LECTURE NOTES ON THE FORMATION AND EARLY EVOLUTION OF PLANETARY SYSTEMS*

Philip J. Armitage

JILA, 440 UCB, University of Colorado, Boulder, CO80309-0440

These notes provide an introduction to the theory of the formation and early evolution of planetary systems. Topics covered include the structure, evolution and dispersal of protoplanetary disks; the formation of planetesimals, terrestrial and gas giant planets; and orbital evolution due to gas disk migration, planetesimal scattering, and planet-planet interactions.

arXiv:astro-ph/0701485v2 11 Aug 2009

Contents

I. Introduction 1
A. Critical Solar System Observations 2
1. Architecture 2
2. Mass and angular momentum 2
3. Minimum mass Solar Nebula 2
4. Resonances 3
5. Minor bodies 3
6. Ages 3
7. Satellites 4
B. Extrasolar Planets 4
1. Detection methods and biases 4
2. Observed properties 6

II. Protoplanetary Disks 7
A. The star formation context 7
B. Passive circumstellar disks 8
1. Vertical structure 9
2. Radial temperature profile 9
3. Spectral energy distribution (SED) 10
4. Sketch of more complete models 11
C. Actively accreting disks 11
1. Diffusive evolution equation 12
2. Solutions 12
3. Temperature profile 14
4. Origin of angular momentum transport 15
5. Layered disks 17
6. Disk dispersal 18
D. The condensation sequence 20

III. Planet Formation 21
A. Planetesimal formation 22
1. Dust settling 22
2. Settling with coagulation 23
3. Radial drift of particles 24
4. The Goldreich-Ward mechanism 26
B. Growth beyond planetesimals 30
1. Gravitational focusing 30
2. Growth versus fragmentation 31
3. Shear versus dispersion dominated encounters 32
4. Growth rates 33
5. Isolation mass 34
6. Coagulation equation 34
7. Overview of terrestrial planet formation 35
C. Gas giant formation 36
1. Core accretion model 36
2. Gravitational instability model 38
3. Comparison with observations 40

IV. Evolution of Planetary Systems 40
A. Gas disk migration 41
1. Conditions for resonance 41
2. Gravitational torques at resonances 42
3. Type I migration 43
4. Type II migration 44
5. The Type II migration rate 45
6. Stochastic migration 46
7. Eccentricity evolution during migration 47
8. Observational evidence for inner holes 47
B. Planetesimal disk migration 47
1. Solar System evidence 48
2. The Nice model 48
C. Planet-planet scattering 49
1. Hill stability 49
2. Scattering and exoplanet eccentricities 51
D. Predictions of migration theories 52

Acknowledgements 53

References 53

I. INTRODUCTION

The theoretical study of planet formation has a long history. Many of the fundamental ideas in the theory of terrestrial planet formation were laid out by Safronov (1969) in his classic monograph 'Evolution of the Protoplanetary Cloud and Formation of the Earth and the Planets', while the core accretion theory for gas giant formation, which was discussed by Cameron in the early 1970's (Perri & Cameron, 1973), had been quantitatively developed in recognizable detail by 1980 (Mizuno, 1980). A wealth of new data over the last decade — including observations of protoplanetary disks, the discovery of the Solar System's Kuiper Belt, and the detection of numerous extrasolar planetary systems — has led to renewed interest in the problem. Although these observations have confirmed some existing predictions, they have also emphasized the need to explore new theoretical avenues. The major questions that work in this field seeks to answer include:

- How did the terrestrial and giant planets form?
- How much evolution in the orbits of planets takes place at early times?

*To be published in greatly expanded form as *Astrophysics of Planet Formation* by Cambridge University Press (November 2009).

TABLE I Basic properties of planets in the Solar System

	a/AU	e	M_p/g
Mercury	0.387	0.206	3.3×10^{26}
Venus	0.723	0.007	4.9×10^{27}
Earth	1.000	0.017	6.0×10^{27}
Mars	1.524	0.093	6.4×10^{26}
Jupiter	5.203	0.048	1.9×10^{30}
Saturn	9.537	0.054	5.7×10^{29}
Uranus	19.189	0.047	8.7×10^{28}
Neptune	30.070	0.009	1.0×10^{29}

- Is the architecture of the Solar System typical?
- How common are habitable planets?

The main goal of these notes is to provide a succinct introduction to the critical concepts necessary to understand the astrophysics of planet formation. Before delving into theory, however, we first briefly review the basic observational properties of the Solar System and of extrasolar planetary systems that a theory of planet formation might aspire to explain.

