitude of the larger mass m_β . Accordingly, runaway accretion by these may occur (see above).

The condition for the larger masses to dominate the relaxation is determined by equating the relaxation rates given by equations (32) and (37) and is, neglecting any variation in $\ln(\Lambda)$,

$$3m_\beta \Sigma_\beta > m_\alpha \Sigma_\alpha. \quad (38)$$

Simulations (see, e.g. Ida and Makino (1993), Thommes *et al* (2002)) indicate that, once the larger masses start to dominate the relaxation, the system enters the orderly mode of oligarchic accretion in which the masses m_β remain in near circular orbits separated by $\sim 10f$ times the Hill radius $r_H = [m_\beta/(3M_\star)]^{1/3}r$, with f being a number of order unity, and grow by capturing the low-mass material in their neighbourhood. Then $\Sigma_\beta = m_\beta/(20f\pi r r_H)$. Using this, the condition (38) implies that for oligarchic accretion we require

$$m_\beta > 2.2 \times 10^{-7} f^{3/5} \left(\frac{r}{1 \text{ AU}} \right)^{6/5} \left(\frac{\Sigma_\alpha m_\alpha}{10^{19} \text{ g}^2 \text{ cm}^{-2}} \right)^{3/5} M_\oplus. \quad (39)$$

This strongly suggests that the transition to oligarchic accretion occurs at quite an early stage in the core accumulation process.

Damping processes. Gas drag. We consider a planetesimal of mass m_α , density ρ_α and radius a moving relative to the disc gas with speed v . The equation of motion for v is

$$m_\alpha \frac{dv}{dt} = -\pi a^2 C_D \rho v^2, \quad (40)$$

where C_D , being a dimensionless number normally in the range 0.1–1, is the drag coefficient. This leads to a relative velocity damping time, $t_{\text{gd}} \equiv v/(dv/dt)$, equivalent to

$$\frac{1}{\Omega t_{\text{gd}}} = \frac{(36\pi)^{1/3} \Sigma C_D e r}{8 \rho_\alpha^{2/3} m_\alpha^{1/3} H}, \quad (41)$$

where we have set $v = er\Omega \sim \sigma_x$ with e corresponding to an orbital eccentricity for a planetesimal. In cgs units, this gives

$$t_{\text{gd}} = 5.27 \times 10^3 \left(\frac{m_\alpha}{10^{18} \text{ g}} \right)^{1/3} \frac{H}{er} \frac{50 \rho_\alpha^{2/3}}{\Sigma C_D} \left(\frac{r}{3/2 \text{ AU}} \right) \text{ yr}. \quad (42)$$

Note that this time can be small for small masses, which is related to the need to evolve fast through the size range where decoupling from the gas first occurs to objects of $\sim 10^{18}$ g (see section 3.1.2).

Relative importance of gas drag and disc tides: because of the dependence on protoplanet mass, gas drag is more effective for smaller masses while disc tides take over for larger masses. The crossover mass, which is also the mass for which the damping time is a maximum, can be estimated by equating the gas drag time given by equation (42) with the eccentricity decay time resulting from disc tides given by equation (78) (see below). This gives

$$\frac{m_\alpha}{M_\oplus} = 5.2 \times 10^{-4} \frac{f_s^{15/8} C_D^{3/4}}{\rho_\alpha^{1/2}} \left(\frac{H/r}{0.05} \right)^3 \left(\frac{H/r}{e} \right)^{-3/4} \left(\frac{r}{5.2 \text{ AU}} \right)^{-3/2}. \quad (43)$$

Although there is significant sensitivity to protoplanet parameters, disc location, eccentricity and aspect ratio (going from 5.2 to 1 AU, other things being equal, increases it by an order of magnitude), the crossover mass is characteristically in the range 10^{-4} – 10^{-2} earth masses. Thus gas drag will be the dominant eccentricity damping process when the planetesimals first form. However, if many larger mass cores form and then interact during the later stages of core accumulation, disc tides will dominate.

Velocity dispersion equilibrium. From the above estimates we anticipate the following evolutionary sequence: in the early phases, when the core masses are comparable to the planetesimal mass, the balance is between relaxation by self-interaction among the small masses against gas drag. In this case the velocity dispersion depends only on the small unchanging mass m_α . A few masses can then undergo runaway accretion growing to large values m_β . When this is high enough, the larger masses dominate the gravitational scattering of the planetesimals which is again balanced against gas drag. The velocity dispersion in this phase increases with m_β and the core accumulation moves into the more orderly oligarchic regime (Ida and Makino 1993).

Scattering due to larger bodies balanced against gas drag. In this case we have

$$1/t_R = 1/t_{gd}, \quad (44)$$

where we use equation (37) for t_R . This gives

$$\frac{16rm_\beta \ln(\Lambda_\beta) \Sigma_\beta}{M_\star^2} \left[\frac{\sqrt{3}}{2} \ln(2 + \sqrt{3}) - 1 \right] = \left(\frac{\sigma_x}{r\Omega} \right)^5 \frac{(36\pi)^{1/3} \Sigma C_D}{16\rho_\alpha^{2/3} m_\alpha^{1/3} H}. \quad (45)$$

Numerically for $\ln(\Lambda_\beta) = 1$

$$\left(\frac{\sigma_x}{r\Omega} \right)^5 = 4 \times 10^{-6} \times \left(\frac{\Sigma_\beta}{C_D \Sigma} \right) \left(\frac{10^{18} \rho_\alpha}{m_\alpha} \right)^{2/3} \left(\frac{r}{5.2 \text{ AU}} \right)^2 \left(\frac{20H}{r} \right) \left(\frac{m_\alpha m_\beta M_\odot^2}{10^{18} M_\oplus M_\star^2} \right). \quad (46)$$

Thus, even when m_β has grown to the order of an Earth mass, the dispersion velocity σ_x plausibly remains such that $\sigma_x \lesssim H/r$, for r in the AU range, justifying the application of gas drag and maintaining a strong degree of gravitational focusing. For example, when $m_\alpha = 10^{18} \text{ g}$, $m_\beta = 1 M_\oplus$ and $\Sigma_\beta = 2.5 \times 10^{-3} C_D \Sigma$, we find $H_\beta/r \sim 0.025$, which gives an enhancement factor due to gravitational focusing (see above) of ~ 400 at 5.2 AU. This order of enhancement is needed in order to accumulate cores in the Earth mass range.

In addition, the velocity dispersion σ_x increases as $m_\beta^{1/5}$ and k , defined in the paragraph on runaway accretion, is equal to 14/15. Since this is less than unity, the accumulation does not occur in a runaway manner in which a single core outgrows everything else. Instead, a number of cores of comparable mass, or oligarchs, that are adequately separated, grow together. But, nonetheless, gravitational focusing can and must remain important if core accumulation to the Earth mass range is to occur within the gas disc lifetime.

Scattering due to isolated larger bodies balanced against disc tides when gas is present.

When all or most of the small mass planetesimals are accumulated, the likely situation that

develops is one for which large mass cores, or oligarchs, of comparable mass m_β are in near circular orbits separated by $10f$ Hill radii in isolation (Chambers *et al* 1996). The isolation mass may be estimated from the mean surface density of the accumulated solid material as

$$m_\beta = \left[\frac{20f\pi r^2 \Sigma_\beta}{(3M_\odot)^{1/3}} \right]^{3/2} = 0.46f^{3/2} \left(\frac{\Sigma_\beta}{1 \text{ g cm}^{-2}} \right)^{3/2} \left(\frac{r}{5.2 \text{ AU}} \right)^3 M_\oplus, \quad (47)$$

giving objects characteristically in the Earth mass range for a minimum mass solar nebula at 5.2 AU. But note that, for such a nebula, the isolation mass $\propto r^{3/4}$, being $\sim 0.11f^{3/2} M_\oplus$ at 1 AU.

Such masses may slowly interact building up velocity dispersion and resulting in further accumulation (Papaloizou and Larwood 2000). For such large planetesimal masses m_β , tidal interaction with the disc, should there be gas present, becomes more important than gas drag. Then we should use the timescale t_e as given by equation (78) (see below) rather than t_{gd} in equation (44) which determines the velocity dispersion equilibrium. The timescale t_e is obtained from the protoplanet–disc interaction theory described below: as this timescale is significantly shorter than the associated possible type I migration timescale t_1 , we expect that a quasi-equilibrium is set up before significant migration occurs and that this then drives the evolution of the system, which will proceed under conditions of quasi-equilibrium. We recall the relaxation timescale for a central point solar mass potential in the form:

$$t_R = \frac{5}{\ln(\Lambda)} \frac{M_\odot^2}{\pi \Sigma_\beta r^2 m_\beta} \left(\frac{H_\beta}{r} \right)^4 \left(\frac{r}{1 \text{ AU}} \right)^{3/2} \text{ yr}, \quad (48)$$

where $\Lambda = (3M_\star/m_\beta)(H_\beta/r)^3$. Equating t_e from equation (78) and t_R , assuming $e < H/r$ and $f_s = 1$, gives the equilibrium semi-thickness of the protoplanet distribution H_β in terms of the gas disc semi-thickness H :

$$\frac{H_\beta}{H} = 0.6[\ln(\Lambda)]^{1/4} \left(\frac{\Sigma_\beta}{4\Sigma} \right)^{1/4} \left(\frac{r}{1 \text{ AU}} \right)^{-1/8}. \quad (49)$$

Since H_β/H depends only very weakly on Λ , it is essentially independent of the protoplanet mass m_β , and it depends only weakly on other parameters. For the representative values $m_\beta = 0.1 M_\oplus$, $\Sigma_\beta/(4\Sigma) = 10^{-3}$ and $H_\beta/r = 0.01$, we obtain $H_\beta = 0.15H$ at 1 AU. Hence the protoplanet swarm is expected to remain thin and confined within the gaseous nebula as is confirmed by numerical simulations (Papaloizou and Larwood 2000). Note too that this implies that gravitational focusing will be important, enabling core mass build-up to the Earth mass regime. However, disc tide induced type I orbital migration then becomes important on a 10^6 yr timescale (Ward 1997, Tanaka *et al* 2002), threatening the survival of the cores against falling into the central star (see below).

The gas free case. When gas is absent so is the problem of inward core migration due to type I migration. However, should all the mass of solids be within the massive cores, there is nothing to balance velocity dispersion growth other than the effects of inter-core collisions. These are only likely to be limited when the relative velocity is comparable to the escape velocity (Safronov 1969). If the final phase of the accumulation of the terrestrial planets occurs in this way (Chambers and Wetherill 1998), we may estimate the evolution timescale for attaining $1 M_\oplus$ at 1 AU using equation (15), as gravitational focusing is of marginal importance, to be $\sim 2 \times 10^8$ yr. This timescale being $\propto r^3$ implies an unacceptably long time to form Uranus and Neptune for minimum mass solar nebula models. In addition, the escape velocity from a $1 M_\oplus$ core exceeds the escape velocity from the sun if $r > 12$ AU, implying that escapers will be generated in this region. Goldreich *et al* (2004) suggest that a residue of small particles derived from collisions amongst the original planetesimals, such as has been produced in

calculations directed towards explaining the distribution of Kuiper belt objects by Kenyon and Bromley (2004), may remain and damp the velocity dispersions, reducing both evolution times and the orbital eccentricities of the remaining growing cores, once their number has been adequately reduced.

3.2. Formation of giant planets: the core accretion model

In order to produce a Jovian mass giant planet, the solid cores discussed above must accrete gas. According to the *critical core mass* model first developed by Perri and Cameron (1974) and Mizuno (1980) (see also Stevenson (1982), Wuchterl (1995), Papaloizou and Terquem (1999)), the solid core grows in mass along with the atmosphere in quasi-static and thermal equilibrium until it reaches the critical mass, M_{crit} , above which no equilibrium solution can be found for the atmosphere. As long as the core mass is smaller than M_{crit} , the energy radiated from the envelope into the surrounding nebula is compensated for by the gravitational energy which the planetesimals entering the atmosphere release when they collide with the surface of the core. During this phase of the evolution, both the core and the atmosphere grow in mass relatively slowly. By the time the core mass reaches M_{crit} , the atmosphere has grown massive enough so that its energy losses can no longer be compensated for by the accretion of planetesimals alone. At that point, the envelope has to contract gravitationally to supply more energy. This is a runaway process, leading to the very rapid accretion of gas onto the protoplanet and to the formation of giant planets such as Jupiter. In earlier studies, it was assumed that this rapid evolution was a dynamical collapse, hence the designation ‘core instability’ for this model. The critical mass can be estimated to be in the range $5\text{--}15 M_{\oplus}$ depending on physical conditions and assumptions about grain opacity etc (see, e.g. Pollack *et al* (1996), Ikoma *et al* (2000), Papaloizou and Nelson (2005)). There is thus a need to evolve beyond a typical isolation mass appropriate to a minimum solar nebula model, and some authors have suggested that a nebula up to ten times more massive than the minimum mass solar nebula is needed (e.g. Pollack *et al* (1996)).

However, this late evolutionary phase is not yet well studied because it depends on uncertain global phenomena such as inhibition of planetesimal accretion due to isolation, competition with other cores, gap formation (Thommes *et al* 2002) and how mobile the cores are under a modified type I migration which may be stochastic in the presence of turbulence (Nelson and Papaloizou 2004) or be reversed by large scale toroidal magnetic fields (Terquem 2003). Indeed, it is believed that in order for the giant planet formation to occur and to be able to fit the observed distribution of extrasolar planets, type I migration rates obtained for a laminar disc have to be reduced by such processes (e.g. Alibert *et al* (2004)). However, this has yet to be explored fully.

Time-dependent numerical calculations of protoplanetary evolution by Bodenheimer and Pollack (1986) and Pollack *et al* (1996) support the critical core mass model, although they show that the core mass beyond which runaway gas accretion occurs, which is referred to as the ‘crossover mass’, is slightly larger than M_{crit} and that the very rapid gravitational contraction of the envelope is not a dynamical collapse. The designation ‘crossover mass’ comes from the fact that rapid contraction of the atmosphere occurs when the mass of the atmosphere is comparable to that of the core. Once the crossover mass is reached, the core no longer grows significantly through accretion of planetesimals.

The evolutionary calculations by Pollack *et al* (1996) show that the evolution of a protoplanet with a gaseous envelope is governed by three distinct phases. During phase 1, runaway planetesimal accretion, as described above, occurs, bringing about the depletion of the feeding zone of the protoplanet leading to isolation. At this point, when phase 2 begins,

the atmosphere is massive enough such that the location of its outer boundary is determined by both the mass of gas and the mass of the core. As more gas is accreted, this outer radius moves out, so that more planetesimals can be captured enabling yet more gas to enter the atmosphere. The protoplanet grows in this way until the core reaches the crossover mass, which should be identified with the Mizuno (1980) critical core mass, at which point runaway gas accretion occurs and phase 3 begins. The timescale for planet formation is determined almost entirely by phase 2, i.e. by the rate at which the core can accrete planetesimals, and is found to be a few million years at 5 AU, being comparable to the gas disc lifetime. Note that although phase 3 is relatively rapid compared with phase 2 for cores as large as $15 M_{\oplus}$, it may become much longer for smaller mass cores which become critical for reduced planetesimal accretion rates (see Pollack *et al* (1996), Papaloizou and Terquem (1999) and below).

Note that isolation of the protoplanetary core, as it occurs at the end of phase 1, may be prevented under some circumstances by tidal interaction with the surrounding gaseous disc (Ward and Hahn 1995) or by interactions with other cores (Chambers *et al* 1996).

Conditions appropriate to Jupiter's present orbital radius are normally considered, and then for standard 'interstellar' dust opacities the critical or crossover mass is found to be around $15 M_{\oplus}$. This is rather large as recent observations indicate that Jupiter has a solid core not exceeding $\sim 5 M_{\oplus}$.

The similarity between the critical and crossover masses is due to the fact that, although there is some liberation of gravitational energy as the atmosphere grows in mass together with the core, the effect is small as long as the atmospheric mass is small compared with that of the core. Consequently, the hydrostatic and thermal equilibrium approximation for the atmosphere is a good one for core masses smaller than the critical value.

3.2.1. Equations governing an envelope at equilibrium. We denote R the spherical polar radius in a frame with the origin at the centre of the protoplanet. We assume that we can model the protoplanet as a spherically symmetric nonrotating object. We consider the envelope at the hydrostatic and thermal equilibrium. The equation of hydrostatic equilibrium is

$$\frac{dP}{dR} = -g\rho, \quad (50)$$

where P is the pressure, ρ is the mass density per unit volume and $g = GM(R)/R^2$ is the acceleration due to gravity, $M(R)$ being the mass contained in the sphere of radius R (this includes the core mass if R is larger than the core radius). We also have the definition of density:

$$\frac{dM}{dR} = 4\pi R^2 \rho. \quad (51)$$

At the base of a protoplanetary envelope, the densities are so high that the gas cannot be considered to be ideal. Thus we adopt the equation of state for a hydrogen and helium mixture given by Chabrier *et al* (1992). We take the mass fractions of hydrogen and helium to be 0.7 and 0.28, respectively. We also use the standard equation of radiative transport in the form

$$\frac{dT}{dR} = \frac{-3\kappa\rho}{16\sigma T^3} \frac{L}{4\pi R^2}. \quad (52)$$

Here T is the temperature, κ is the opacity, which in general depends on both ρ and T , σ is the Stefan–Boltzmann constant and L is the radiative luminosity. Denoting the radiative and adiabatic temperature gradients by ∇_{rad} and ∇_{ad} , respectively, we have

$$\nabla_{\text{rad}} = \left(\frac{\partial \ln T}{\partial \ln P} \right)_{\text{rad}} = \frac{3\kappa L_{\text{core}} P}{64\pi\sigma G M T^4} \quad (53)$$

and

$$\nabla_{\text{ad}} = \left(\frac{\partial \ln T}{\partial \ln P} \right)_s \quad (54)$$

with the subscript s denoting evaluation at constant entropy. In writing equation (53), we have assumed that the only energy source comes from the accretion of planetesimals onto the core. The gravitational energy of the planetesimals going through the planet atmosphere is released near the surface of the core which, as a result, outputs a total luminosity L_{core} given by (see, e.g. Mizuno (1980), Bodenheimer and Pollack (1986))

$$L_{\text{core}} = \frac{GM_{\text{core}}\dot{M}_{\text{core}}}{r_{\text{core}}}. \quad (55)$$

Here M_{core} and r_{core} are, respectively, the mass and the radius of the core and \dot{M}_{core} is the planetesimal accretion rate onto the core (note that some of the planetesimals entering the atmosphere may not end up colliding with the core).