A. Critical Solar System Observations

1. Architecture

The orbital properties and masses of the planets in the Solar System are listed in Table I (the values here are taken from the JPL web site). The basic architecture of our Solar System comprises 2 *gas giants* (Jupiter and Saturn) composed primarily of hydrogen and helium, though not of Solar composition. Saturn is known to have a substantial core. Descending in mass there are then 2 *ice giants* (Uranus and Neptune) composed of water, ammonia, methane, silicates and metals, plus low mass hydrogen / helium atmospheres; 2 large *terrestrial planets* (Earth and Venus) plus two smaller terrestrial planets (Mercury and Mars). Apart from Mercury, all of the planets have low eccentricities and orbital inclinations. They orbit in a plane that is approximately perpendicular to the Solar rotation axis (though there is a small misalignment of 7°).

In the Solar System the giant and terrestrial planets are clearly segregated in orbital radius, with the inner zone occupied by the terrestrial planets being separated from the outer giant planet region by the main asteroid belt. The orbital radii of the giant planets coincide with where we expect the protoplanetary disk to have been cool enough for ices to have been present. This is a significant observation in the classical theory of giant planet formation, since in that theory the time scale for giant planet formation is a sensitive function of the mass of condensible materials.

2. Mass and angular momentum

The mass of the Sun is $M_\odot = 1.989 \times 10^{33}$ g, made up of hydrogen (fraction by mass $X = 0.73$), helium ($Y = 0.25$) and 'metals' ($Z = 0.02$). One observes immediately that most of the heavy elements in the Solar System *are found in the Sun* rather than in the planets. The import of this trivial observation is that if most of the mass in the Sun passed through a disk at some juncture during the star formation process the planet formation need not be very efficient.

The angular momentum budget for the Solar System is dominated by the orbital angular momentum of the planets. The angular momentum in the Solar rotation is,

$$L_\odot \simeq k^2 M_\odot R_\odot^2 \Omega, \tag{1}$$

assuming for simplicity solid body rotation. Taking $\Omega = 2.9 \times 10^{-6}$ s^{-1} and adopting $k^2 = 0.1$ (roughly appropriate for a star with a radiative core), $L_\odot \simeq 3 \times 10^{48}$ gcm^{2}s^{-1}. By comparison, the orbital angular momentum of Jupiter is,

$$L_J = M_J \sqrt{G M_\odot a} = 2 \times 10^{50} \mathrm{gcm^2 s^{-1}}. \tag{2}$$

The significance of this result is that it implies that substantial segregation of mass and angular momentum must have taken place during (and subsequent to) the star formation process. We will look into how such segregation arises during disk accretion later, but the broader question of exactly how the angular velocity of low mass stars evolves at early times remains a subject of active research (Herbst et al., 2007).

3. Minimum mass Solar Nebula

We can use the observed masses and compositions of the planets to derive a *lower limit* to the amount of gas that must have been present when the planets formed. This is called the Minimum Mass Solar Nebula (Weidenschilling, 1977). The procedure is:

1. Start from the known mass of heavy elements (say iron) in each planet, and augment this mass with enough hydrogen and helium to bring the mixture to Solar composition. This is a mild augmentation for Jupiter, but a lot more for the Earth.
2. Then divide the Solar System into annuli, with one planet per annulus. Distribute the augmented mass for each planet uniformly across the annuli, to yield a characteristic gas surface density Σ (units gcm^{-2}) at the location of each planet.

The result is that between Venus and Neptune (and ignoring the asteroid belt) $\Sigma \propto r^{-3/2}$. To derive a precise normalization from such a hand-waving procedure is somewhat pointless, but if one needs a specific number the most common value used is that due to Hayashi

(1981),

$$\Sigma = 1.7 \times 10^3 \left(\frac{r}{\mathrm{AU}}\right)^{-3/2} \ \mathrm{gcm}^{-2}. \quad (3)$$

Integrating this expression out to 30 AU the enclosed mass works out to be around 0.01 $M_\odot$, comparable to the 'typical' mass estimated for protoplanetary disks around other stars as inferred from mm observations of the dust.

As the name should remind you, this is a *minimum* mass. It is not an estimate of the disk mass at the time the Solar Nebula formed, nor is there any reason to believe that the $\Sigma \propto r^{-3/2}$ scaling represents the steady-state surface density profile for a protoplanetary disk. Most theoretical models of disks, in fact, predict a significantly shallower slope more akin to $\Sigma \propto r^{-1}$ (Bell et al., 1997).