Note however that, as the interiors of these protoplanet models are convective and approximately isentropic, the above formalism can be used when the energy source corresponds to the settling of the accreting gaseous atmosphere. One can use the conservation of mass and energy to determine the evolution of the protoplanet (see Papaloizou and Nelson (2005) and sections 3.2.3 and 3.2.4).

When $\nabla_{\text{rad}} < \nabla_{\text{ad}}$, the atmosphere is stable to convection and thus all the energy is transported by radiation, i.e. $L = L_{\text{core}}$. When $\nabla_{\text{rad}} > \nabla_{\text{ad}}$, there is instability to convection. Then, part of the energy is transported by convection and $L_{\text{core}} = L + L_{\text{conv}}$, where L_{conv} is the luminosity associated with convection. We use mixing length theory to evaluate L_{conv} (Cox and Giuli 1968). This gives

$$L_{\text{conv}} = \pi R^2 C_p \Lambda_{\text{ml}}^2 \left[\left(\frac{\partial T}{\partial R} \right)_s - \left(\frac{\partial T}{\partial R} \right) \right]^{3/2} \times \sqrt{\frac{1}{2} \rho g \left| \left(\frac{\partial \rho}{\partial T} \right)_p \right|}, \quad (56)$$

where $\Lambda_{\text{ml}} = |\alpha_{\text{ml}} P / (dP/dR)|$ is the mixing length, α_{ml} being a constant of order unity, $(\partial T / \partial R)_s = \nabla_{\text{ad}} T (d \ln P / dR)$ and the subscript P means that the derivative has to be evaluated for a constant pressure. All the required thermodynamic parameters are given by Chabrier *et al* (1992). In the numerical calculations presented below we take $\alpha_{\text{ml}} = 1$.

Inner boundary. We suppose that the planet core has a uniform mass density ρ_{core} . The composition of the planetesimals and the high temperatures and pressures at the surface of the core suggest $\rho_{\text{core}} = 3.2 \text{ g cm}^{-3}$ (e.g. Bodenheimer and Pollack (1986) and Pollack *et al* (1996)), which is the value we will adopt throughout. The core radius, which is the inner boundary of the atmosphere, is then given by

$$r_{\text{core}} = \left(\frac{3M_{\text{core}}}{4\pi\rho_{\text{core}}} \right)^{1/3}. \quad (57)$$

At $R = r_{\text{core}}$, the total mass is equal to M_{core} .

Outer boundary. In the models that will be discussed below, the outer boundary of the atmosphere will be taken to be either free or at the Roche lobe radius r_{L} of the protoplanet given by

$$r_{\text{L}} = \frac{2}{3} \left(\frac{M_{\text{pl}}}{3M_{\star}} \right)^{1/3} r, \quad (58)$$

where $M_{\text{pl}} = M_{\text{core}} + M_{\text{atm}}$ is the planet mass, M_{atm} being the mass of the atmosphere, and r is the orbital radius of the protoplanet in the disc.

3.2.2. Models enveloped by the disc with luminosity derived from planetesimal accretion. These models must match to the state variables of the accretion disc at the surface of the Roche lobe. We denote the disc midplane temperature, pressure and mass density at the distance r from the central star by T_m , P_m and ρ_m , respectively. To obtain the needed matching, we require that at $R = r_L$, the mass is equal to M_{pl} , the pressure is equal to P_m and the temperature is given by

$$T = \left(T_m^4 + \frac{3\tau_L L_{core}}{16\pi\sigma r_L^2} \right)^{1/4}, \quad (59)$$

where we approximate the additional optical depth above the protoplanet atmosphere, through which radiation passes, by

$$\tau_L = \kappa(\rho_m, T_m)\rho_m r_L. \quad (60)$$

For a particular disc model, at a chosen radius r , for a given core mass M_{core} and planetesimal accretion rate \dot{M}_{core} , Papaloizou and Terquem (1999) solve equations (50)–(52) with the boundary conditions described above. The opacity, taken from Bell and Lin (1994), has contributions from dust grains assumed to be ‘interstellar’, molecules, atoms and ions.

For a fixed \dot{M}_{core} at a given radius, there is a critical core mass M_{crit} above which no solution can be found, i.e. there can be no atmosphere in hydrostatic and thermal equilibrium confined between the radii r_{core} and r_L around cores with mass larger than M_{crit} , as explained above. For masses below M_{crit} , there are (at least) two solutions, corresponding to a low-mass and a high-mass envelope, respectively.

To estimate how large M_{core} should be before a significant atmosphere can start to form, we suppose it to be isothermal, which is the most favourable situation for retaining a massive atmosphere. When the atmosphere starts to form, we may neglect its mass, so that the gas density will be given by

$$\rho_b = \rho_m \exp\left(\frac{GM_{core}}{H^2\Omega^2 r_{core}}\right). \quad (61)$$

We estimate that the atmosphere can start to be significant when $4\pi r_{core}^3 \rho_b / 3 = f M_{core}$, where $f < 1$ is an estimated mass fraction contained in the atmosphere. This gives

$$M_{core} = \left(\frac{H^2\Omega^2}{G}\right)^{3/2} \left(\frac{3}{4\pi\rho_{core}}\right)^{1/2} \left(\ln\left(\frac{2fH\rho_{core}}{\Sigma}\right)\right)^{3/2}. \quad (62)$$

Taking $\Sigma = 200 \text{ g cm}^{-2}$, $H/r = 0.05$, $r = 5.2 \text{ AU}$ and $f = 0.01$, we obtain $M_{core} = 0.064 M_{\oplus}$.

Figure 6 shows the total protoplanet mass M_{pl} as a function of core mass M_{core} at different radii in a steady state disc with $\alpha = 10^{-2}$ (Shakura and Sunyaev 1973) and $\dot{M} = 10^{-7} M_{\odot} \text{ yr}^{-1}$ (see Papaloizou and Terquem (1999)). In each frame, the different curves correspond to planetesimal accretion rates in the range 10^{-11} – $10^{-6} M_{\oplus} \text{ yr}^{-1}$. The critical core mass is attained at the point where the curves start to loop backwards. It was demonstrated by Stevenson (1982) that the qualitative behaviour of the curves shown in figure 6 can be produced using a simplified model with only radiative energy transport, ideal gas equation of state and constant opacity. It thus appears to result from the core plus envelope nature of the models rather than from the details of the physics.

When the core first begins to gravitationally bind some gas, the protoplanet is on the left of the lower branch of these curves. Assuming \dot{M}_{core} to be constant, as the core and the atmosphere grow in mass, the protoplanet moves along the lower branch up to the right, until

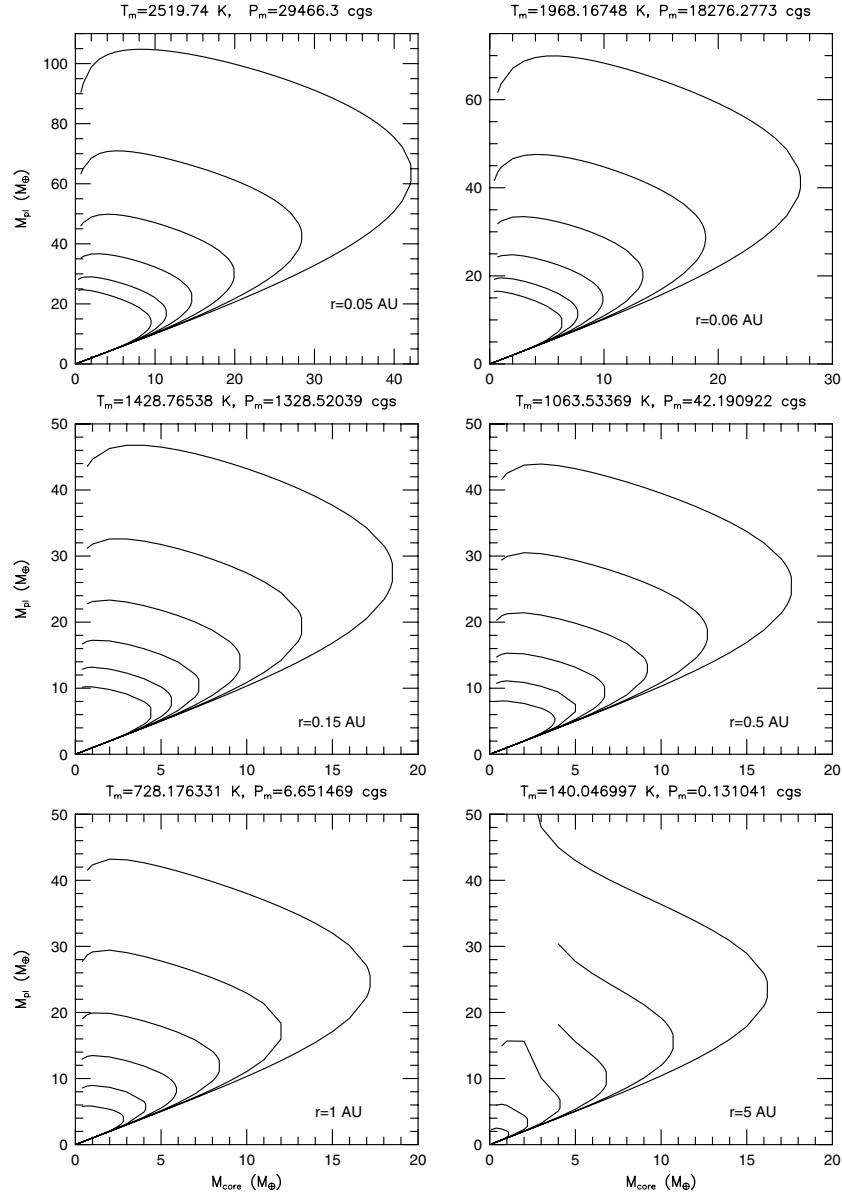


Figure 6. Plots of total mass M_{pl} , in units M_{\oplus} , as a function of core mass M_{core} , in units M_{\oplus} , at different locations r in a steady state disc model with $\alpha = 10^{-2}$ and gas accretion rate $\dot{M} = 10^{-7} M_{\odot} \text{ yr}^{-1}$. From left to right and top to bottom, the frames correspond to $r = 0.05$ AU, 0.06 AU, 0.15 AU, 0.5 AU, 1 AU and 5 AU, respectively. The midplane temperature and pressure at these locations are indicated above each frame. Each frame contains six curves which, moving from left to right, correspond to core luminosities derived from planetesimal accretion rates of $\dot{M}_{\text{core}} = 10^{-11} M_{\oplus} \text{ yr}^{-1}$, $10^{-10} M_{\oplus} \text{ yr}^{-1}$, $10^{-9} M_{\oplus} \text{ yr}^{-1}$, $10^{-8} M_{\oplus} \text{ yr}^{-1}$, $10^{-7} M_{\oplus} \text{ yr}^{-1}$ and $10^{-6} M_{\oplus} \text{ yr}^{-1}$, respectively. The critical core mass is attained when the curves first begin to loop backwards when moving from left to right. These plots are taken from Papaloizou and Terquem (1999).

the core reaches M_{crit} . At that point, the hydrostatic and thermal equilibrium approximation can no longer be used for the atmosphere, which begins to undergo very rapid contraction. Figure 6 indicates that when the core mass reaches M_{crit} , the mass of the atmosphere is comparable to that of the core, in agreement with Bodenheimer and Pollack (1986). Since the atmosphere in complete equilibrium is supported by the energy released by the planetesimals accreted onto the protoplanet, the critical core mass decreases as \dot{M}_{core} is reduced.

For $\alpha = 10^{-2}$ and $\dot{M} = 10^{-7} M_{\odot} \text{ yr}^{-1}$, the critical core mass at 5 AU varies between 16.2 and 1 M_{\oplus} as the planetesimal accretion rate varies between the largest and smallest value. The former result is in good agreement with that of Bodenheimer and Pollack (1986). Note that there is a tendency for the critical core masses to increase as the radial location moves inwards, the effect being most marked at small radii. At 1 AU, the critical mass varies from 17.5 to 3 M_{\oplus} as the accretion rate varies between the largest and smallest value, while at 0.05 AU these values increase still further to 42 M_{\oplus} and 9 M_{\oplus} , respectively.

In figure 7 we plot the critical core mass M_{crit} versus the location r for three different steady disc models. These models have $\alpha = 10^{-2}$ and $\dot{M} = 10^{-7} M_{\odot} \text{ yr}^{-1}$, $\alpha = 10^{-2}$ and $\dot{M} = 10^{-8} M_{\odot} \text{ yr}^{-1}$ and $\alpha = 10^{-3}$ and $\dot{M} = 10^{-8} M_{\odot} \text{ yr}^{-1}$, respectively. Here again, in each frame, the different curves correspond to planetesimal accretion rates in the range 10^{-11} – $10^{-6} M_{\oplus} \text{ yr}^{-1}$. Similar qualitative behaviour is found for the three disc models, but the critical core masses are smaller for the models with $\dot{M} = 10^{-8} M_{\odot} \text{ yr}^{-1}$, being reduced to 27 M_{\oplus} and 6 M_{\oplus} at 0.05 AU for the highest and lowest accretion rate, respectively.

3.2.3. Models enveloped by the disc with no planetesimal accretion. For these models, subsequently denoted as of type A, just as for those with planetesimal accretion as energy source, we assume that the structure extends to the Roche lobe or boundary of the Hill sphere beyond which the material must be gravitationally unbound from the protoplanet. There it matches to the external disc. Thus the boundary conditions for these models are the same as those with planetesimal accretion as energy source.

Papaloizou and Nelson (2005) adopt disc parameters appropriate to 5 AU from the disc model of Papaloizou and Terquem (1999) with $\alpha = 10^{-3}$ (Shakura and Sunyaev 1973) and steady state accretion rate of $10^{-7} M_{\oplus} \text{ yr}^{-1}$. Accordingly, $T_{\text{m}} = 140.047 \text{ K}$ and $P_{\text{m}} = 0.131 \text{ dyn cm}^{-2}$.

For models of type A, specification of the core mass and the total mass M_{pl} is enough to enable it to be constructed. This is because, although the luminosity is apparently an additional parameter, the degree of freedom associated with it is specified by the requirement that the radius of the planet atmosphere should be the Hill radius which, for a fixed orbital radius, depends only on M_{pl} . Thus, the luminosity and the total energy of the protoplanet may be written as $L = L(M_{\text{pl}})$ and $E = E(M_{\text{pl}})$.

For small changes in mass, the change in E is $dE = (dE/dM_{\text{pl}})dM_{\text{pl}}$. If the change occurs over a time interval dt , we must have

$$\frac{dE}{dM_{\text{pl}}} \frac{dM_{\text{pl}}}{dt} = -L. \quad (63)$$

This determines how the mass and hence the luminosity of the protoplanet vary with time and therefore the evolutionary track is specified.

3.2.4. Models with free boundary accreting from the protostellar disc. In contrast to embedded models of type A, we can consider models that have boundaries detached from and interior to the Roche lobe which still accrete material from the external protoplanetary

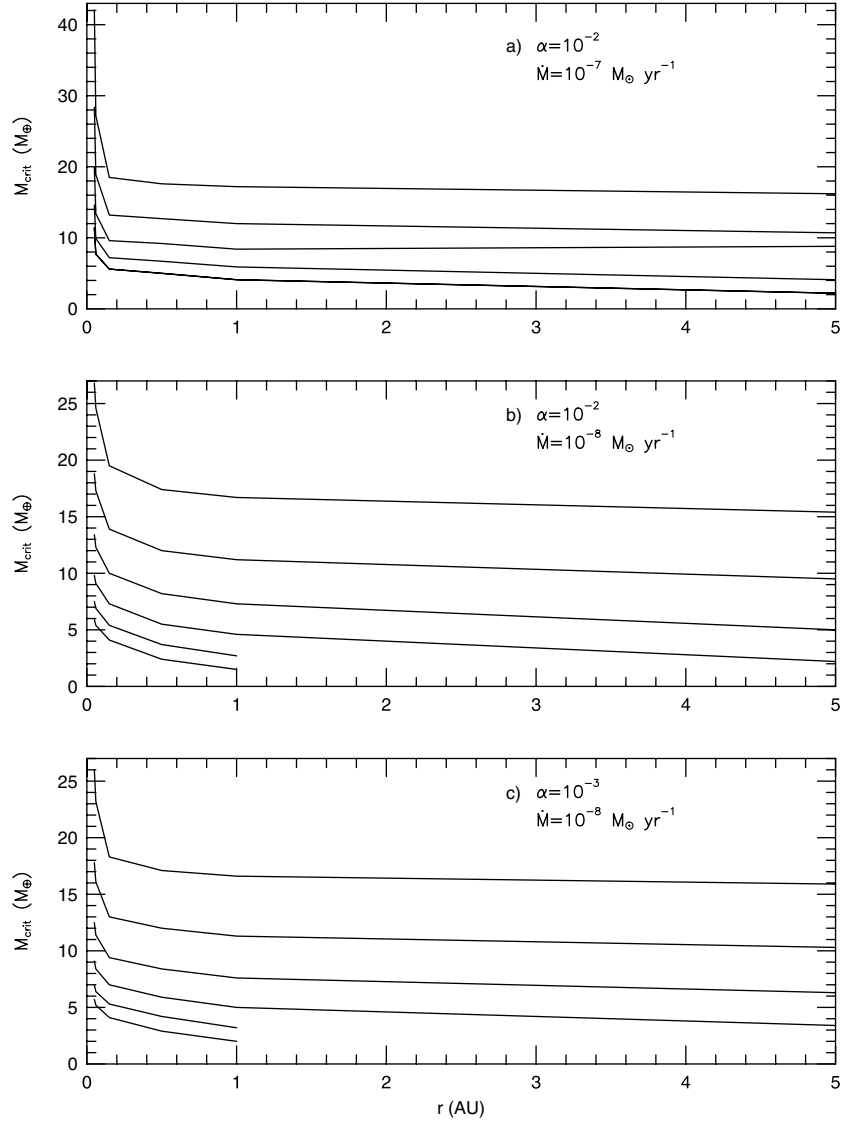


Figure 7. Critical core mass M_{crit} , in units M_{\oplus} , versus location r , in units AU, in a steady state disc model with viscous parameter α and gas accretion rate \dot{M} . The different plots correspond to $\alpha = 10^{-2}$ and $\dot{M} = 10^{-7} M_{\odot} \text{ yr}^{-1}$ (a), $\alpha = 10^{-2}$ and $\dot{M} = 10^{-8} M_{\odot} \text{ yr}^{-1}$ (b) and $\alpha = 10^{-3}$ and $\dot{M} = 10^{-8} M_{\odot} \text{ yr}^{-1}$ (c). Each plot contains six curves which, moving from bottom to top, correspond to core luminosities derived from planetesimal accretion rates of $\dot{M}_{\text{core}} = 10^{-11} M_{\oplus} \text{ yr}^{-1}$, $10^{-10} M_{\oplus} \text{ yr}^{-1}$, $10^{-9} M_{\oplus} \text{ yr}^{-1}$, $10^{-8} M_{\oplus} \text{ yr}^{-1}$, $10^{-7} M_{\oplus} \text{ yr}^{-1}$ and $10^{-6} M_{\oplus} \text{ yr}^{-1}$, respectively. These plots are taken from Papaloizou and Terquem (1999).

disc that orbits the central star. Detachment from the Roche lobe and the development of a free boundary are expected once the planet goes into a rapid accretion phase. This is also expected because numerical simulations of disc–planet interactions have shown that, once it becomes massive enough, a protoplanet forms a gap in the disc but is still able to accrete from it through a circumplanetary disc (see, e.g. Bryden *et al* (1999), Kley (1999), Lubow *et al* (1999), R P Nelson *et al* (2000)). We thus consider models with free boundaries which are

able to increase their mass and liberate gravitational energy through its settling. These models are referred to as type B.

In contrast to models of type A, for an assumed externally supplied accretion rate, models of type B form a two parameter family in which, without specification of the energy source, and given their freedom to determine their own radius, they require specification of both M_{pl} and L in order for a model to be constructed. Thus $E = E(M_{\text{pl}}, L)$. Accordingly, for small changes in mass and luminosity, the change in E is $dE = (\partial E / \partial M_{\text{pl}}) dM_{\text{pl}} + (\partial E / \partial L) dL$.

Now, for these models, matter is presumed to join the protoplanet on its equator after having accreted through a circumplanetary disc. In this case, we assume the accretion rate to be prescribed by the dynamics of the disc–planet interaction while gap formation is taking place. This is in agreement with simulations of disc–planet interactions where it is found that an amount of material comparable to that flowing through the disc may be supplied to the protoplanet (Bryden *et al* (1999), Kley (1999), Lubow *et al* (1999), R P Nelson *et al* (2000) and simulations presented in section 4). On arriving at the planet equator, all available gravitational binding energy of $-GM_{\text{pl}}/r_s$ per unit mass, r_s being the surface radius, has been liberated, and so an amount of energy $-GM_{\text{pl}} dM_{\text{pl}}/r_s$ must be subtracted from dE in order to obtain the energy available to replace radiation losses. Therefore, if the changes occur over an interval dt , we must have

$$dE + \frac{GM_{\text{pl}} dM_{\text{pl}}}{r_s} = \frac{\partial E}{\partial M_{\text{pl}}} dM_{\text{pl}} + \frac{\partial E}{\partial L} dL + \frac{GM_{\text{pl}} dM_{\text{pl}}}{r_s} = -L dt. \quad (64)$$

Thus, total energy conservation for models of type B enables the calculation of evolutionary tracks through

$$\left[\frac{\partial E}{\partial M_{\text{pl}}} + \frac{GM_{\text{pl}}}{r_s} \right] \frac{dM_{\text{pl}}}{dt} + \frac{\partial E}{\partial L} \frac{dL}{dt} = -L. \quad (65)$$

Note that, as we regard the accretion rate dM_{pl}/dt as specified for these models, equation (65) enables the evolution of L to be calculated.

Thus, equations (63) and (65) constitute the basic equations governing the evolution of models of type A and type B, respectively.

Note that any input from planetesimal accretion during and after the phase when the core becomes critical is neglected. We recall that after this point planetesimal accumulation proceeds by oligarchic growth (Ida and Makino 1993, Kokubo and Ida 1998). N-body simulations of protoplanetary core formation indicate that obtaining cores of a high enough mass $\sim 15 M_{\oplus}$ to obtain evolutionary timescales within the gas disc lifetime is not an easy task to achieve during the oligarchic growth phase. This is in part due to planetary cores of a few M_{\oplus} repelling the surrounding planetesimals and opening gaps in the planetesimal disc (e.g. Thommes *et al* (2002)). In addition, it is reasonable to suppose that the formation of a series of neighbouring cores may deplete planetesimals in the local neighbourhood. We also note that if the slowdown in evolution rate due to the long thermal timescale of the atmosphere can be obviated, a depletion of planetesimals brings about a favourable situation for shortening the evolutionary timescale as a potential energy source for sustaining the models without gas accretion is removed. The thermal timescale may be reduced by reducing the contribution of grain opacity from the ‘interstellar value’ (see figure 8, Ikoma *et al* (2000) and Papaloizou and Nelson (2005)). If the grain opacity in the outer layers of the atmosphere can be reduced by a factor of 100, the evolutionary timescale for a protoplanet with a $5 M_{\oplus}$ core can be reduced from 3×10^8 yr to a value within the expected gas disc lifetime. Such a reduction may occur either because of grain depletion or because the grains have been accumulated into particles of larger sizes than the original (Ikoma *et al* 2000).

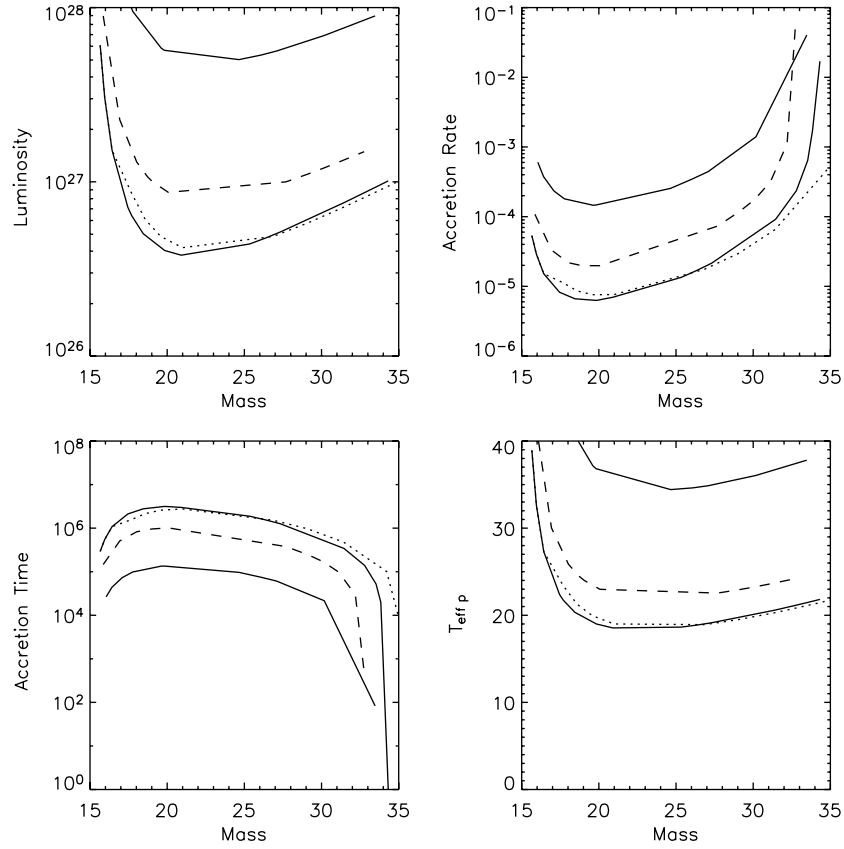


Figure 8. This figure illustrates the evolution of protoplanet models of **type A** which maintain contact with the protoplanetary disc and fill their Roche lobes while they accrete from it. They have fixed solid core masses of $15 M_{\oplus}$ and are situated at 5 AU. The upper left panel shows the luminosity, in cgs units, as a function of their increasing mass M , in cgs units. The upper right panel shows the gas accretion rate, in $M_{\oplus} \text{ y}^{-1}$, as a function of mass, while the lower left panel gives the accretion time M/\dot{M} , in year, as a function of mass. The lower right panel gives the temperature $T_{\text{eff p}}$, being the radiation temperature on the Roche lobe required account for the luminosity of the protoplanet, as a function of mass. The model illustrated with the full curve is embedded in a standard disc while the model illustrated with a dotted curve is embedded in a disc with the same temperature but with a density ten times larger. These two protoplanet models with standard opacity show very similar behaviour indicating lack of sensitivity to the detailed boundary conditions. In addition, we illustrate two models with this core mass embedded in a standard disc but with opacities which have reduction factors of ten and one hundred (dashed curves and triple dot dashed curves, respectively) that are constant for $T < 1600 \text{ K}$ and which then decrease linearly to unity at $T = 1700 \text{ K}$. Note the rapid evolutionary speed up that occurs when these models attain $\sim 35 M_{\oplus}$. At this point, detachment from the Roche lobe and transition to a model of type B is expected. (These plots are taken from Papaloizou and Nelson (2005).)

However, regardless of opacity, it is found that the evolutionary timescales for the type A models decrease rapidly once the gas mass is about twice the core mass (see figure 8). At this point, detachment from the Roche lobe and transition to a model of type B is expected (Papaloizou and Nelson 2005). Some of these are illustrated in figure 9. It is found that these models rapidly attain a radius $\lesssim 2 \times 10^{10} \text{ cm}$ for any reasonable accretion rate from the disc. Furthermore, these protoplanets may attain luminosities of $\sim 10^{-3} - 10^{-4} L_{\odot}$ for a time $\sim 10^5 - 10^6 \text{ yr}$.

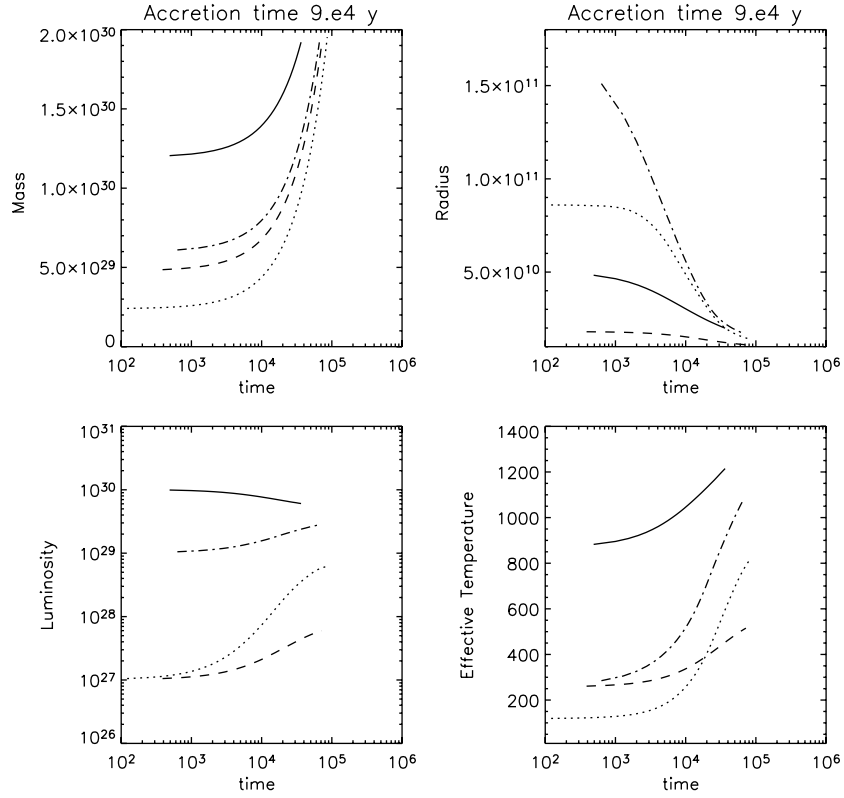


Figure 9. This figure illustrates the evolution of protoplanet models of **type B** which accrete from the protoplanetary disc at an assumed rate of one Jupiter mass in 9×10^4 yr but are detached from their Roche lobes. They have fixed solid core masses of $15 M_{\oplus}$ and are situated at 5 AU. The upper left panel shows the planet mass, in cgs units, as a function of time, in year. The upper right panel shows the protoplanet radius, in cgs units, as a function of time. The lower left panel gives the luminosity, in cgs units, as a function of time. The lower right panel gives the effective temperature as a function of time. The four models shown correspond to differing initial conditions corresponding to different starting masses and luminosities. The resulting evolutionary tracks tend to show convergence as time progresses. (These plots are taken from Papaloizou and Nelson (2005).)

3.3. Formation of giant planets: gravitational instabilities

Gravitational instabilities in protostellar discs have been considered as a mechanism for forming giant planets since Kuiper (1949) proposed that planets were formed through gravitational breakup of gaseous rings. The advantage of this mechanism is that it leads to giant planets on a very short timescale, on the order of the dynamical timescale. This is an important argument, as we have seen above that the build-up of a massive core may be prohibitively long in some parts of the protostellar nebula. Also, some of the extrasolar planets discovered so far are very massive, with a mass several times that of Jupiter. It is not clear that these objects can form through the core-accretion scenario, as gap formation inhibits gas accretion above some critical mass (see below).

It was initially believed that axisymmetric instabilities in discs would lead to the formation of rings that would further break-up into condensations. Using the theory of viscous accretion discs developed by Lynden-Bell and Pringle (1974), Cameron (1978) calculated evolution models of the primitive solar nebula which were unstable against axisymmetric perturbations.

The parameter governing the importance of disc self-gravity is the Toomre Q parameter which, in a Keplerian disc, can be written as $Q \simeq \Omega c_s / (\pi G \Sigma) \sim M_\star H / (M_d r)$, with M_d being the disc mass contained within radius r , H the disc semi-thickness and c_s the sound speed. Typically $H/r \sim 0.1$ for protostellar discs (Stapelfeldt *et al* 1998) such that the condition for the importance of self-gravity, $Q \sim 1$, gives $M_d \sim 0.1 M_\star$. The characteristic scale associated with growing density perturbations in a disc undergoing gravitational instability with $Q \sim 1$ is on the order of H , and the corresponding mass scale is $M_d (H/r)^2 \sim M_\star (H/r)^3$, which is on the order of one Jupiter mass for $H/r \sim 0.1$ and $M_\star = 1 M_\odot$.

There are, however, problems with forming planets through gravitational instabilities. Fragmentation, which follows the development of axisymmetric modes, requires $Q \leq 1$, whereas non-axisymmetric modes develop for values of Q slightly larger than unity. As noted by Laughlin and Bodenheimer (1994), it is unlikely that a disc would ever get massive enough for Q to decrease below unity. The disc builds up in mass as a result of the collapse of an envelope. The Q parameter, therefore, decreases with time from high values. When it reaches values slightly larger than unity, non-axisymmetric (spiral) modes develop. These lead to outward angular momentum transport (Papaloizou and Savonije 1991, Heemskerk *et al* 1992, Laughlin and Bodenheimer 1994, Pickett *et al* 1998) that results in additional mass growth of the central star. Therefore Q is decreased and gravitational stability is restored. Since the redistribution of mass may occur on the dynamical timescale (a few orbits) of the outer part of the disc, it follows that Q can never decrease to below unity unless the disc can be cooled down (so that c_s decreases) or matter can be added (so that Σ increases) on the dynamical timescale, which is unlikely to happen in the regions of planet formation (it could happen, however, in the optically thin parts of the disc, typically beyond ~ 40 AU).

Nonetheless, condensations may form even if the disc is unstable only against non-axisymmetric perturbations, provided the density gets high enough in the spiral arms (Boss 1998, Mayer *et al* 2004). There are indications, however, that the clumps that appear in the numerical simulations of gravitationally unstable discs are not long-lived (see, e.g. Durisen (2001) and references therein). The outcome of gravitational instabilities is very sensitive to the thermal physics of the gas. Chemical composition dependence is expected but the correlation of the existence of planets with metallicity is yet to be demonstrated. When the equation of state is adiabatic, heating of the gas prevents the formation of clumps. In contrast, condensations form in an isothermal disc. But if the outer radius of the disc is allowed to expand under the effect of gravitational torques, these clumps tend to disappear, dissolving back into the gas (Pickett *et al* 1998, 2000a, 2000b, Boss 2000). Simulations including heating via dynamical processes and radiative cooling produce spiral structures that are less distinct than in isothermal discs and that do not collapse into condensed objects (A F Nelson *et al* 2000). Although the question of whether the clumps can survive or not is still being debated, it seems unlikely that Jupiter mass planets can be formed through gravitational instabilities in optically thick regions of discs. However, as noted in section 3.1.2, the accumulation of solid particles into planetesimals may be assisted by gravitational instabilities in the gas disc through particle accumulation at pressure maxima in spiral structures (Rice *et al* 2004, Durisen *et al* 2005).

Note that fragmentation may also occur before a disc is completely formed, during the initial collapse of the protostellar envelope. Such opacity limited fragmentation has been estimated to produce objects with a lower mass limit of 7 Jupiter masses (Low and Lynden-Bell 1976, Masunaga and Inutsuka 1999), but there is no definitive argument to rule out somewhat smaller masses (Bodenheimer *et al* 2000). It is possible that both a disc and fragments may form simultaneously out of the protostellar envelope, the relative importance of the two processes depending, for instance, on the angular momentum content of the envelope, on the strength

of any magnetic field (so far neglected in disc fragmentation calculations) and possibly on the initial clumpiness. Large scale observations of class 0 envelopes, so far, do not rule out the presence of clumps with masses smaller than about 10 Jupiter masses (Motte and André 2001). To explain the large eccentricities of many of the extrasolar planets, Papaloizou and Terquem (2001) and Terquem and Papaloizou (2002) have considered a scenario in which a population of planetary mass objects are assumed to form rapidly through a fragmentation process occurring in a disc or protostellar envelope on a scale of 100 AU (see section 5).

4. Disc–planet interactions and orbital migration

The discovery of the ‘hot Jupiters’ led to the realization of the importance of orbital migration. This comes about because of the perceived difficulties in accumulating a core directly at small radii interior to the so called ‘snow line’ (Bodenheimer *et al* 2000). Instead, it is possible that the protoplanet formed at larger radii and migrated inwards.

So far, three mechanisms have been proposed to explain the location of planets at very short orbital distances. One of them relies on the gravitational interaction between two or more Jupiter mass planets, which may lead to orbit crossing and to the ejection of one planet while the other is left in a smaller orbit (Rasio and Ford 1996, Weidenschilling and Marzari 1996). However, this mechanism cannot account for the relatively large number (about 20%) of short-period planets observed. Another mechanism is the so-called ‘migration instability’ (Murray *et al* 1998, Malhotra 1993). It involves resonant interactions between the planet and planetesimals located inside its orbit which lead to the ejection of a fraction of them while simultaneously causing the planet to migrate inwards. To move a Jupiter mass planet from about 5 AU down to very small radii through this process, a disc containing about 1 Jupiter mass of planetesimals, and thus about $0.1 M_{\odot}$ of gas, inside 5 AU is required. Such a massive disc is unlikely and furthermore it would be only marginally gravitationally stable. The third mechanism, that we are going to focus on here, involves the tidal interaction between the protoplanet and the gas in the surrounding protoplanetary nebula (Goldreich and Tremaine (1979, 1980), Lin and Papaloizou (1979, 1993) and references therein, Papaloizou and Lin (1984), Ward (1986, 1997b)).