4. Resonances

A *resonance* occurs when there is a near-exact relation between characteristic frequencies of two bodies. For example, a mean-motion resonance occurs between two planets with orbital periods P_1 and P_2 when,

$$\frac{P_1}{P_2} \simeq \frac{i}{j}, \quad (4)$$

with i, j integers (the resonance is typically important if i and j are *small* integers). In the Solar System Neptune and Pluto (along with many other Kuiper Belt objects) are in a 3:2 resonance, while Jupiter and Saturn are close to a 5:2 mean-motion resonance (known as 'the great inequality') which influences their motion (Lovett, 1895). There are no simple resonances among the major planets. There are, however, many resonant pairs among planetary moons. Jupiter's satellites Io, Europa and Ganymede, for example, form a resonant chain in which Io is in 2:1 resonance with Europa, which itself is in a 2:1 resonance with Ganymede. The existence of these non-trivial configurations is generally regarded as strong circumstantial evidence that dissipative processes (such as tides) resulted in orbital evolution at some point in the past history of the systems (Goldreich, 1965).

5. Minor bodies

As a very rough generalization the Solar System is dynamically full, in the sense that most locations where test particle orbits would be stable for 5 Gyr are in fact occupied by minor bodies. In the inner and middle Solar System the main asteroid belt is the largest reservoir of minor bodies. The asteroid belt displays considerable structure, most notably in the form of sharp decreases in the number of asteroids in zones known as the *Kirkwood gaps*. The existence of these gaps provides a striking illustration of the importance of resonances (in this case with Jupiter) in influencing dynamics.

The properties of objects beyond Neptune (Chiang et al., 2007; Jewitt & Luu, 1993) provide critical constraints on both the early evolution of the outer Solar System (Malhotra, 1993) and collisional models for planet formation (Kenyon, 2002). Kuiper Belt properties include:

1. A large population of objects in Pluto-like orbits in 3:2 resonance with Neptune ('plutinos').

2. A dearth of KBOs in orbits with 36 AU $< a <$ 39 AU.

3. An apparent edge to the distribution of Classical KBOs at about 50 AU (Trujillo, Jewitt & Luu, 2001).

4. A differential size distribution (deduced indirectly from the measured luminosity function) that is roughly a power-law for large bodies with diameters $D \gtrsim 100$ km (Trujillo, Jewitt & Luu, 2001). A recent determination by Fraser & Kavelaars (2009) infers a power-law slope $q \simeq 4.8$ for large bodies together with a break to a much shallower slope at small sizes.

Kuiper Belt Objects are commonly classified into several dynamically distinct families. *Resonant KBOs* are those — like Pluto — that exhibit mean-motion resonances with Neptune. *Centaurs* are non-resonant KBOs which have perihelion distances interior to the orbit of Neptune. *Classical KBOs* are objects further out whose orbits have been little influenced by Neptune. Finally, *scattered disk KBOs* are bodies with perihelia beyond the orbit of Neptune that do not fall into the other classes.

Beyond the Kuiper belt the most intriguing object known is *Sedna*, a large object with semi-major axis $a = 480 \pm 40$ AU, eccentricity $e = 0.84 \pm 0.01$, and inclination $i = 12°$ (Brown, Trujillo & Rabinowitz, 2004). Since Sedna was discovered close to perihelion, it is highly likely to represent the first of a substantial new class of objects with perihelion distances substantially exterior to the orbit of Neptune. It may represent an object in an inner extension of the Oort cloud.

6. Ages

Radioactive dating of meteorites provides an absolute measure of the age of the Solar System, together with constraints on the time scales of some phases of planet formation. The details are beyond the scope of these lectures; typical numbers quoted are a Solar System age of 4.57 Gyr, a time scale for the formation of large bodies within the asteroid belt of < 5 Myr (Wadhwa et al., 2007), and a time scale for final assembly of the Earth of ~ 100 Myr.

7. Satellites

Most of the planets possess satellite systems, some of which are very extensive. Aspects of their formation or capture are discussed by Canup & Ward (2002) and by Nesvorný et al. (2003). The *regular* satellites (those which have prograde orbits approximately coincident with the equatorial plane of their host planet) clearly provide evidence that the giant planets themselves were surrounded by sub-disks of at early times, and much of the theory of planet formation is at least qualitatively transferable to the satellite problem. Although interesting in its own right, the study of regular satellite formation has not yet yielded any particularly dramatic insights into the formation of planets within larger scale disks.

B. Extrasolar Planets

1. Detection methods and biases

The most important current methods for detecting and characterizing extrasolar planets are:

1. Radial velocity surveys of nearby, typically Solar-type stars (Butler et al., 1996). More than 300 planets have been found with this technique.

2. Blind transit searches, and follow-up of radial velocity discovered planets that happen to show transits (Charbonneau et al., 2007). There has been explosive growth in the success of this technique over the last couple of years, and more than 60 such planets are now known. The pace of discovery will accelerate further as more results from the space-based *COROT* mission (Baglin et al., 2002) are released, and when the yet more sensitive *Kepler* (Borucki et al., 2003) results arrive.