Below we review the possible main types of orbital migration associated with this tidal interaction. These are (i) type I migration, which applies to an embedded protoplanet of small mass for which the disc response can be considered using linear analysis (e.g. Ward (1997)), (ii) type II migration, which applies when the protoplanet is massive enough to open a gap (Lin and Papaloizou 1986) and (iii) runaway or type III migration (Masset and Papaloizou 2003, Artymowicz 2004), which is a new form of potentially fast migration, applicable to massive discs, that could be driven by coorbital torques.

Orbital migration in each of these cases is induced by the angular momentum exchange with disc material that occurs through disc–planet interaction. This may be through wave excitation, as in type I migration, a combination of wave excitation and shock dissipation, as occurs in type II migration, or direct exchange with disc material traversing the orbit, as has been suggested for type III migration.

In the [appendix](#) we give a simple description of the global angular momentum balance and conservation when a protoplanet in circular orbit is embedded in a gaseous disc with which it interacts gravitationally and consequently migrates radially. We show that the angular momentum exchange with the protoplanet can be accounted for by considering viscous and wave fluxes of angular momentum, that can be measured a long way from the protoplanet, as well as the difference between the rate of advection of angular momentum by gas flowing through the orbit and the rate of increase in the angular momentum content of regions in the

neighbourhood of the protoplanet because of non-steady conditions. The angular momentum carried in wave/viscous fluxes is associated with type I and type II migration, whereas the advected flux is associated with the more recently proposed type III migration.

We consider the timescales for the various migration processes and the potential problems introduced for planet formation theory.

4.1. Angular momentum transfer due to scattering and type II migration

4.1.1. The impulse approximation applied to a protoplanet in a gap. Here, we adopt a very simple picture of the angular momentum exchange process between a protoplanet in circular orbit and the disc. We work in a frame in which the protoplanet appears at rest at the origin. For a protoplanet in circular orbit, this frame would have to be in a state of uniform rotation and we should include centrifugal and Coriolis forces to describe the motion of the disc matter. However, we shall neglect these and make a local approximation, supposing that a particle of disc matter approaches the protoplanet from a large distance moving in a straight line with impact parameter a which we shall take to be much less than the orbital radius of the protoplanet (Lin and Papaloizou 1979).

On passing the protoplanet, the gravitational interaction produces an angle of deflection δ between the asymptotic tangents to the trajectory of the passing element of disc matter. In the local approximation, the disc matter recedes from the protoplanet along a straight line inclined to the line of approach by an angle δ .

We adopt local Cartesian coordinates with origin at the centre of mass of the protoplanet. The x -axis points radially outwards, the y -axis points in the azimuthal direction in the direction of rotation while the z -axis points in the vertical direction. We may assume that the initial trajectory is along $x = a$ approaching from $y = -\infty$ with relative speed u . This corresponds to a disc with circular streamlines.

As a result of the interaction, a velocity component v_x , parallel to the x -axis, is induced. This can be found from the x -component of the equation of motion in the form:

$$\frac{dv_x}{dt} = -GM_p x / (x^2 + y^2)^{3/2},$$

where M_p is the planet mass. Assuming the interaction to be weak, we set $x = a$, $y = ut$ and integrate with respect to time between $-\infty$ and $+\infty$. Hence, the final induced x -velocity is

$$v_x = -GM_p a \int_{-\infty}^{\infty} \frac{dt}{[a^2 + (ut)^2]^{3/2}} = -\frac{2GM_p}{au}.$$

The angle of deflection is thus simply

$$\delta = \left| \frac{v_x}{u} \right| = \frac{2GM_p}{au^2}.$$

Associated with this interaction is a transfer of angular momentum between the disc matter and protoplanet. To see this, we first recall that, as the interaction between protoplanet and disc is conservative, energy conservation implies that the ultimate speed of recession is also u . Accordingly, there must have been a parallel momentum transfer per unit mass of magnitude $\Delta u = u(1 - \cos \delta) \simeq u\delta^2/2$. If the orbital radius of the protoplanet is R , the associated angular momentum transferred is

$$\Delta J = Ru\delta = \frac{2RG^2M_p^2}{a^2u^3}. \quad (66)$$

We should, of course, note that this transfer is directed along the original direction of motion as seen by the protoplanet. Thus, for disc matter on circular orbits interior to the protoplanet,

the transfer is positive, while for exterior disc matter, the transfer is in the opposite sense and negative. The interaction is accordingly frictional, with interior matter speeding up the protoplanet and exterior matter slowing it down.

Note too that the trajectory prior to the local scattering interaction is assumed linear (corresponding to circular orbit when viewed globally). Subsequent returns of disc matter on circular orbit are required in order to make the frictional interaction persistent. The scattering itself, of course, disturbs that situation, so that there is an implicit assumption that dissipative or other processes work to restore circular orbits for returning trajectories. The indications from many numerical simulations are that disc viscosity is able to provide such an effect.

The simple calculation of ΔJ given above neglects the fact that the calculation should be done in a rotating frame. However, this has been incorporated in calculations by Goldreich and Tremaine (1980) and Lin and Papaloizou (1993). As these calculations add nothing new with regard to the essential physics, we refer the reader to those works for the details. The end result is that the expression given in equation (66) is multiplied by a correction factor

$$C_0 = \frac{4}{9}[2K_0(2/3) + K_1(2/3)]^2,$$

where K_0 and K_1 denote modified Bessel functions. Thus

$$\Delta J = C_0 \frac{2RG^2M_p^2}{a^2u^3}. \quad (67)$$

4.1.2. Angular momentum exchange rate with the disc. We may use equation (67) to evaluate the total angular momentum exchange rate with the disc. To do this, we replace the relative speed u by $u = R|\Omega - \omega|$, where ω and Ω are the angular rotation rates of the protoplanet and disc matter, respectively. Given that we have implicitly assumed the interaction occurs close to the protoplanet, we adopt the first order expansion $|\Omega - \omega| = 3a\omega/(2R)$.

Furthermore, each disc particle orbiting a distance a from the protoplanet suffers impulses separated by a time interval $2\pi/|\Omega - \omega|$. We can thus write down the rate of transfer of angular momentum between the disc matter interior to the protoplanet orbit and the protoplanet by integrating over the disc mass as

$$\frac{dJ}{dt} = \int_{a_{\min}}^{\infty} R \Sigma \Delta J |\Omega - \omega| da. \quad (68)$$

This gives

$$\frac{dJ}{dt} = \frac{8G^2M_p^2R\Sigma C_0}{27\omega^2a_{\min}^3}, \quad (69)$$

where a_{\min} is the minimum distance between disc matter and the protoplanet orbit. Thus, there is an inherent assumption of a gap in the disc, or lack of coorbital material, which would have to be taken into account if the protoplanet is embedded in the disc and the interaction is linear. Therefore, the above analysis is applicable only in a nonlinear regime when there is a significant gap, corresponding to type II migration. Note that the sign is positive for the more rapidly moving inner disc material. For the outer disc material, one can see from symmetry considerations that the same expression applies but with a sign reversal because it is more slowly moving than the protoplanet and thus attempts to slow it down.

4.1.3. Gap formation. A gap forms because angular momentum exchange occurs locally in the disc, on each side of the protoplanet, in such a way as to cause the disc material to move away from the protoplanet.

In order for a gap to be maintained, we should expect that the angular momentum exchange rate given by equation (69) is able to balance the angular momentum flow rate due to viscosity. The viscous flow rate is given by

$$\left(\frac{dJ}{dt}\right)_{\text{visc}} = 3\pi\nu\Sigma R^2\omega, \quad (70)$$

where ν is the kinematic viscosity and Σ is the surface density of gas in the disc. Of course, we recall that ν is not a real ‘viscosity’ but a way of representing the flow of angular momentum resulting from turbulent stresses. From what we have indicated above, we expect the condition for gap formation to be

$$\frac{dJ}{dt} > \left(\frac{dJ}{dt}\right)_{\text{visc}}. \quad (71)$$

This is equivalent to

$$\frac{8M_p^2 R^3 C_0}{81\pi M_\star^2 a_{\min}^3} > \frac{\nu}{R^2\omega}, \quad (72)$$

where M_\star is the mass of the central star, around which the planet orbits. To work with equation (72), we need to know what value of a_{\min} to use. Considerations of the pressure response in type I migration (Goldreich and Tremaine 1980, Artymowicz 1993) suggest that there should be a torque cut-off leading to $a_{\min} \sim H$. This is also a statement that pressure effects are able to smooth out smaller scale perturbations in the vicinity of the protoplanet. A similar conclusion was reached by Papaloizou and Lin (1984) who considered the linear response of a protoplanet orbiting in a disc already with a gap with lengthscale H , which was found to be the physical scale that led to the maximum angular momentum exchange rate.

However, the condition for strong nonlinearity, which is required for local gap formation, is that the radial lengthscale of the flow cannot be less than the Hill radius, r_H , which being the largest possible scale of significant influence by the protoplanet gravitational field (Lin and Papaloizou 1979), must exceed H . Accordingly, we should have $a_{\min} \geq r_H$. Taking $a_{\min} = 2r_H$ and absorbing uncertainties in the constant coefficient C_0 , equation (72) gives

$$\frac{3M_p C_0}{81\pi M_\star} > \frac{\nu}{R^2\omega}. \quad (73)$$

Taking $C_0 = 2/3$, this leads to

$$\frac{M_p}{M_\star} > \frac{40\nu}{R^2\omega}. \quad (74)$$

This condition is given by Lin and Papaloizou (1986, 1993). But note that it is only expected to be accurate to within a constant of order unity. For a typical disc model, $\nu/(R^2\Omega) \sim 10^{-5}$ and we expect gap formation for planets slightly above the mass of Saturn.

However, note that Rafikov (2002) argues that gap formation may occur for planets of mass as low as $2 M_\oplus$. So far, gap formation at such low masses has not been seen in simulations (e.g. D’Angelo *et al* (2003)). This may be because Rafikov (2002) suggested that a gap can be produced as the density waves excited by a planet become weak shock waves and dissipate. However, this occurs several scale heights away from the planet, beyond the region where a gap would be expected. Furthermore, the angular momentum deposition from the waves also extends over a region of several scale heights, being wider than an expected gap. In this context, note that an annulus of width $5H$ at 5.2 AU in a minimum mass solar nebula contains a Jupiter mass, so that a planet of several earth masses could not redistribute its angular momentum without significant migration.

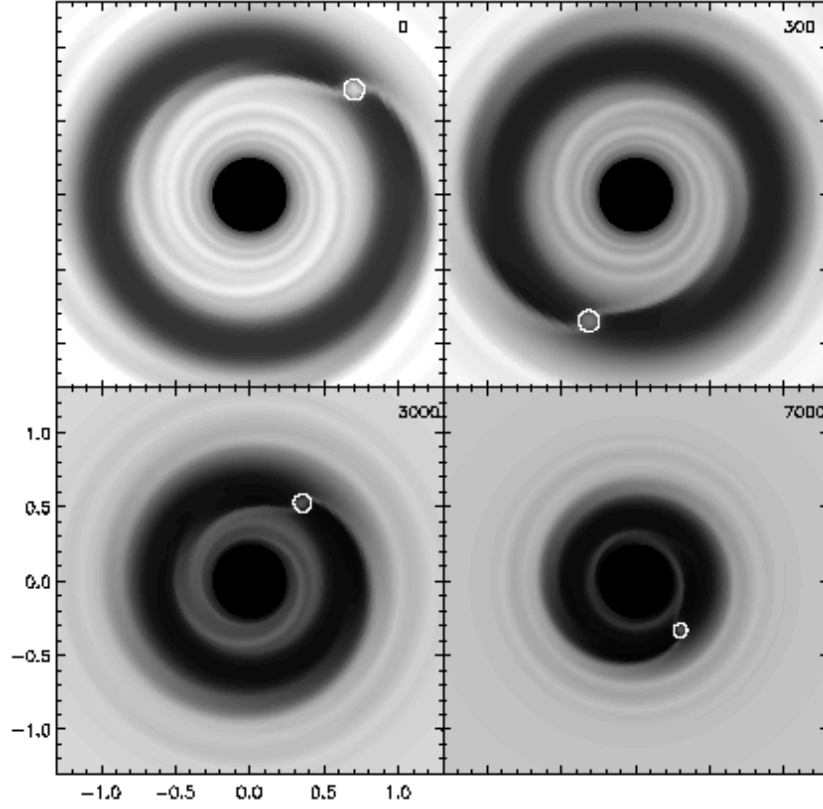


Figure 10. This figure shows the time evolution of the disc surface density contours and orbital radius as a function of time measured in orbital periods for a protoplanet with initial mass of $1 M_J$. As the protoplanet undergoes type II migration inwards, its mass grows to $3.5 M_J$ under an assumption of maximal accretion. (The plots are taken from R P Nelson *et al* (2000).)

4.1.4. Type II migration. When the conditions for gap formation are satisfied, the outward flux of angular momentum due to viscous transport in the outer disc section is supplied by the planet. Similarly, the outward flux from the inner section is taken up by the planet. When these contributions do not balance, the planetary orbit changes its angular momentum and thus migrates. Normally, simulations indicate that the outer disc has the dominant effect, causing inward migration. This is illustrated in figure 10, taken from R P Nelson *et al* (2000), which shows the evolution of a planet of initially $1 M_J$ in a standard disc. In this case, the protoplanet was allowed to accrete as it migrated inwards on the viscous timescale of $\sim 10^5$ yr. Inward migration on the viscous timescale R^2/ν is characteristic of type II migration driven by the evolution of the disc when the protoplanet mass does not exceed the local disc mass. In that case, using the standard α -prescription by Shakura and Sunyaev (1973), the migration timescale is given by

$$t_{\text{II}} (\text{yr}) = \frac{1}{3\alpha} \left(\frac{R}{H} \right)^2 \Omega^{-1} = \frac{0.05}{\alpha} \left(\frac{R}{H} \right)^2 \left(\frac{R}{1 \text{ AU}} \right)^{3/2}. \quad (75)$$

If we adopt $H/r = 0.1$ and consider α in the range 10^{-3} – 10^{-2} , we get $t_{\text{II}} \sim 10^3$ or 10^4 yr at $R = 1$ or 5 AU, respectively. These timescales are much shorter than the disc lifetime or estimated planetary formation timescales. The expression of t_{II} is independent of the

protoplanet's mass and the surface mass density of the disc, but it is implicitly assumed here that the mass of gas within the characteristic orbital radius of the planet is at least comparable to the mass of the planet itself. If the disc is significantly less massive, the inertia of the planet slows down the migration (Syer and Clarke 1995, Ivanov *et al* 1999).

4.1.5. *Stopping type II migration*

Stopping type II migration at small radii. Planets subject to type II migration would stop before falling onto the star if they entered a magnetospheric cavity. Tidal interaction with a rapidly rotating star would also halt planet orbital decay at a few stellar radii (where the interaction becomes significant; see Lin *et al* (2000) and references therein). Both these mechanisms have been put forth to account for the present location of the planet around 51 Pegasi (Lin *et al* 1996) and to explain the location of hot Jupiters more generally.

A planet overflowing its Roche lobe and losing part of its mass to the central star would also halt at small radii. This is because during the transfer of mass the planet moves outwards to conserve the angular momentum of the system (Trilling *et al* 1998). The planet stops at the location where its physical radius is equal to its Roche radius. Recent observations of atomic hydrogen absorption in the stellar Lyman α line during three transits of the planet HD209458b suggest that hydrogen atoms are escaping the planetary atmosphere (Vidal-Madjar *et al* 2003).

'*The last of the Mohicans*'. The mechanisms reviewed above cannot account for the presence of giant planets orbiting their parent star at distances larger than about a tenth of an AU.

It has been suggested that migration of a giant planet could be stopped at any radius if migration and disc dissipation were concurrent (Trilling *et al* 1998, 2002). If the disc dissipates while migration is taking place, then the drift timescale may increase in such a way that the planet stalls at some finite radius. Note, however, that this requires very fine tuning of the parameters (disc mass, disc lifetime etc), as for a given disc mass the migration (viscous) timescale given by equation (75) decreases as the orbital radius decreases. Also, a major problem with this mechanism is to explain how the disc dissipates. Within this scenario, there is initially enough gas in the disc to push the planet down to some orbital radius. For typical disc parameters, only part of this gas may be accreted by the planet or leak through the gap to be accreted onto the star (Bryden *et al* 1999, Kley 1999, Lubow *et al* 1999, R P Nelson *et al* 2000). It is therefore not clear how the gas disappears.

A giant planet could survive if after it formed there were not enough material left in the disc for significant migration to occur. It has been suggested that a series of giant planets could actually assemble in the disc and disappear onto the star (Gonzalez 1997, Laughlin and Adams 1997). Then at some point the disc mass may be such that one more planet can be formed but not migrate (Lin 1997). This survivor is sometimes referred to as the last of the Mohicans.

Note that the prediction that the more massive giant planets that are able to form gaps should be predominantly at larger radii is supported by the observations.

Planets locked in resonances. Orbital migration occurring for different planets at different rates, dependent on local disc parameters, can produce convergent migration and locking on to mean motion commensurabilities in a similar manner as is believed to occur as a result of tidally induced migration of satellites in the solar system (Goldreich 1965).

Pairs of extrasolar planets have been discovered in such commensurabilities. Gliese 876 (Marcy *et al* 2001) and HD 82943 (Mayor *et al* 2004) are in 2:1 commensurabilities while 55 Cancri (McArthur *et al* 2004) exhibits a 3:1 commensurability. The origin of these commensurabilities is satisfactorily accounted for by disc planet interaction (e.g. Snellgrove *et al* (2001), Lee and Peale (2002), Papaloizou (2003), Kley *et al* (2004)). However, there is no indication that the existence of these commensurabilities slowed down or halted inward migration. On the other hand, it is possible that for some planet masses and disc conditions,

resonant trapping of planets could lead to a reversal of type II migration. This has been suggested to occur when a giant planet (e.g. Jupiter) migrates inwards and captures a lighter Saturn mass planet in 2:3 resonance (Masset and Snellgrove 2001). However, it is unclear how often and for how long such situations can be realized.

We also comment that apsidal resonances in which the apsidal lines of two planet orbits are aligned, but not commensurable, could also be produced through disc planet interactions which are dissipative and therefore tend to select a particular normal mode involving non-zero eccentricity of the planetary and disc orbits jointly, while the latter is present (e.g. Papaloizou (2003)).

4.2. Linear torques and type I migration

In this case, one is concerned with protoplanets that are not massive enough to form a gap in the disc. We are therefore interested in calculating the linear response of the disc to an embedded orbiting protoplanet and using that to calculate the angular momentum exchange rate with the protoplanet.

The perturbation exerted by the protoplanet on the disc propagates as density waves outside the Lindblad resonances and is evanescent inside these resonances, in the corotation region. The protoplanet exerts a torque on the density waves, which, together with the torque exerted at corotation, is responsible for the exchange of angular momentum between the disc rotation and the planet's orbital motion. Away from the Lindblad resonances, the waves have a small wavelength, so that the contribution to the net torque is small. Therefore, most of the torque is exerted in the vicinity of the Lindblad resonances, where the perturbation has a large wavelength, or in the vicinity of the corotation resonance (see Terquem *et al* (2000) and references therein).

Thus, the torque exerted by the protoplanet on the disc can be calculated either by performing a full linear analysis of the excited waves in two or three dimensions (see, e.g. Korycansky and Pollack (1993), Tanaka *et al* (2002), Terquem (2003), Tanaka and Ward (2004)) or by summing up the contribution of point like torques exerted at Lindblad and corotation resonances (see, e.g. Goldreich and Tremaine (1979), Artymowicz (1993), Ward (1997), Papaloizou and Larwood (2000)). The results of full numerical simulations in general confirm the linear analysis (e.g. D'Angelo *et al* (2003), Bate *et al* (2003)). Both analytical and numerical calculations find a positive (negative) torque on the disc (protoplanet) when the disc is laminar, circular and does not contain any magnetic field. This corresponds to inward migration of the protoplanet.

The conditions for non-linearity, or gap formation, have been explored by Papaloizou *et al* (2004) using both local shearing box simulations and fully global simulations. A sequence of local shearing box simulations, incorporating MHD turbulence with zero net magnetic flux, which have increasing $M_p R^3 / (M_* H^3)$, is illustrated in figure 11. The transition from being embedded to opening a gap is seen to occur when this parameter is around unity and it corresponds to the nonlinearity condition mentioned above that the radius of the Hill sphere should exceed the disc thickness. When this condition is satisfied, the viscous transport condition given by equation (74) is also satisfied in these simulations.

4.2.1. Timescales. The resulting dependence of the linear (inward) migration time t_1 and eccentricity damping time t_e , calculated for a laminar inviscid disc as a sum of contributions from Lindblad resonances, on the gas disc thickness is such that $t_1 \equiv |J/\dot{J}| \propto (H/R)^2$ and $t_e \equiv |e/\dot{e}| \propto (H/R)^4$. This scaling results in t_e being much shorter than t_1 . Accordingly, we expect circular orbits to be set up quickly and then migrate. But note that this may not be

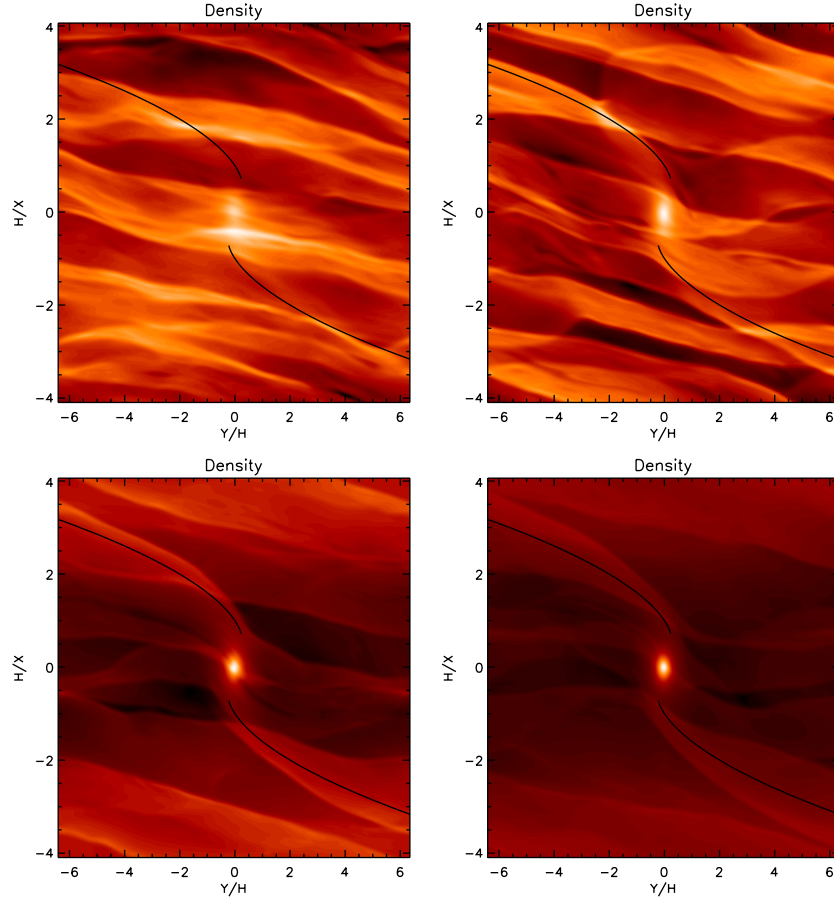


Figure 11. Density contour plots in a typical (x, y) plane for simulations with $M_p R^3 / (M_* H^3) = 0.1$ (top left panel), 0.3 (top right panel), 1.0 (bottom left panel) and 2.0 (bottom right panel). As $M_p R^3 / (M_* H^3)$ increases, the wake becomes more prominent, material is accreted by the protoplanet and is pushed towards the radial boundaries as a gap is formed. The location of the characteristic ray defining the wake is also plotted (see Papaloizou *et al.* (2004)). The qualitative structure seen in these simulations is maintained once a quasi-steady state has been attained. (These plots are taken from Papaloizou *et al.* (2004).)

the case when the disc is eccentric. Then the disc interaction is expected to lead to a finite eccentricity in the protoplanet orbit (Papaloizou 2002). Furthermore, the migration direction may be affected.

Tanaka *et al.* (2002) performed a local three dimensional calculation for a planet in circular orbit embedded in an isothermal axisymmetric disc that summed contributions from both Lindblad and corotation resonances introduced through the vertical dependence. When $\Sigma \propto R^{-\gamma}$ and there is a central solar mass, they obtained

$$t_1 = 2(2.7 + 1.1\gamma)^{-1} \frac{M_\odot}{M_p} \frac{M_\odot}{\Sigma R^2} \left(\frac{H}{R} \right)^2 \Omega^{-1}. \quad (76)$$

Results obtained from three-dimensional calculations can be reproduced in two-dimensional flat disc calculations if a gravitational softening parameter is adopted. When softening is introduced, the $1/R$ potential for the protoplanet is changed to $1/(R^2 + b^2)^{1/2}$, the softening

parameter being b . This can be regarded as accounting for the vertical averaging of the potential over the vertical thickness, and accordingly b should be $\sim H$.

We note that for their two-dimensional calculations, Papaloizou and Larwood (2000) found that the torque results are dependent on the gravitational softening parameter used. Then t_1 and t_e scale as $b^{1.75}$ and $b^{2.5}$, respectively, for b in the range $0.4H-H$. Papaloizou and Larwood (2000) find the approximate fits to t_1 and t_e when there is a central solar mass and $\gamma = 3/2$ to be

$$t_1 \text{ (yr)} = 3.5 \times 10^5 f_s^{1.75} \left[1 + \left(\frac{eR}{1.3H} \right)^5 \right] \left[1 - \left(\frac{eR}{1.1H} \right)^4 \right]^{-1} \times \left(\frac{H/R}{0.07} \right)^2 \left(\frac{2M_J}{M_{\text{GD}}} \right) \left(\frac{M_{\oplus}}{M_p} \right) \left(\frac{R}{1 \text{ AU}} \right), \quad (77)$$

and

$$t_e \text{ (yr)} = 2.5 \times 10^3 f_s^{2.5} \left[1 + \frac{1}{4} \left(\frac{e}{H/R} \right)^3 \right] \times \left(\frac{H/R}{0.07} \right)^4 \left(\frac{2M_J}{M_{\text{GD}}} \right) \left(\frac{M_{\oplus}}{M_p} \right) \left(\frac{R}{1 \text{ AU}} \right). \quad (78)$$

Here the gas disc has a mass M_{GD} contained within 5 AU and $f_s \equiv 2.5b/H$, where b is the gravitational softening length. Equations (76) and (77) agree when $f_s = 2.3$ and give the characteristic timescale for migration of an Earth mass core in the minimum mass solar nebula as about 10^6 yr. This timescale decreases linearly with mass, resulting in a survival problem for proto-Jovian mass planet cores. In this context, even the smallest cores considered of $5 M_{\oplus}$ have an infall timescale of only 2×10^5 yr.

The resolution of this problem may be provided by magnetic fields. When they are present and produce turbulence with density fluctuations that can be of order unity, type I migration may be stochastic (e.g. Nelson and Papaloizou (2004), Laughlin *et al* (2004)). In this case, the migration direction of small mass cores becomes indeterminate over the timescales; simulations can be performed and the final outcome may depend on long term global fluctuations in the MHD turbulence which in turn may depend on the global disc environment. We comment here that it is not clear that the disc behaves as a laminar disc with the time averaged density, upon which noisy fluctuations, with a characteristic time comparable to the orbital period and which obey Gaussian statistics, are superposed. If that were the case, it would be very difficult to affect the long term trend determined by the average laminar disc because of the large ratio of disc lifetime to orbital period. An additional recent investigation by Nelson (2005) suggests that turbulent fluctuations occur on all timescales up to the global diffusion or viscous time. Furthermore, the background perturbed by the planet is very different from laminar, with large density fluctuations encountering the planet and providing stochastic torques even when unperturbed by it. Consider a density fluctuation of order unity with lengthscale of order H , a distance of order H from the planet. The characteristic specific torque acting on it is $G\Sigma R$. Given the stochastic nature of turbulence, one might expect the specific torque acting on the planet to oscillate between $\pm G\Sigma R$. Note that this exceeds the net specific torque implied by equation (76) by a factor $N_t \sim [(M_{\odot}/M_p)(H/R)^3](R/H)$. From the discussion in section 4.2, the first factor should exceed unity for an embedded planet. Thus, such an object is inevitably subjected to large torque fluctuations. The strength of the perturbation of the planet on the disc is measured by $(M_p/M_{\odot})(R/H)^3$. This perturbation only needs to produce a bias in the turbulent fluctuations resulting in a non-zero mean, corresponding to the typical fluctuation reduced by a factor N_t , for it to become comparable to the laminar type I value.

Also, global magnetic fields may be present and have significant effects on migration (Terquem 2003) or the disc may have long lived global distortions corresponding to non-zero eccentricity (Papaloizou 2002) which may also significantly affect migration. The migration

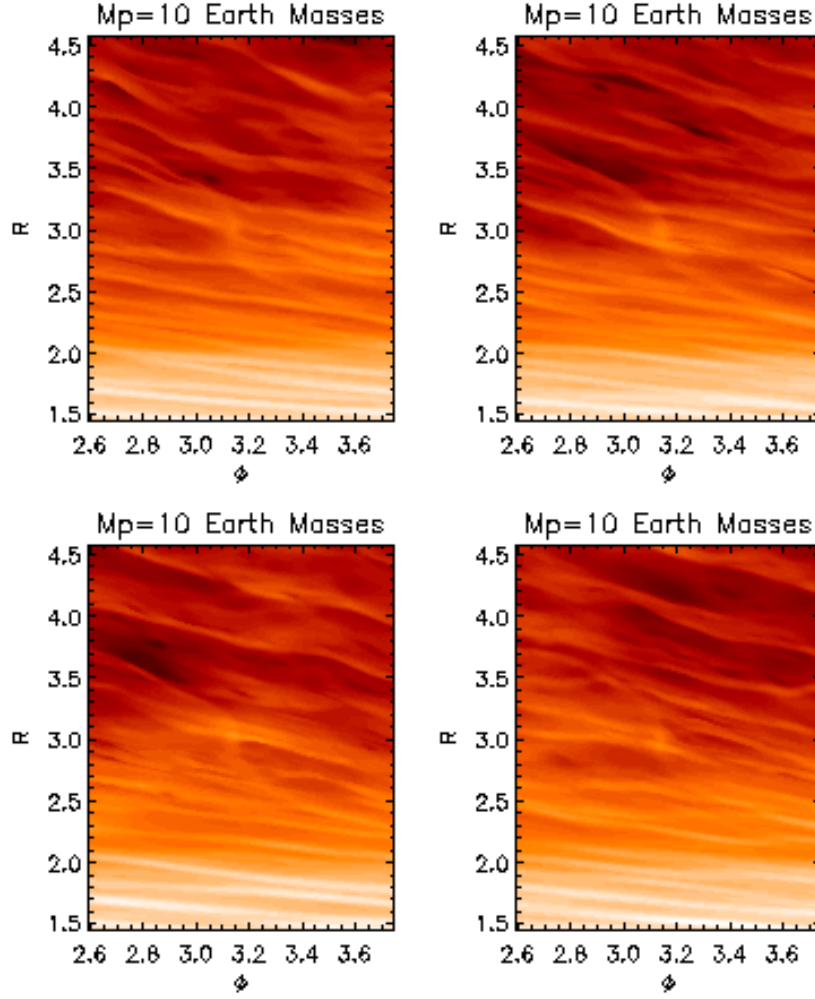


Figure 12. This figure shows midplane density contours in the region close to the protoplanet. The protoplanet is located at $(R, \phi) = (3, \pi)$ in a disc with fully developed MHD turbulence with $H/R = 0.07$ and no net magnetic flux. The panels correspond to times 429.0687 (*top left*), 496.8698 (*top right*), 542.0710 (*bottom left*) and 634.4136 (*bottom right*). The orbital period at $R = 1$ is 2π in these units. Note that the presence of the protoplanet of mass $10 M_{\oplus}$ is just detectable, although the perturbation it makes to the disc is of lower amplitude than the perturbations generated by the turbulence. (These plots are taken from Nelson and Papaloizou (2004).)

rates may also be affected by opacity transitions in the disk (Menou and Goodman 2004). Under some circumstances, significant fluctuations that can affect migration may occur in the non-magnetic case for planets in the Saturn mass range (Koller *et al* 2003) through the production of instabilities in inviscid gap edges.

4.2.2. Type I migration in a turbulent disc. To illustrate stochastic migration when fully developed MHD turbulence is present in a disc with $H/R = 0.07$ and no net magnetic flux, we show midplane density contours in figure 12, taken from Nelson and Papaloizou (2004). The presence of the protoplanet of mass $10 M_{\oplus}$ is just detectable, although the perturbation it makes to the disc is of lower amplitude than the perturbations generated by the turbulence.

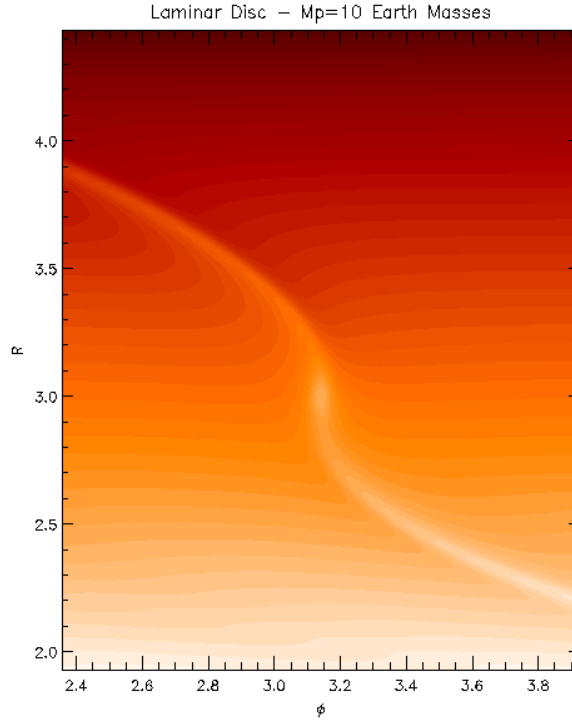


Figure 13. This figure shows midplane density contours for a laminar disc run with disc and protoplanet parameters otherwise identical to the simulation shown in figure 12. (This plot is taken from Nelson and Papaloizou (2004).)

Midplane density contours for the same protoplanet in a laminar disc with identical parameters is given in figure 13. The disc response to the protoplanet is clearly much better defined in this case.

The behaviour of the torque per unit mass exerted by the disc on the protoplanet in the laminar case is illustrated in the left panel of figure 14, taken from Nelson and Papaloizou (2004). The upper line shows the torque due to the inner disc, the lower line shows the torque due to the outer disc and the middle line shows the total torque. A well defined torque is produced, with an associated inward migration timescale in agreement with the standard linear theory.

By way of contrast, the right panel of figure 14 shows the running time average of the torque per unit mass exerted by the disc in the turbulent case. The upper line is the running time average of the torque acting on the planet due to the inner disc. The lower line is that due to the outer disc. The middle (not straight) line is the running time average of the total torque. The straight line is the total torque exerted on the protoplanet in the laminar disc run. The total time averaged torque does not converge to a well defined value over the simulation run time of 20 orbits.

These simulations indicate that, if the disc is not strictly laminar, type I migration may be significantly modified to the extent that the survival of protoplanetary cores may occur (this is also the case if the disc is magnetic or/and with gas not strictly on circular orbits). Furthermore, the enhanced mobility throughout the disc of a core undergoing stochastic migration may act to counteract the effects of core isolation (Rice and Armitage 2003). However, considerable

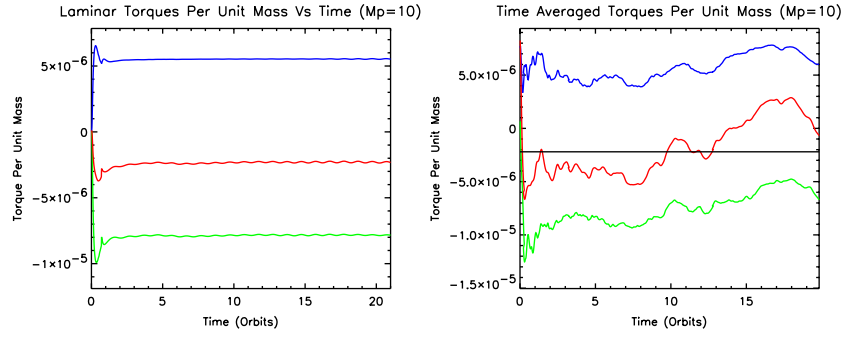


Figure 14. *Left panel:* torque per unit mass exerted by the disc on the protoplanet for the laminar disc simulation shown in figure 13. The upper line shows the torque due to the inner disc, the lower line shows the torque due to the outer disc and the middle line shows the total torque. It is clear that a well defined torque is produced, with an associated migration timescale. *Right panel:* running time average of the torque per unit mass exerted by the disc for the simulation shown in figure 12. The upper line is the running time average of the torque acting on the planet due to the inner disc. The lower line is that due to the outer disc. The middle (not straight) line is the running time average of the total torque. The straight line is the total torque exerted on the protoplanet in the comparable laminar disc run. We note that the total time averaged torque does not converge to a well defined value. (These plots are taken from Nelson and Papaloizou (2004).)

theoretical developments need to be made in order to confidently assess the outcome of these phenomena.