3. Gravitational lensing (Beaulieu et al., 2006).

4. Direct imaging, which although presently limited to searches for massive planets at large orbital radii, has already uncovered the exceptionally interesting multiple system surrounding the star HR 8799 (Marois et al., 2008). This system consists of three very massive planets orbiting at projected separations between 20 and 70 AU from the star.

5. Pulsar timing (Wolszczan & Frail, 1992).

In addition to these methods astrometry, and transit timing (Agol et al., 2005; Holman & Murray, 2005) have significant future potential.

Of the extant methods, the most important is radial velocity surveys. 51 Peg, the first known extrasolar planet orbiting a normal star, was discovered this way (Mayor & Queloz, 1995), and most of our current knowledge of the extrasolar planet population derives from radial velocity surveys (Marcy et al., 2005). The observable

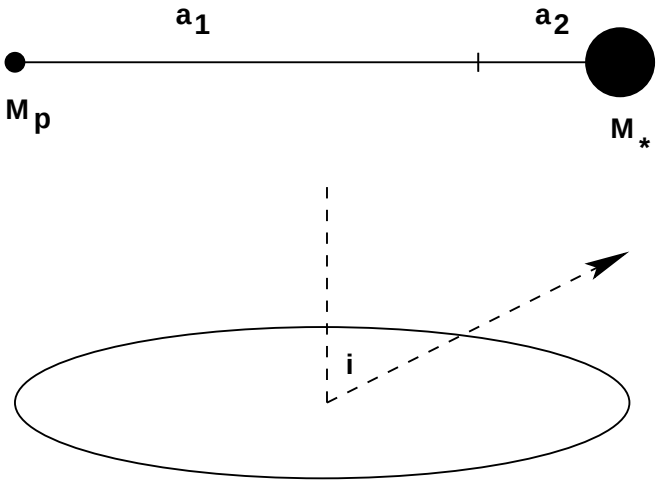

FIG. 1 A planet of mass M_p orbits the common center of mass at distance a_1, while the star of mass M_* orbits at distance a_2. The system is observed at inclination angle i.

is the time dependence of the radial velocity of a star due to the presence of an orbiting planet. For a planet on a circular orbit the geometry is shown in Figure 1. The star orbits the center of mass with a velocity,

$$v_* \simeq \left(\frac{M_p}{M_*}\right)\sqrt{\frac{GM_*}{a}}. \tag{5}$$

Observing the system at an inclination angle i, we see the radial velocity vary with a semi-amplitude $K = v_* \sin i$,

$$K \propto M_p \sin i a^{-1/2}. \tag{6}$$

If the inclination is unknown, what we measure (K) determines a *lower limit* to the planet mass M_p. Note that M_* is not determined from the radial velocity curve, but must instead be determined from the stellar spectral properties. If the planet has an eccentric orbit, e can be determined by fitting the non-sinusoidal radial velocity curve.

The noise sources for radial velocity surveys comprise photon noise, intrinsic jitter in the star (e.g. from convection or stellar oscillations), and instrumental effects. The magnitude of these effects vary (sometimes dramatically) from star to star. However, if we imagine an idealized survey for which the noise per observation was a constant, then the selection limit would be defined by,

$$M_p \sin i|_{\rm minimum} = Ca^{1/2}, \tag{7}$$

with C a constant. Planets with masses below this threshold would be undetectable, as would planets with orbital periods exceeding the duration of the survey (since orbital solutions are typically poorly constrained when only part of an orbit is observed unless the signal to noise of the observations is very high). The selection boundary defined by these limits is shown schematically in Figure 2.

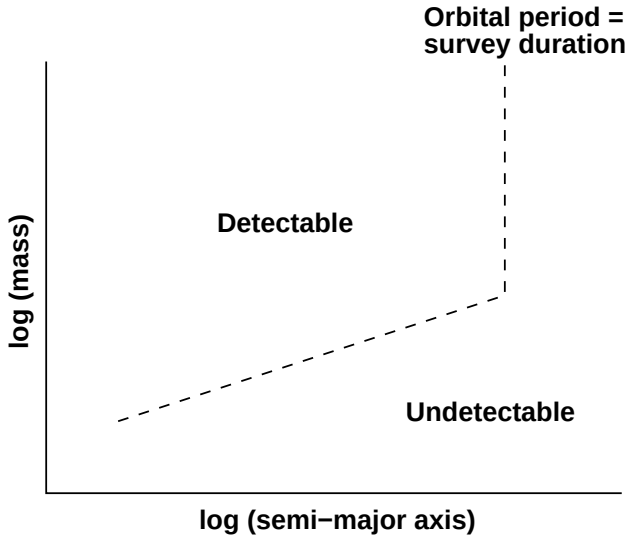

FIG. 2 Highly schematic illustration of the selection function of an idealized radial velocity survey. The minimum mass planet that can be detected scales with semi-major axis as $a^{1/2}$ until the orbital period of the planet exceeds the duration of the survey.