4.3. Coorbital torques and type III migration

By coorbital torque, we mean a torque exerted on the protoplanet by disc material flowing through the orbit. This has been discussed recently by Masset and Papaloizou (2003) and Artymowicz (2004) in a global context and by Papaloizou (2005) in a local shearing box context.

One starts by considering a migrating protoplanet. When gap formation is only partial, material will flow through the orbit. As it does so, a torque will be exerted on the protoplanet. This may either act to assist the migration (positive feedback) or retard it (negative feedback).

D'Angelo *et al* (2005) have recently studied coorbital torques acting on a Saturn mass planet in a massive disc under a similar set-up to Masset and Papaloizou (2003). They used a multigrid system for which the resolution could be increased close to the planet, *but not elsewhere*. They report only slow migration differing only marginally from type II migration. This was claimed to result from important torques acting within the Roche lobe. We firstly comment that their computational set-up allowed mass flow into the Roche lobe, effectively increasing the protoplanet mass by more than a factor of three. However, the increase in mass was not incorporated in the action of the protoplanet on the disc to determine its response, making the calculation inconsistent. Furthermore the results, checked with the highest resolution, shown in figure 14, which differ from those used to draw final conclusions, indicate a fast inward migration time of ~ 600 orbits. This would not indicate important contributions from inside the Roche lobe as one would expect when there is not an increasing accumulation of mass there (see the [appendix](#)). However, the study of the migration of partially gap forming planets of modest mass, with a disc mass comparable to their own in the coorbital region, is difficult numerically and clearly further work on the problem is required.

Masset and Papaloizou (2003) argue for positive feedback from coorbital torques because, for an outwardly migrating protoplanet, material traverses the coorbital zone from outside to inside. As it does so, it moves along the outer boundary of the coorbital zone occupied by material librating around the coorbital equilibrium (Lagrange) points passing close to the protoplanet. As this passage occurs, angular momentum is transferred to the protoplanet, a process which acts to assist the migration of the protoplanet and which accordingly gives a positive feedback. We now look at this process in more detail for an outwardly migrating protoplanet. Similar considerations apply with appropriate sign reversals to an inwardly migrating protoplanet.

The expected force exerted on the protoplanet by material flowing through the coorbital region is estimated as follows. The rate at which the mass flows through is $2\pi R\Sigma \, dR/dt$. The outward momentum per unit mass imparted to the protoplanet when the disc matter moves across the coorbital region of width w is $w\omega/2$, where ω is the rotation rate of the protoplanet. We adopt $w = 2H$, being appropriate for marginal gap formation. Then, the rate of transfer of momentum to the planet is $F_{\text{cr}} = 2\pi R\omega H\Sigma (dR/dt)$. This can be expressed as

$$F_{\text{cr}} = 2\pi R\Sigma\omega^2 H^2 \frac{1}{c_s} \frac{dR}{dt}, \quad (79)$$

where c_s is the sound speed. Using $c_s \simeq \omega H$, it may also be written in the form

$$F_{\text{cr}} = \frac{1}{2} M_d \omega \frac{dR}{dt}, \quad (80)$$

where $M_d = 4\pi R\Sigma H$ is the disc mass that would fill the coorbital zone of width $2H$ were it to do so at the background surface density.

However, torques on the protoplanet do not only arise from material passing through the coorbital zone. Material that is forced to comove with the protoplanet, either because it has been accreted by it or because it librates about coorbital equilibrium points, has to be acted on by the same force per unit mass as the protoplanet so as to maintain its migration. This results in an additional force acting on the protoplanet given by

$$F_{\text{crb}} = -\frac{1}{2} M_b \omega \frac{dR}{dt}, \quad (81)$$

where M_b is the coorbital bound mass. The total force so far is thus

$$F_{\text{cr}} = \frac{1}{2} (M_d - M_b) \omega \frac{dR}{dt}. \quad (82)$$

However, there are other forces. The forces acting due to density waves from the two sides will be affected by the migration and flow through and thus differ from the non-migrating case. In this context, note that there is an associated asymmetry in the surface density profile. Thus, we should expect a wave torque component which is proportional to the migration speed. The indication from our simulations is that this acts as a drag on the protoplanet, as does M_b . Accordingly, we shall consider the effects of asymmetric wave torques as modifying the value of M_b and continue to use equation (82).

Suppose now that the protoplanet is acted on by some external torque T_{ext} . The equation of motion governing the migration of the protoplanet of mass M_p with speed dR/dt , obtained by considering the conservation of angular momentum, is

$$\frac{1}{2} M_p R \omega \frac{dR}{dt} = \frac{1}{2} (M_d - M_b) R \omega \frac{dR}{dt} + T_{\text{ext}}. \quad (83)$$

Accordingly, we can consider the planet to move with an effective mass:

$$M_{\text{eff}} = M_p - (M_d - M_b). \quad (84)$$

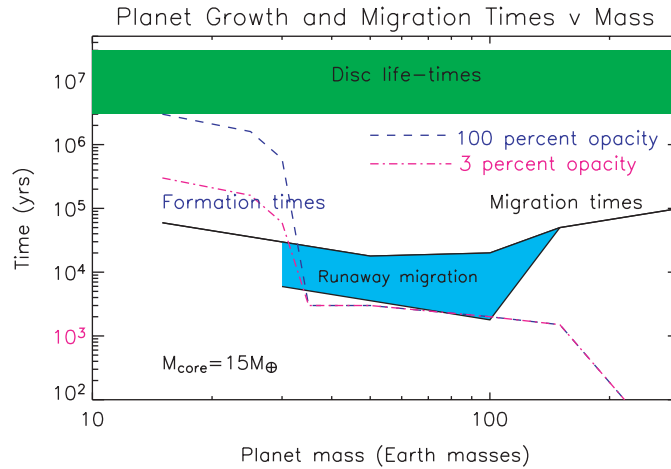


Figure 15. This diagram shows the migration time estimated for a minimum mass solar nebula (—) and the mass growth time for protoplanet models with $15 M_{\oplus}$ cores. The lowest runaway migration times, forming the lower boundary to the lower shaded region, are estimated for a model five times more massive than the minimum mass solar nebula. The dashed line shows the protoplanet growth time as a function of mass for models with standard opacity. The dashed-dotted line shows the growth time for models with 3 per cent opacity. (This plot is taken from Papaloizou and Nelson (2005).)

The quantity $(M_d - M_b)$ has been called the coorbital mass deficit by Masset and Papaloizou (2003). When this is positive, there is an effective reduction in the inertia of the protoplanet. If there were no asymmetry in wave torques, the coorbital mass deficit would be the amount of mass evacuated from the gap region were it to be initially filled with the background density. Gap filling accordingly reduces the coorbital mass deficit. It is also clear that, because at least a partial gap is required, the protoplanet must be massive enough to produce a nonlinear response in the disc.

Masset and Papaloizou (2003) and Papaloizou (2005) indeed found the coorbital mass deficit to be positive, resulting in positive feedback from coorbital torques. They also found that, in some circumstances, the coorbital mass deficit could become as large as the planet mass, so reducing the effective inertia to zero. Under these circumstances, a very fast or runaway migration may occur. Because the coorbital mass deficit is proportional to the disc surface density, the phenomenon requires a massive disc, typically an order of magnitude larger than the minimum mass solar nebula.

An interesting feature of this type of migration is that it is most effective for protoplanets that bare marginally gap forming, and thus is in the mass range $0.1\text{--}0.3 M_J$. Such forming protoplanets may undergo a period of fast migration that brings them into the inner regions of the disc, where gap formation is easier, while they are still of sub-Jovian mass. It may thus form an explanation of why the ‘hot Jupiters’ are in general sub-Jovian in mass (see Masset and Papaloizou (2003) and Papaloizou and Nelson (2005)). But note in this context that this explanation does not necessarily require type III migration, as migration rates tend to be maximal in the mass range $0.1\text{--}0.3 M_J$ for low-mass discs too (see D’Angelo *et al* (2003)) while lower masses may have migration inhibited by turbulence (Nelson and Papaloizou 2004).

A graphical summary of the different types of migration is shown in figure 15. This shows the migration timescale as a function of protoplanet mass assuming a core of a constant

mass of $15 M_{\oplus}$. It allows for the possibility of type I, type II and type III migration. It indicates that in order for a planet to form, type I migration, which produces radial infall timescales much shorter than the protoplanet growth time, has to be strongly suppressed relative to the prediction derived from a non-magnetic, laminar and circular disc. However, beyond type I migration, survivability, at least to enable the formation of the planet, is not threatened on account of the shortened growth times. This is the case even if type III migration operates.

5. Multiplanet systems and their interactions

The presence of high orbital eccentricities amongst extrasolar planets is suggestive of a strong orbital relaxation or scattering process. For this to happen, formation must occur on a timescale short enough that strong dynamical interactions may take place subsequently. The gaseous environment of a disc may act to inhibit such interactions until it is removed. Gas free dynamical interactions of coplanar protoplanets formed on neighbouring circular orbits have been considered by Weidenschilling and Marzari (1996) and Rasio and Ford (1996). These may produce close scatterings and high eccentricities, but the observed distribution of extrasolar planets is not reproduced.

5.1. Orbital relaxation in young planetary systems

Papaloizou and Terquem (2001) (see also Adams and Laughlin (2003)) investigate a scenario in which $5 \leq N \leq 100$ planetary objects in the range of several Jupiter masses are assumed to form rapidly through fragmentation or gravitational instability occurring in a disc or protostellar envelope on a scale of $R_{\max} = 100$ AU. If these objects are put down in circular orbits about a solar mass star, at random in a volume contained within a spherical shell with inner and outer radii of $0.1 R_{\max}$ and R_{\max} , respectively, a strong relaxation on a timescale ~ 100 orbits occurs, leading to independence of details of initial conditions.

For the range of values of N considered, the evolution is similar to that of a star cluster. Most objects escape, leaving at most 3 bound planets, the innermost with semi-major axis in the range $0.01 R_{\max} - 0.1 R_{\max}$. However, close encounters or collisions with the central star occur for about 10% of the cases. Tidal interaction, giving orbital circularization at fixed pericenter distance and leading to the formation of a very closely orbiting giant planet, is then a possibility.

An example of a run with $N = 8$ masses selected uniformly at random in the interval $(0, 5 \times 10^{-3} M_{\odot})$ and with central stellar radius $R_{\star} = 1.337 \times 10^{-4} R_{\max}$ is illustrated in figure 16. At the end of this run, only 2 planets remain bound to the central star.

The relaxation processes discussed above are more likely to apply to the more massive extrasolar planets exceeding $1 M_J$. Their observed number increases with distance (Zucker and Mazeh (2002) and section 2.2.1 above), as found in the simulations. So far, there are 23 isolated candidate objects with mass $M_p \sin i > 4.5 M_J$, semi-major axes a in the range 0.3–3.9 AU and eccentricities e in the range 0.03–0.76. Amongst the ‘hot Jupiters’ detected so far, τ Boo is a possible candidate, being unusually massive with $M_p \sin i \sim 4 M_J$. Note that, in multiple systems, only 3 planets have been detected with a mass larger than $4.5 M_J$. They are actually significantly more massive than this limit, being between 6 and $13 M_J$. One of these systems contains a $7.7 M_J$ planet with $a = 0.29$ AU and $e = 0.529$ and a $16.9 M_J$ brown dwarf with $a = 2.85$ AU and $e = 0.228$. Such a system could well have formed through the process described above. The two other systems will be discussed below.

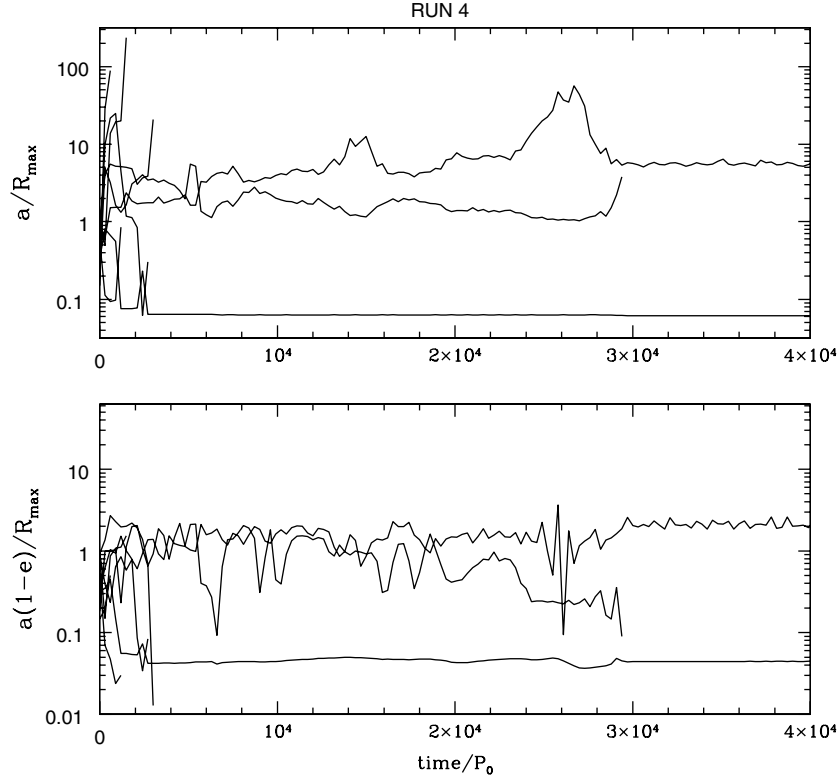


Figure 16. This figure shows the evolution of the semi-major axes (*upper plot*) and pericenter distances (*lower plot*) of the $N = 8$ planets in the system versus time (measured in units of P_0 , being the orbital period at R_{\max}). The lines correspond to the different planets. A line terminates just prior to the escape of a planet. (These plots are taken from Papaloizou and Terquem (2001).)

5.2. Effect of an outer relaxing distribution of protoplanets on a low-mass planet formed in an inner disc

Terquem and Papaloizou (2002) have considered the effects of an outer relaxing distribution of protoplanets on an inner disc in which one inner planet with a mass $m_p = 0.3 M_J$ is formed on a timescale of 10^6 yr. During the formation process, the eccentricity is assumed to be damped by tidal interaction with the disc while the planet is built up progressively with orbital radius a in the range 0.3–10 AU.

An example of a run with $N = 9$ outer planets of mass $8 M_J$ and $R_{\max} = 100$ AU is illustrated in figure 17. After $\sim 10^6$ yr, one outer planet with $a \sim 8$ AU and $e \sim 0.6$ remains. The inner planet enters into a cycle in which e varies between 0.1 and 0.24. The mutual inclination oscillates between 0° and 30° . More extreme cycles have been produced in other examples.

Thus, an outcome of an outer relaxing system could be an inner lower mass protoplanet with high orbital eccentricity. Among the observed planets, there are several candidates for which no companion has been detected so far. If the above scenario applies, there should be an outer massive planet with high eccentricity. Among the candidate systems with large radial velocity drifts potentially due to a companion selected by Fischer *et al* (2001) was HD 38529, which has a planet with $m_p \sin i = 0.76 M_J$, $a = 0.13$ AU and $e = 0.27$. This is similar to the system illustrated in figure 17. More recently, the discovery of a companion with $M_p \sin i = 12.7 M_J$, $a = 3.7$ AU and $e = 0.36$ has been announced (Fischer *et al* 2003).

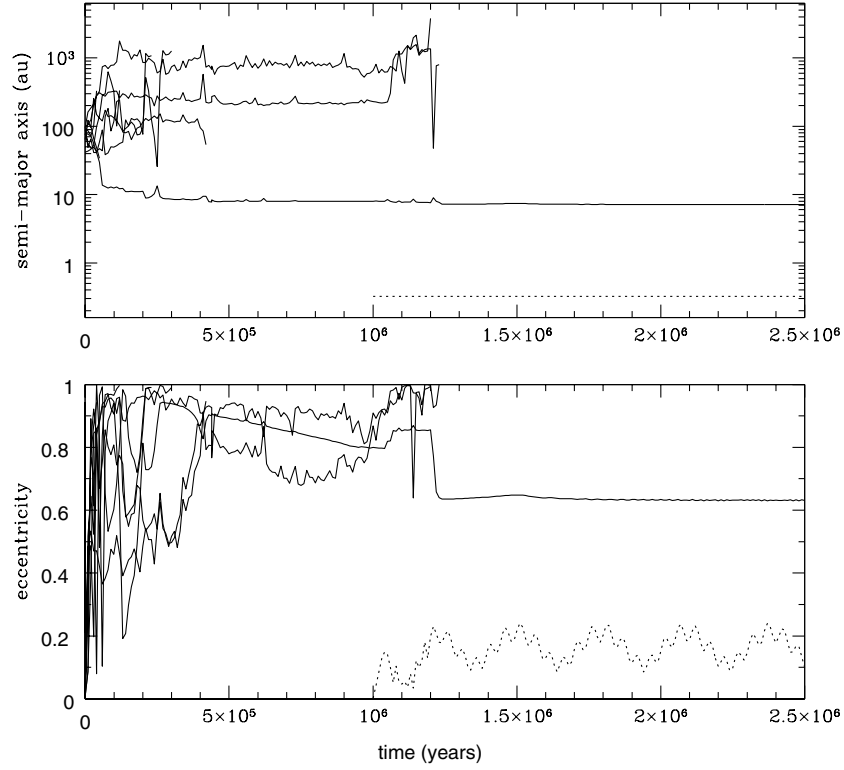


Figure 17. Semi-major axes (in AU, *upper plot*) and eccentricities (*lower plot*) of the $N = 9$ outer planets (—) and low-mass planet (·····) versus time (in year). Here, for the innermost low-mass planet, $a = 0.3$ AU initially. The eccentricity of the low-mass planet varies between 0.1 and 0.24 while a is almost constant. (These plots are taken from Terquem and Papaloizou (2002).)

If $\sin i = 0.8$ is adopted for this system, a significant eccentricity is not excited in an initially circular inner planet orbit if the system is assumed coplanar. But, as illustrated in figure 18, for a mutual orbital inclination of 60° , a cycle is found in which the inner eccentricity reaches its observed value.

As mentioned above, three of the observed multiple systems have a planet with a mass larger than $4.5 M_J$. In two of these systems (HD 74156 and HD 38529), the massive planet has a semi-major axis on the order of 3.5 AU and a significant eccentricity. The companion has a lower mass, a semi-major axis in the range 0.1–0.3 AU and also a significant eccentricity. These systems may be explained by the scenario presented here.

6. Concluding remarks

The recent observations of extrasolar planets strongly suggest that gravitational interactions between forming planets and the nebula in which they are embedded form a key element in the understanding of the configurations of the systems. Given that the hot Jupiters have almost certainly not formed where they are observed today, they must have been subject to some orbital migration resulting from planet/disc tidal interaction. This type of interaction also leads to commensurable pairs of planets and apsidal resonances, features that are also observed in some extrasolar systems.

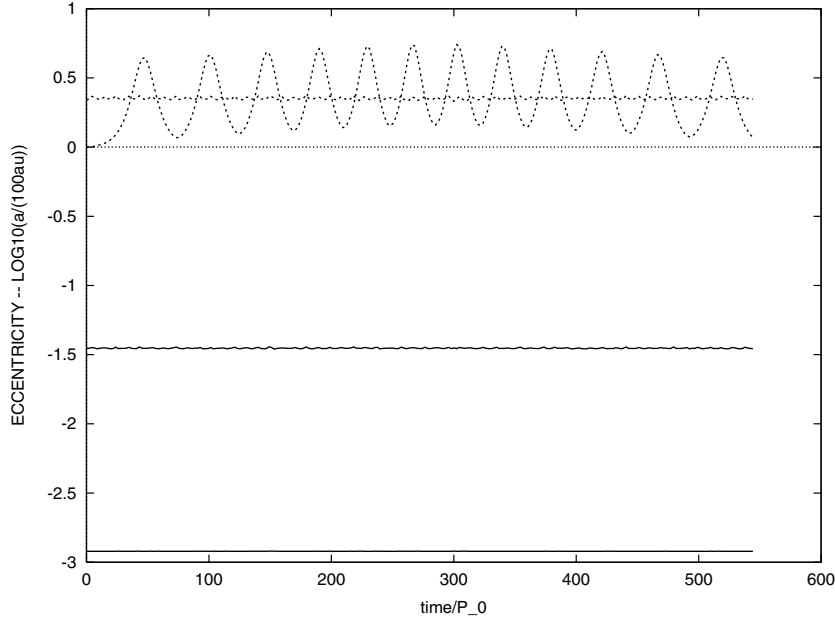


Figure 18. Motion in HD 35829 for mutual inclinations of 0° and 60° . The unchanging $\log_{10}(a/R_{\max})$ for the two planets are given by the lower curves. The eccentricities by the upper curves. The oscillatory curve indicates a Kozai like cycle for the inner planet in the high inclination case. In the low inclination case the eccentricity of the inner planet remains close to zero.

A substantial amount of work aimed at gaining a better understanding of planet/disc tidal interactions has been done in the last ten years. Three-dimensional numerical simulations incorporating internally generated MHD turbulence rather than anomalous viscosity prescriptions have begun. They show that orbital migration is strongly affected by such processes, especially when the mass is in the Earth mass range. Analytical and numerical work has also shown that the presence of a large scale magnetic field or a finite orbital eccentricity can have a significant effect on migration.

Strong gravitational interactions amongst a population of distant giant planets formed early in the life of the protostellar disc through gravitational fragmentation may also produce close orbiters and planets on highly eccentric orbits. The latter may belong to the original distribution or be formed by a later accumulation in an inner disc.

We are currently at a stage where the important processes (gravitational instability, dust sedimentation, core accumulation, gas accretion and disc/planet interaction) can be studied in detail individually through numerical simulations. However, it is still not possible to take into account all these processes acting simultaneously in a system with many embryos, to try to model the evolution of planetary systems from birth to final state. This would require very complex global simulations that cannot yet be performed. Also, for planetary systems, the evolutionary outcome, in the form of planetary masses and orbital configuration, depends very much on the initial conditions. The amount of mass present in the disc when planets begin to form is important in relation to gravitational instabilities, solid core masses that can be accumulated and migration rates. Improved understanding will come as more powerful computers become available and observations put more constraints on realizable initial conditions.

Acknowledgments

This work was supported in part by the European Community's Research Training Networks Programme under Contract HPRN-CT-2002-00308, 'PLANETS'. CT acknowledges partial support from the Institut Universitaire de France.

Appendix A. Disc-protoplanet torques and angular momentum conservation

We here consider a protoplanet in either a fixed circular orbit or migrating in a circular orbit with slowly varying radius, that is embedded in and interacting with a gaseous disc. We show that under very general assumptions, the torque exerted between protoplanet and disc can be expressed as a combination of advective wave and viscous fluxes evaluated far from the protoplanet, together with a contribution accounting for the rate of change of angular momentum of the disc domain in the neighbourhood of the protoplanet when conditions are non-stationary. In this way, the contributions relevant to different types of migration can be identified.

In order to proceed, we need to consider the problem set-up and distinguish between disc and protoplanet. It has been a common practice to treat the protoplanet as providing a forcing of the disc with a softened point mass potential. In that case, one may regard all the gas as belonging to the disc and identify the protoplanet mass as that associated with the potential, but only as long as the gas mass within a softening lengthscale of the protoplanet centre is negligible compared with that of the protoplanet. Another possibility would be to regard the protoplanet as residing interior to some domain within its Roche lobe. In that case, we consider the outer disc material to be unable to enter this domain or transmit angular momentum interior to it. Thus, the protoplanet itself does not change its mass or angular momentum through interaction with disc material at its boundary. But note that when this domain is well inside the Roche lobe, mass and angular momentum can still accumulate inside the Roche lobe but exterior to the protoplanet.

Adopting cylindrical coordinates (r, φ, z) with origin at the central star of mass M_\star , we consider a protoplanet orbiting in a disc with uniform angular velocity $\Omega \hat{\mathbf{k}}$, where $\hat{\mathbf{k}}$ is the unit vector in the vertical z direction, relative to an inertial frame.

The basic equations governing the disc-protoplanet interaction are the conservation of mass and momentum in the form:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{v}), \quad (85)$$

and

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} + 2\Omega \hat{\mathbf{k}} \times \mathbf{v} = -\frac{\nabla P}{\rho} - \mathbf{f} - \nabla \psi_0 - \nabla \psi_1, \quad (86)$$

respectively. Here, ρ and $\mathbf{v} = (v_r, v_\varphi, v_z)$ are the disc density and velocity, respectively. The pressure is P , $\psi_0 = -GM_\star/\sqrt{r^2 + z^2} - \Omega^2 r^2/2$ is the gravitational potential due to the central object together with the centrifugal potential and ψ_1 is the gravitational potential due to the orbiting protoplanet.

The force per unit volume $\rho \mathbf{f} = \rho(f_r, f_\varphi, f_z)$ is taken to be the divergence of a stress tensor, t_{ij} . Thus, in Cartesian coordinates $\mathbf{r} \equiv (x_1, x_2, x_3)$ for which $\mathbf{f} = (f_1, f_2, f_3)$, $\rho f_i = \partial t_{ij}/\partial x_j$, where we have used the Einstein summation convention. Different effects may contribute to t_{ij} , including Navier Stokes viscosity in a laminar disc, self-gravity in a massive disc or Maxwell stresses in a magnetized disc.

Appendix A.1. Conservation of angular momentum

We may derive the relationship between the total torque acting between disc and protoplanet and wave and advected angular momentum fluxes in the disc by considering global angular momentum conservation. To do this, we start from the local conservation expressed by the azimuthal component of the equation of motion (86) in the form

$$\frac{\partial(r v_\varphi)}{\partial t} + \mathbf{v} \cdot \nabla(r v_\varphi) + 2\Omega r v_r = -\frac{1}{\rho} \frac{\partial P}{\partial \varphi} - r f_\varphi - \frac{\partial \psi_1}{\partial \varphi}. \quad (87)$$

On multiplying by ρr and integrating over a fixed domain \mathcal{H} , including the protoplanet but with a non-interacting core cut out if needed, we obtain a statement of the total angular momentum conservation which, for an assumed steady state, takes the form

$$T = \Delta(F_{Re} + F_{str}). \quad (88)$$

Here, T is the total torque acting between protoplanet and disc:

$$T = - \int_{\mathcal{H}} \rho \frac{\partial \psi_1}{\partial \varphi} r \, dz \, d\varphi \, dr. \quad (89)$$

$\Delta(F_{Re} + F_{str})$ denotes the difference of the total flux $F_{Re} + F_{str}$ (see below) evaluated on cylindrical bounding surfaces exterior to and interior to the protoplanet orbit, respectively.

F_{Re} is the angular momentum flux associated with Reynolds stresses:

$$F_{Re} = \int_{\Sigma} \rho v_r (v_\varphi + r\Omega) r^2 \, dz \, d\varphi, \quad (90)$$

with the integral being taken over a cylindrical bounding surface, Σ , assumed distant from the protoplanet. F_{str} is the contribution from the stress term:

$$F_{str} = - \int_{\Sigma} t_{r\varphi} r^2 \, dz \, d\varphi. \quad (91)$$

The mass flow through a cylindrical surface Σ' , normal to a particular radius r in the disc, is given by

$$\dot{M} = \int_{\Sigma'} \rho v_r r \, dz \, d\varphi. \quad (92)$$

Using this in equation (90), we obtain:

$$F_{Re} = \int_{\Sigma} \rho v_r (v_\varphi - \bar{v}_\varphi) r^2 \, dz \, d\varphi + \dot{M} (\bar{v}_\varphi + r\Omega) r. \quad (93)$$

Here the overline denotes azimuthal and vertical averaging with weight ρ . The first term can be identified as a wave or viscous flux (see, e.g. Papaloizou and Lin (1984)), as can F_{str} . The second term may also be written as $\dot{M} j$, with j being the averaged specific angular momentum as viewed from the inertial frame. We shall assume the boundaries are far enough away from the protoplanet such that j may be approximated as the Keplerian value. Note that, in a steady state, \dot{M} is a constant. Further, turbulent or fluctuating states may be dealt with through appropriate time averaging, and we assume this to be carried out where needed. Equation (88) then gives the torque exerted on the disc by the protoplanet as

$$T = \dot{M} \Delta(j) + \Delta(F_{Re} + F_{str}). \quad (94)$$

In this case, the torque can be expressed as the difference between boundary wave and viscous flux terms together with the difference between direct boundary output and input associated with the constant mass flux. Note that, when the protoplanet mass is reduced to zero, $T = 0$ and we have $\dot{M} \Delta(j) + \Delta(F_{str0}) = 0$. Here F_{str0} is the boundary viscous stress in the limit

of vanishing protoplanet mass. This is just a statement that the advected angular momentum flow in a standard steady state disc is balanced by torques exerted through viscous, magnetic or gravitational stresses. Thus, when a protoplanet is present, we have

$$T = \Delta(F_{Re} + F_{str} - F_{str0}). \quad (95)$$

We see that, in this case, the torque exerted by the protoplanet on the disc can be expressed entirely in terms of boundary flux terms.

Appendix A.2. A migrating protoplanet

To consider a migrating protoplanet, we suppose now that the angular velocity Ω is a function of time, as is the orbital radius of the protoplanet $R(t)$. We adopt a coordinate system such that the protoplanet appears stationary. Thus, this rotates with the time dependent angular velocity Ω , while the other coordinate directions scale with the orbital radius $R(t)$. When conditions are quasi-steady, this corresponds to a self-similar solution.

We define new dimensionless coordinates $x'_i = x_i/R(t)$. We also adopt a scaled dimensionless time coordinate τ , which is such that $d\tau = \Omega dt$.

Further, we adopt new dimensionless velocity components given by

$$\mathbf{v}' = \frac{\mathbf{v}}{R\Omega} - \frac{\mathbf{r}}{R^2\Omega} \frac{dR}{dt}. \quad (96)$$

In terms of the new variables, the continuity equation can be written as

$$\frac{1}{\rho_3} \left(\frac{\partial \rho_3}{\partial \tau} + \mathbf{v}' \cdot \nabla' \rho_3 \right) = \frac{1}{\rho_3} \frac{D\rho_3}{D\tau} = -\nabla' \cdot (\mathbf{v}'), \quad (97)$$

where $\rho_3 = \rho R^3/R_0^3$, where R_0 is a constant with the dimensions of length that can be taken to be the value of R at some time $t = t_0$ that can be chosen for convenience and the prime in ∇' denotes differentiation with respect to x'_i .

Similarly, the equation of motion becomes

$$\begin{aligned} \frac{1}{R^2\Omega} \frac{D}{D\tau} (R^2\Omega \mathbf{v}') + 2(\hat{\mathbf{k}} \times \mathbf{v}') + \frac{(\hat{\mathbf{k}} \times \hat{\mathbf{r}})r'}{\Omega^2 R^2} \frac{d(R^2\Omega)}{dR} \frac{dR}{dt} + \frac{\mathbf{r}'}{\Omega^2 R} \frac{d^2 R}{dt^2} \\ = -\frac{1}{R^2\Omega^2} \left(\nabla'(\psi_0 + \psi_1) + \frac{1}{\rho} \nabla' P + \mathbf{f}R \right). \end{aligned} \quad (98)$$

Note that we assume here that Ω is a function of R . In fact, we shall use the Keplerian form $\Omega^2 = GM_*/R^3$. In addition, we shall adopt $dR/dt = \epsilon R\Omega$, where ϵ is a constant of small magnitude corresponding to a slow radial migration. In that case, when the pressure and \mathbf{f} can be scaled appropriately, it is possible to seek a steady solution of equations (97) and (98) for the primed state variables that is independent of τ . When the equation of state is $P = \rho c_s^2$, with c_s being the sound speed, assumed to be $\propto r^{-1/2}$, the stress tensor corresponding to \mathbf{f} scales as P , and the gravitational softening length scales as R if it is being used; R may be scaled out of the basic equations thus enabling a steady state to be sought.

We comment that the last two terms on the left-hand side of equation (98) arise from the transformation to an expanding/contracting rotating frame. In this frame, they appear as the additional forces/torques required to induce the apparent modified disc flow.

Appendix A.3. Conservation of angular momentum

As in the fixed orbit case, we formulate the global conservation of angular momentum derived from equation (98). On multiplying by $\rho_3 r'$ and integrating over a fixed domain \mathcal{D} , which is

similar to \mathcal{H} but is defined in the primed coordinates, the azimuthal component of equation (98) gives the total angular momentum conservation in the form

$$\Omega \frac{dJ}{d\tau} + \Delta(F_{Re} + F_{str}) = T. \quad (99)$$

Here the total torque acting between protoplanet and disc is

$$T = -R_0^3 \int_{\mathcal{D}} \rho_3 \frac{\partial \psi_1}{\partial \varphi'} r' dz' d\varphi' dr'. \quad (100)$$

The angular momentum content within \mathcal{D} , as seen in the inertial frame, is

$$J = R^2 \Omega R_0^3 \int_{\mathcal{D}} \rho_3 (v'_\varphi + r') r'^2 dr' dz' d\varphi', \quad (101)$$

while the angular momentum flux associated with Reynolds stresses is

$$F_{Re} = R^2 \Omega^2 R_0^3 \int_S \rho_3 v'_r (v'_\varphi + r') r'^2 dz' d\varphi', \quad (102)$$

with the integral being taken over a bounding surface S , assumed distant from the protoplanet and fixed in the primed coordinates. The contribution from the stress term is

$$F_{str} = -R^3 \int_S t_{r',\varphi'} r'^2 dz' d\varphi'. \quad (103)$$

As before, the mass flow through a cylindrical surface normal to a particular radius, r' , in the disc which remains fixed in the primed coordinates, is given by

$$\dot{M} = \Omega R_0^3 \int_S \rho_3 v'_r r' dz' d\varphi'. \quad (104)$$

Using this in equation (102), we obtain

$$F_{Re} = R^2 \Omega^2 R_0^3 \int_S \rho_3 v'_r (v'_\varphi - \bar{v}'_\varphi) r'^2 dz' d\varphi' + \dot{M} R^2 \Omega (\bar{v}'_\varphi + r') r'. \quad (105)$$

Here the overline denotes azimuthal and vertical averaging with weight ρ_3 . The first term can be identified as a wave or viscous flux, as can F_{str} . The second term may also be written as $\dot{M} j$, with j being the averaged specific angular momentum. Again we assume the boundaries to be sufficiently far from the protoplanet such that j may be approximated as Keplerian. In general, \dot{M} is a function of both r' and τ or time. However, in a putative steady state, as seen in the corotating comoving coordinate system (primed variables independent of τ), it is a function of time only.

Using the above, equation (99) gives the torque exerted on the disc by the protoplanet as

$$T = \Omega \frac{dJ}{d\tau} + \Delta(\dot{M} j) + \Delta F_{wv}, \quad (106)$$

where we have combined the viscous and wave fluxes in F_{wv} . Thus the torque can be expressed as the difference between boundary wave and viscous flux terms together with the difference between direct output and input associated with the mass throughput in combination with the rate of growth of the angular momentum contained in the domain \mathcal{D} . It is the latter two terms that are responsible for the coorbital torques. When the angular momentum change associated with mass throughput is not accounted for by an increase or decrease in the angular momentum content of the domain traversed, it is accounted for by the protoplanet disc torque, which means that there has to be an angular momentum exchange with the protoplanet.

Let us look at this more closely. Although conclusions do not depend on this, for simplicity we consider an assumed steady state self-similar case, then \dot{M} does not vary

in space. But note that, from equation (101), because of the dependence on R and Ω , $\Omega dJ/d\tau = J/(2R)(dR/dt) \neq 0$. In this case we have

$$T = \frac{J}{2R} \frac{dR}{dt} + \dot{M}(\Delta j) + \Delta F_{wv}. \quad (107)$$

Suppose now viscous effects are small, so that \dot{M} as viewed from the comoving coordinate frame which we use is due to the flow of matter induced by the migration. When the mass of the protoplanet is negligible, but the migration maintained, we expect that, as then there are no excited waves and $T = 0$:

$$\frac{J_0}{2R} \frac{dR}{dt} + \dot{M}(\Delta j) = 0. \quad (108)$$

Here J_0 is the angular momentum content of \mathcal{D} in this limit. Using this in equation (108) gives

$$T = \frac{(J - J_0)}{2R} \frac{dR}{dt} + \Delta F_{wv}. \quad (109)$$

Now, conservation of angular momentum of the planetary orbit gives

$$T = -\frac{J_p}{2R} \frac{dR}{dt}, \quad (110)$$

where J_p is the angular momentum of the protoplanet. Thus we have

$$\frac{(J_p - (J_0 - J))}{2R} \frac{dR}{dt} = -\Delta F_{wv}. \quad (111)$$

Thus, as far as the orbital evolution is concerned, the protoplanet orbital angular momentum is reduced by the *coorbital angular momentum deficit* $J_0 - J$. Other things being equal, this is expected to be positive if the protoplanet generates a partial gap, and in that case, once the effect of any asymmetric wave torques is taken into account, is clearly related to the *coorbital mass deficit* considered in section 4.3. There, it leads to a positive feedback enhancing the migration. On the other hand, if $J_0 - J$ is negative on account of excess material accreted in the vicinity of the planet increasing the angular momentum content interior to the Roche lobe, then the reverse effect of negative feedback occurs.