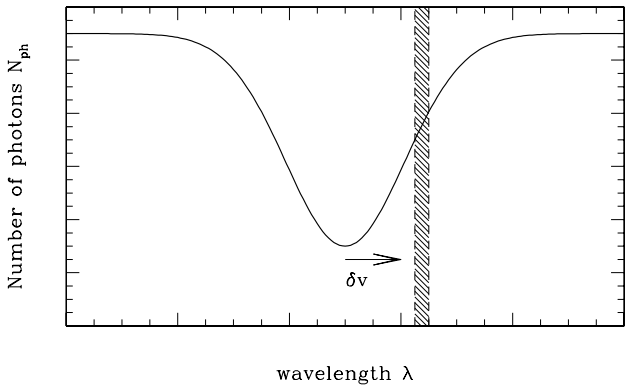

FIG. 3 Schematic spectrum in the vicinity of a single spectral line of the host star. The wavelength range that corresponds to a single pixel in the observed spectrum is shown as the vertical shaded band. If the spectrum shifts by a velocity δv the number of photons detected at that pixel will vary by an amount that depends upon the local slope of the spectrum.

Extremely accurate radial velocity measurements are a prerequisite for discovering planets via this technique. For the Solar System,

$$\begin{aligned} v_* &\approx 12 \text{ ms}^{-1} \text{ (Jupiter)} \\ v_* &\approx 0.1 \text{ ms}^{-1} \text{ (Earth).} \end{aligned} \quad (8)$$

Given that astronomical spectrographs have a resolving power of the order of 10^5 (which corresponds, in velocity units, to a precision of the order of *kilometers* per second) it might seem impossible to find planets with such small radial velocity signatures. To appreciate how detection of small (sub-pixel) shifts is possible, it is useful to consider the precision that is possible against the background of shot noise (i.e. uncertainty in the number of photons due purely to counting statistics). An estimate of the photon noise limit can be derived by considering a very simple problem: how accurately can velocity shifts be estimated given measurement of the flux in a single pixel on the detector? To do this, we follow the basic approach of Butler et al. (1996) and consider the spectrum in the vicinity of a spectral line, as shown in Figure 3. Assume that, in an observation of some given duration, $N_{\rm ph}$ photons are detected in the wavelength interval corresponding to the shaded vertical band. If we now imagine displacing the spectrum by an amount (in velocity units) δv the change in the mean number of photons is,

$$\delta N_{\rm ph} = \frac{{\rm d}N_{\rm ph}}{{\rm d}v}\delta v. \quad (9)$$

Since a 1σ detection of the shift requires that $\delta N_{\rm ph} \approx N_{\rm ph}^{1/2}$, the minimum velocity displacement that is detectable is,

$$\delta v_{\rm min} \approx \frac{N_{\rm ph}^{1/2}}{{\rm d}N_{\rm ph}/{\rm d}v}. \quad (10)$$

This formula makes intuitive sense – regions of the spectrum that are flat are useless for measuring δv while sharp spectral features are good. For Solar-type stars with photospheric temperatures $T_{\rm eff} \approx 6000$ K the sound speed at the photosphere is around 10 kms^{-1}. Taking this as an estimate of the thermal broadening of spectral lines, the slope of the spectrum is at most,

$$\frac{1}{N_{\rm ph}}\frac{{\rm d}N_{\rm ph}}{{\rm d}v} \sim \frac{1}{10 \text{ kms}^{-1}} \sim 10^{-4} \text{ m}^{-1}\text{s}. \quad (11)$$

Combining Equations (10) and (11) with knowledge of the number of photons detected per pixel yields an estimate of the photon-limited radial velocity precision. For example, if the spectrum has a signal to noise ratio of 100 (and there are no other noise sources) then each pixel receives $N_{\rm ph} \sim 10^4$ photons and $\delta v_{\rm min} \sim 100$ ms^{-1}. If the spectrum contains $N_{\rm pix}$ such pixels the combined limit to the radial velocity precision is,

$$\delta v_{\rm shot} = \frac{\delta v_{\rm min}}{N_{\rm pix}^{1/2}} \sim \frac{100 \text{ ms}^{-1}}{N_{\rm pix}^{1/2}}. \quad (12)$$

Obviously this discussion ignores many aspects that are practically important in searching for planets from radial velocity data. However, it suffices to reveal the key feature: given a high signal to noise spectrum and stable wavelength calibration, photon noise is small enough that a radial velocity measurement with the ms^{-1} precision needed to detect extrasolar planets is feasible.

Records for the smallest amplitude radial velocity signal that can be extracted from the noise are regularly bested. Currently some of the highest precision radial velocity measurements have an RMS scatter of around

0.5 ms^{-1}, and it has been suggested that 0.2 ms^{-1} might be attainable for some quiescent stars (Mayor & Udry, 2008). The lowest stellar velocity semi-amplitude is slightly below 2 ms^{-1}. It is important to remember that these are best-case values – complete samples of extrasolar planets that are suitable for statistical studies only exist for much larger $K \approx 30$ ms^{-1} (Fischer & Valenti, 2005).

Detailed modeling is necessary in order to assess whether a particular survey has a selection bias in eccentricity. Naively you can argue it either way – an eccentric planet produces a larger perturbation at closest stellar approach, but most of the time the planet is further out and the radial velocity is smaller. A good starting point for studying these issues is the explicit calculation for the Keck Planet Search reported by Cumming et al. (2008). These authors find that the Keck search is complete for sufficiently massive planets (and thus trivially unbiased) for $e \lesssim 0.6$.

2. Observed properties

For most known extrasolar planets, our information is limited to those quantities derived from the radial velocity observables: a lower limit on the mass $M_p \sin i$, the semi-major axis a, the eccentricity e, and the longitude of pericenter ϖ. In addition, estimates of the host star's mass and metallicity are available. The distribution of planets in $M_p \sin i$, a and e is depicted in Figures 4, 5 and 6, using data for radial velocity detected planets from an updated version of the Butler et al. (2006) catalog.

Marcy et al. (2005) quote the following results from the Lick / Keck / AAT survey, which has monitored 1330 FGKM stars for the better part of a decade:

1. The detected giant planet frequency within $a \sim 5$ AU is $\simeq 7\%$. This is certainly a lower limit as many giant planets would fall below the selection threshold at larger orbital radii (c.f. Figure 5).

2. The frequency of hot Jupiters with $a < 0.1$ AU is approximately 1%. The abundance of planets with orbital radius – measured as $dN_p/d\log a$ – increases to large a.

3. Eccentric orbits are common beyond the radius where tidal circularization is significant (Figure 4). The median eccentricity of planets orbiting between 1 and 3 AU is $\langle e \rangle \simeq 0.28$. Some extremely eccentric planets exist. There is no strong trend of eccentricity with planet mass (Figure 6).

4. The planet mass function declines toward large masses (Butler et al., 2006; Tabachnik & Tremaine, 2002).

5. Planet frequency rises rapidly with host metallicity. This trend, shown in Figure 7 using data from Fischer & Valenti (2005) is dramatic — relatively modest increases in metallicity substantially enhance

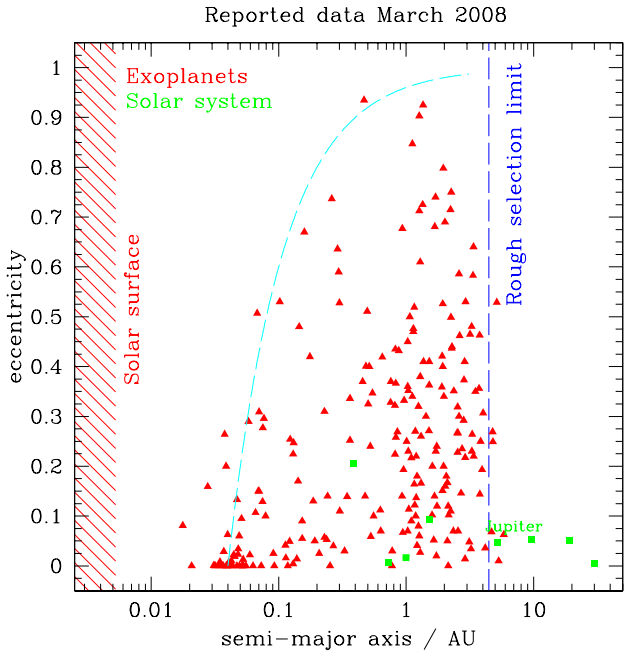

FIG. 4 The distribution of known extrasolar planets in semi-major axis and eccentricity (red triangles). Solar System planets are shown for comparison as the green squares. The blue curve denotes a line of constant periastron distance. The figure uses data from an updated version of the Butler et al. (2006) catalog, and includes planets that have $M_p \sin i < 10 M_J$.