Note further that, if the structure interior to the Roche lobe is quasi-steady and if the total mass of material there, not counted as part of the protoplanet, is small compared with that of the protoplanet, only a small contribution to the disc–protoplanet torque is expected from these regions.

References

- Adams F C and Laughlin G 2003 *Icarus* **163** 290
 Alibert Y, Mordasini C and Benz W 2004 *Astron. Astrophys.* **417** L25
 Alonso R, Brown T M, Torres G *et al* 2004 *Astrophys. J.* **613** L153
 Artymowicz P 1992 *Publ. Astron. Soc. Pac.* **679** 769
 Artymowicz P 1993 *Astrophys. J.* **419** 155
 Artymowicz P 2004 *Astron. Soc. Pac.* **324** 39
 Backer D C, Foster R S and Sallmen S 1993 *Nature* **365** 817
 Balbus S A and Hawley J F 1991 *Astrophys. J.* **376** 214
 Balbus S A and Hawley J F 1998 *Rev. Mod. Phys.* **70** 1
 Barge P and Sommeria J 1995 *Astron. Astrophys.* **295** L1
 Bate M R, Lubow S H, Ogilvie G I and Miller K A 2003 *Mon. Not. R. Astron. Soc.* **341** 213
 Béjar V J S, Martín E L, Zapatero Osorio M R, Rebolo R, Barrado y Navascués D, Bailer-Jones C A L, Mundt R, Baraffe I, Chabrier C and Allard F 2001 *Astrophys. J.* **556** 830
 Bell K R and Lin D N C 1994 *Astrophys. J.* **427** 987

- Bennett D P and Rhie S H 1996 *Astrophys. J.* **472** 660
- Binney J and Tremaine S 1987 *Galactic Dynamics* (Princeton, NJ: Princeton University Press)
- Bodenheimer P, Hubickyj O and Lissauer J J 2000 *Icarus* **143** 2
- Bodenheimer P and Pollack J B 1986 *Icarus* **67** 391
- Bond I A, Udalski A, Jaroszyński M *et al* 2004 *Astrophys. J.* **606** L155
- Boss A P 1998 *Astrophys. J.* **503** 923
- Boss A P 2000 *Astrophys. J.* **536** L101
- Bryden G, Chen X, Lin D N C, Nelson R P and Papaloizou J C B 1999 *Astrophys. J.* **514** 344
- Burgasser A J, Kirkpatrick J D, McGovern M R, McLean I S, Prato L, Reid I N 2004 *Astrophys. J.* **604** 827
- Cameron A G W 1978 *Moon Planets* **18** 5
- Chabrier G, Saumon D, Hubbard W B and Lunine J I 1992 *Astrophys. J.* **391** 817
- Chambers J E, Wetherill G W and Boss A P 1996 *Icarus* **119** 261
- Chambers J E and Wetherill G W 1998 *Icarus* **136** 304
- Charbonneau D, Brown T M, Latham D W and Mayor M 2000 *Astrophys. J.* **529** L45
- Charbonneau D, Brown T M, Noyes R W and Gilliland R L 2002 *Astrophys. J.* **568** 377
- Charbonneau D *et al* 2005 *Astrophys. J.* **626** 523
- Cox J P and Giuli R T 1968 *Principles of Stellar Structure: Physical Principles* (New York: Gordon and Breach)
- Cuzzi J N, Dobrovolskis A R and Champney J M 1993 *Icarus* **106** 102
- Cuzzi J N, Hogan R C, Paque J M and Dobrovolskis A R 2001 *Astrophys. J.* **546** 496
- D'Angelo G, Bate M and Lubow S 2005 *Mon. Not. R. Astron. Soc.* **358** 316
- D'Angelo G, Kley W and Henning T 2003 *Astrophys. J.* **586** 540
- Deming D, Seager S, Richardson L J and Harrington J 2005 *Nature* **434** 740
- Durisen R H 2001 *Proc. IAU Symp. 200 The Formation of Binary Stars (Potsdam) (ASP Conf. Series vol 200)* ed H Zinnecker and R D Mathieu (San Francisco) p 381
- Durisen R H, Cai K, Mejía A C and Pickett M K 2005 *Icarus* **173** 417
- Eggenberger A, Udry S and Mayor M 2004 *Astron. Astrophys.* **417** 353
- Fischer D A, Marcy G W, Butler R P, Vogt S S, Frink S and Apps K 2001 *Astrophys. J.* **551** 1107
- Fischer D A, Marcy G W, Butler R P, Vogt S S, Henry G W, Pourbaix D, Walp B, Misch A A and Wright J T 2003 *Astrophys. J.* **586** 1394
- Gaudi B S *et al* 2002 *Astrophys. J.* **566** 463
- Gilliland R L *et al* 2000 *Astrophys. J.* **545** L47
- Goldreich P 1965 *Mon. Not. R. Astron. Soc.* **130** 159
- Goldreich P and Lynden-Bell D 1965 *Mon. Not. R. Astron. Soc.* **130** 125
- Goldreich P, Lithwick Y and Sari R 2004 *Annu. Rev. Astron. Astrophys.* **42** 549
- Goldreich P and Tremaine S 1979 *Astrophys. J.* **233** 857
- Goldreich P and Tremaine S 1980 *Astrophys. J.* **241** 425
- Goldreich P and Ward W R 1973 *Astrophys. J.* **183** 1051
- Gonzalez G 1997 *Mon. Not. R. Astron. Soc.* **285** 403
- Goodman J and Pindor B 2000 *Icarus* **148** 537
- Gould A and Loeb A 1992 *Astrophys. J.* **396** 104
- Halbwachs J L, Mayor M and Udry S 2005 *Astron. Astrophys.* **431** 1129
- Hayashi C 1981 *Prog. Theor. Phys. Suppl.* **70** 35
- Heemskerk M H M, Papaloizou J C and Savonije G J 1992 *Astron. Astrophys.* **260** 161
- Ida S and Makino J 1993 *Icarus* **106** 210
- Ikoma M, Nakazawa K and Emori H 2000 *Astrophys. J.* **537** 1013
- Ivanov P B, Papaloizou J C B and Polnarev A G 1999 *Mon. Not. R. Astron. Soc.* **307** 79
- Joshi K J and Rasio F A 1997 *Astrophys. J.* **479** 948
- Kenyon S J and Bromley B C 2004 *Astron. J.* **127** 513
- Klahr H and Henning T 1997 *Icarus* **128** 213
- Kley W 1999 *Mon. Not. R. Astron. Soc.* **303** 696
- Kley W, Peitz J and Bryden G 2004 *Astron. Astrophys.* **414** 735
- Kokubo E and Ida S 1998 *Icarus* **131** 171
- Koller J, Li H and Lin D N C 2003 *Astrophys. J.* **596** L91
- Korycansky D G and Pollack J B 1993 *Icarus* **102** 105
- Kuiper G P 1949 *Astrophys. J.* **109** 308
- Laughlin G and Adams F C 1997 *Astrophys. J.* **491** L51
- Laughlin G and Bodenheimer P 1994 *Astrophys. J.* **ApJ** **436** 335
- Laughlin G, Steinacker A and Adams F C 2004 *Astrophys. J.* **608** 489

- Lee M H and Peale S J 2002 *Astrophys. J.* **567** 59
- Lin D N C 1997 *Proc. IAU Colloq. 163: Accretion Phenomena and Related Outflows (ASP Conf. Series vol 121)* ed D T Wickramasinghe *et al* (San Francisco: ASP) p 321
- Lin D N C, Bodenheimer P and Richardson D C 1996 *Nature* **380** 606
- Lin D N C and Papaloizou J C B 1979 *Mon. Not. R. Astron. Soc.* **186** 799
- Lin D N C and Papaloizou J C B 1986 *Astrophys. J.* **309** 846
- Lin D N C and Papaloizou J C B 1993 *Protostars and Planets III* ed E H Levy and J I Lunine (Tucson: University of Arizona Press) p 749
- Lin D N C, Papaloizou J C B, Terquem C, Bryden G and Ida S 2000 *Protostars and Planets IV* ed V Mannings *et al* (Tucson: University of Arizona Press) p 1111
- Lissauer J J 1993 *Annu. Rev. Astron. Astrophys.* **31** 129
- Low C and Lynden-Bell D 1976 *Mon. Not. R. Astron. Soc.* **176** 367
- Lubow S H, Seibert M and Artymowicz P 1999 *Astrophys. J.* **526** 1001
- Lucas P W and Roche P F 2000 *Mon. Not. R. Astron. Soc.* **314** 858
- Lynden-Bell D and Pringle J E 1974 *Mon. Not. R. Astron. Soc.* **168** 60
- Malhotra R 1993 *Nature* **365** 819
- Mao S and Paczynski B 1991 *Astrophys. J.* **374** L37
- Marcy G W and Butler R P 1995 *187th AAS Meeting Bull. Am. Astron. Soc.* **27** 1379
- Marcy G W and Butler R P 1998 *Annu. Rev. Astron. Astrophys.* **36** 57
- Marcy G W, Butler R P, Fischer D, Vogt S S, Lissauer J J and Rivera E J 2001 *Astrophys. J.* **556** 296
- Marcy G W, Butler R P, Fischer D, Vogt S S, Wright J T, Tinney C G and Jones H R A 2005 *Preprint astro-ph/0505003*
- Masset F and Papaloizou J C B 2003 *Astrophys. J.* **588** 494
- Masset F and Snellgrove M D 2001 *Mon. Not. R. Astron. Soc.* **320** L55
- Masunaga H and Inutsuka S I 1999 *Astrophys. J.* **510** 822
- Mayer L, Quinn T, Wadsley J and Stadel J 2004 *Astrophys. J.* **609** 1045
- Mayor M and Queloz D 1995 *Nature* **378** 355
- Mayor M, Udry S, Naef D, Pepe F, Queloz D, Santos N C and Burnet M 2004 *Astron. Astrophys.* **415** 319
- McArthur B E *et al* 2004 *Astrophys. J.* **614** L81
- Menou K and Goodman J 2004 *Astrophys. J.* **606** S20
- Mizuno H 1980 *Prog. Theor. Phys.* **64** 544
- Motte F and André P 2001 *Astron. Astrophys.* **365** 440
- Murray N, Hansen B, Holman M and Tremaine S 1998 *Science* **279** 69
- Nakagawa Y, Nakazawa K and Hayashi C 1981 *Icarus* **45** 517
- Nelson A F, Benz W and Ruzmaikina T V 2000 *Astrophys. J.* **529** 357
- Nelson R P 2005 *Astron. Astrophys.* at press *Preprint astro-ph/0508486*
- Nelson R P and Papaloizou J C B 2004 *Mon. Not. R. Astron. Soc.* **350** 849
- Nelson R P, Papaloizou J C B, Masset F and Kley W 2000 *Mon. Not. R. Astron. Soc.* **318** 18
- Papaloizou J C B 2002 *Astron. Astrophys.* **388** 615
- Papaloizou J C B 2003 *Celest. Mech. Dyn. Astron.* **87** 53
- Papaloizou J C B 2005 *Celest. Mech. Dyn. Astron.* **91** 33
- Papaloizou J C B and Larwood J D 2000 *Mon. Not. R. Astron. Soc.* **315** 823
- Papaloizou J C B and Lin D N C 1984 *Astrophys. J.* **285** 818
- Papaloizou J C B and Nelson R P 2005 *Astron. Astrophys.* **433** 247
- Papaloizou J C B, Nelson R P and Snellgrove M D 2004 *Mon. Not. R. Astron. Soc.* **350** 829
- Papaloizou J C B and Savonije G J 1991 *Mon. Not. R. Astron. Soc.* **248** 353
- Papaloizou J C B and Terquem C 1999 *Astrophys. J.* **521** 823
- Papaloizou J C B and Terquem C 2001 *Mon. Not. R. Astron. Soc.* **325** 221
- Perri F and Cameron A G W 1974 *Icarus* **22** 416
- Perryman M and Hainault O 2005 Extra-solar planets *ESA-ESO working groups report* (Paris: ESA and ESO) http://www.eso.org/gen-fac/pubs/esaesowg/espwg_report.pdf
- Pickett B K, Cassen P, Durisen R H and Link R 1998 *Astrophys. J.* **504** 468
- Pickett B K, Cassen P, Durisen R H and Link R 2000a *Astrophys. J.* **529** 1034
- Pickett B K, Cassen P, Durisen R H and Link R 2000a *Astrophys. J.* **530** 1106 (erratum)
- Pickett B K, Durisen R H, Cassen P and Mejia A C 2000b *Astrophys. J.* **540** L95
- Pinsonneault M H, DePoy D L and Coffee M 2001 *Astrophys. J.* **556** L59
- Pollack J B, Hubickyj O, Bodenheimer P, Lissauer J J, Podolak M and Greenzweig Y 1996 *Icarus* **124** 62
- Rafikov R 2002 *Astrophys. J.* **572** 566
- Rasio F A and Ford E B 1996 *Science* **274** 54

- Rice W K M and Armitage P J 2003 *Astrophys. J.* **598** L55
- Rice W K M, Lodato G, Pringle J E, Armitage P J and Bonnell I A 2004 *Mon. Not. R. Astron. Soc.* **355** 543
- Sadakane K, Ohkubo M, Takeda Y, Sato B, Kambe E and Aoki W 2002 *Publ. Astron. Soc. Japan* **54** 911
- Safronov V S 1969 *Evolutsiia doplanetnogo oblaka* (Moscow: Nauka)
- Safronov V S 1972 *Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets* (Jerusalem: IPST) (Engl. Transl.)
- Sandquist E L, Taam R, Lin D N C, Burkert A 1998 *Astrophys. J.* **506** L65
- Sandquist E L, Dokter J J, Lin D N C and Mardling R A 2002 *Astrophys. J.* **572** 1012
- Santos N C, García López R J, Israelian G, Mayor M, Rebolo R, García-Gil A, Pérez de Taoro M R and Randich S 2002 *Astron. Astrophys.* **386** 1028
- Santos N C, Israelian G and Mayor M 2001 *Astron. Astrophys.* **373** 1019
- Santos N C, Israelian G, Mayor M, Rebolo R and Udry S 2003 *Astron. Astrophys.* **398** 363
- Sekiya M 1998 *Icarus* **133** 298
- Shakura N I and Sunyaev R A 1973 *Astron. Astrophys.* **24** 337
- Snellgrove M, Papaloizou J C B and Nelson R P 2001 *Astron. Astrophys.* **374** 1092
- Sozzetti A 2004 *Mon. Not. R. Astron. Soc.* **354** 1194
- Stapelfeldt K R, Krist J E, Menard F, Bouvier J, Padgett D L and Burrows C J 1998 *Astrophys. J.* **502** 65
- Stevenson D J 1982 *Planet. Space. Sci.* **30** 755
- Syer D and Clarke C J 1995 *Mon. Not. R. Astron. Soc.* **277** 758
- Tanaka H, Takeuchi T and Ward W R 2002 *Astrophys. J.* **565** 1257
- Tanaka H and Ward W R 2004 *Astrophys. J.* **602** 388
- Tanga P, Babiano A, Dubrulle B and Provenzale A 1996 *Icarus* **121** 158
- Terquem C E J M L J 2003 *Mon. Not. R. Astron. Soc.* **341** 1157
- Terquem C and Papaloizou J C B 2002 *Mon. Not. R. Astron. Soc.* **332** 39
- Terquem C, Papaloizou J C B and Nelson R P 2000 *From Dust to Terrestrial Planets* ed W Benz *et al* (ISSI Space Sciences Series 9) (Reprinted from Space Science Reviews 92) p 323
- Thommes E W, Duncan M J and Levison H F 2002 *Astron. J.* **123** 2862
- Thorsett S E, Arzoumanian Z, Camilo F and Lyne A G 1999 *Astrophys. J.* **523** 763
- Trilling D E, Benz W, Guillot T, Lunine J I, Hubbard W B and Burrows A 1998 *Astrophys. J.* **500** 428
- Trilling D E, Lunine J I and Benz W 2002 *Astron. Astrophys.* **394** 241
- Udry S, Mayor M and Santos N C 2003 *Astron. Astrophys.* **407** 369
- Vidal-Madjar A, Désert J M, Lecavelier des Etangs A, Hébr G, Ballester G E, Ehrenreich D, Ferlet R, McConnell J C, Mayor M and Parkinson C D 2004 *Astrophys. J.* **604** L69
- Vidal-Madjar A, Lecavelier des Etangs A, Désert J M, Ballester G E, Ferlet R, Hébrard G and Mayor M 2003 *Nature* **422** 143
- Ward W R 1986 *Icarus* **67** 164
- Ward W R 1997 *Icarus* **126** 261
- Ward W R and Hahn J M 1995 *Astrophys. J.* **440** L25
- Weidenschilling S J 1977 *Mon. Not. R. Astron. Soc.* **180** 57
- Weidenschilling S J 1980 *Icarus* **44** 172
- Weidenschilling S J 1984 *Icarus* **60** 553
- Weidenschilling S J 2003 *Icarus* **165** 438
- Weidenschilling S J and Marzari F 1996 *Nature* **384** 619
- Wetherill G W and Stewart G R 1989 *Icarus* **77** 330
- Wolszczan A 1996 *Proc. IAU Colloq. 160 Pulsars: problems and progress (Sydney)* (ASP Conf. Series vol 105) ed S Johnston *et al* (San Francisco: ASP) p 91
- Wolszczan A and Frail D A 1992 *Nature* **355** 145
- Wu Y and Murray N 2003 *Astrophys. J.* **589** 605
- Wuchterl G 1995 *Earth Moon Planets* **67** 51
- Youdin A N and Shu F H 2002 *Astrophys. J.* **580** 494
- Zapatero Osorio M R, Béjar V J S, Martín E L, Rebolo R, Barrado y Navascués D, Bailer-Jones C A L and Mundt R 2000 *Science* **290** 103
- Zapatero Osorio M R, Béjar V J S, Martín E L, Rebolo R, Barrado y Navascués D, Mundt R, Eislöffel J and Caballero J A 2002 *Astrophys. J.* **578** 536
- Zucker Shay and Mazeh T 2002 *Astrophys. J.* **568** L113