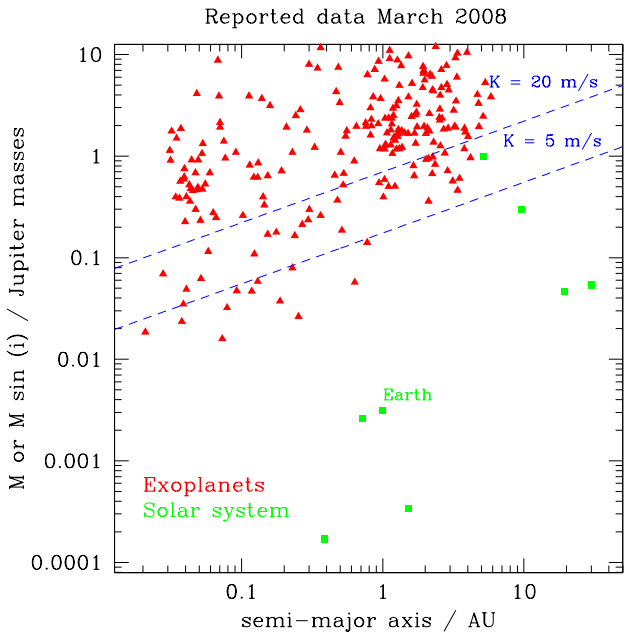

FIG. 5 The distribution of known extrasolar planets in semi-major axis and minimum mass. Lines of constant semi-amplitude radial velocity perturbation are plotted assuming a Solar mass host. It is clear by eye that the typical extrasolar planet detected so far is not a hot Jupiter but rather orbits at $a > 1$ AU.

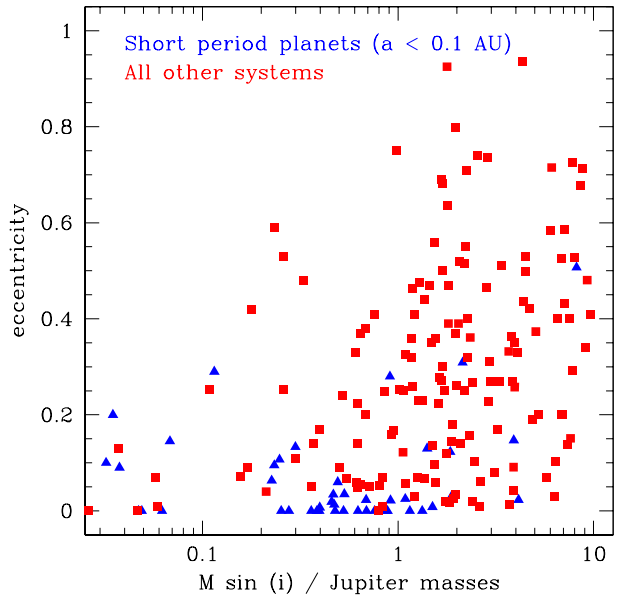

FIG. 6 Eccentricity vs mass for known extrasolar planets, divided into short period planets ($a < 0.1$ AU, shown as blue triangles) and all other systems (shown as red squares). The short period planets have low eccentricity orbits, presumably as a consequence of tidal interactions with their host stars. No strong correlation of eccentricity with mass is seen.

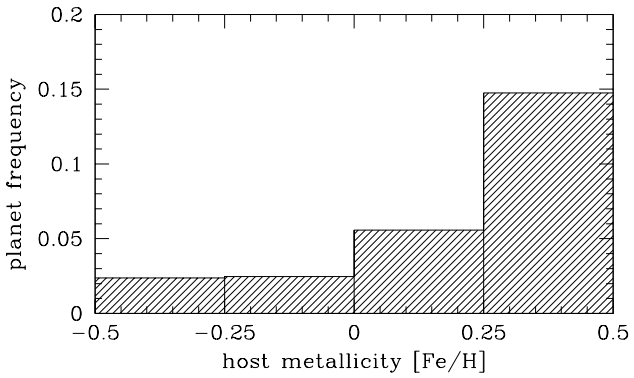

FIG. 7 The fraction of stars that host currently known extrasolar planets is plotted as a function of the stellar metallicity, from data (their Figure 4) reported by Fischer & Valenti (2005).

the probability that currently detectable planets will be found around a star.

6. Multiple planet systems are common, of which perhaps a third exhibit prominent mean-motion resonances.

Transit surveys contribute complementary information about the extrasolar planet population. Ground based surveys – which due to atmospheric limitations on photometric precision are only sensitive to massive planets – have discovered numerous short and ultrashort period planets (e.g. WASP-12b with a period of only 1.09 days) and measured their physical radii. The planetary radii confirm their gas giant nature. Space based transit surveys, together with the highest precision radial velocity measurements, have extended this sample into the so-called *super-Earth* regime of planets with masses $M_p < 10\ M_\oplus$. The properties of the super-Earth population remain somewhat murky, but these may plausibly be rocky bodies that bridge the gap between the Solar System's terrestrial and giant planets.

Transit observations have also provided two surprises. First, although the basic confirmation that massive planets have radii similar to that of Jupiter is reassuring, the scatter in the observed radii is inconsistent with the simplest theoretical predictions. These radius anomalies (Torres, Winn & Holman, 2008) are puzzling, and appear to reflect either limitations in our understanding of planetary structure (Burrows et al., 2007) or the existence of unspecified extra sources of heat in some of these planets. That composition plays some role is relatively clear. Sato et al. (2005) infer that the Saturn mass planet orbiting HD 149026 needs to have a very massive core[1] of $M_{\rm core} \simeq 70 M_\oplus$ in order to match the observed radius for that object. Second, the combination of transit and radial velocity data has permitted a number of measurements of the *orientation* of the orbital plane relative to the stellar spin axis via detection of the Rossiter-McLaughlin effect (Rossiter, 1924)[2]. Both intuition and Solar System precedent would suggest that any misalignment between the stellar and orbital angular momentum vectors ought to be small, but while this is often the case at least some systems show quite large misalignments. The XO-3 system, for example, shows a well measured misalignment of 37 ± 4 degrees (Winn et al., 2009). This is of interest for the possible constraints it places on theoretical models of orbital migration of massive planets.

II. PROTOPLANETARY DISKS

A. The star formation context

Stars form in the Galaxy today from the small fraction of gas that exists in dense, molecular clouds. Molecular clouds are observed in one or more molecular tracers –

[1] Note that interpreting the observations in terms of a 'core' involves the invocation of some theoretical prejudice, since the radius is primarily sensitive to the total heavy element abundance which need not be in the form of a well-defined solid core.

[2] The effect, originally measured for binary stars, consists of a shift in the apparent stellar radial velocity during transit as the planet obscures parts of the photosphere that are rotating either toward or away from the observer.

examples include CO, ^{13}CO and NH_3 – which can be used both to probe different regimes of column density and to furnish kinematic information that can give clues as to the presence of rotation, infall and outflows. Observations of the dense, small scale *cores* within molecular clouds (with scales of the order of 0.1 pc) that are the immediate precursors of star formation show velocity gradients that are of the order of 1 kms^{-1}pc^{-1}. Even if all of such a gradient is attributed to rotation, the parameter,

$$\beta \equiv \frac{E_{\rm rot}}{|E_{\rm grav}|} \tag{13}$$

is small – often of the order of 0.01. Hence rotation is dynamically unimportant during the early stages of collapse. The angular momentum, on the other hand, is large, with a ballpark figure being $J_{\rm core} \sim 10^{54}$ gcm^{2}s^{-1}. This is much larger than the angular momentum in the Solar System, never mind that of the Sun, a discrepancy that is described as the *angular momentum problem* of star formation. The overall solution to this problem is thought to be an undetermined admixture of binary formation, angular momentum loss in outflows, and disk formation. For our purposes, it suffices to note that the specific angular momentum of gas in molecular cloud cores would typically match the specific angular momentum of gas in Keplerian orbit around a Solar mass star at a radius of $\sim 10 - 10^2$ AU.

The bottom line is thus simply that the observed properties of molecular cloud cores are consistent with the formation of large disks – of the size of the Solar System and above – around newly formed stars. At least initially, those disks could be quite massive.

Young Stellar Objects (YSOs) are classified observationally according to the shape of their Spectral Energy Distribution $\lambda F_\lambda(\lambda)$ in the infra-red. As shown schematically in Figure 8, YSOs often display,

1. An infra-red excess (over the stellar photospheric contribution) – this is attributed to hot dust in the disk near the star.

2. An ultra-violet excess, which is ascribed to high temperature regions (probably hot spots) on the stellar surface where gas from the disk is being accreted.

To quantify the magnitude of the IR excess, it is useful to define a measure of the slope of the IR SED,

$$\alpha_{\rm IR} = \frac{\Delta \log(\lambda F_\lambda)}{\Delta \log \lambda} \tag{14}$$

between the near-IR and the mid-IR. Conventions vary, but for illustration we can assume that the slope is measured between the K band (at 2.2μm) and the N band (at 10μm). We can then classify YSOs as,

- **Class 0**: SED peaks in the far-IR or mm part of the spectrum ($\sim$ 100 μm), with no flux being detectable in the near-IR.

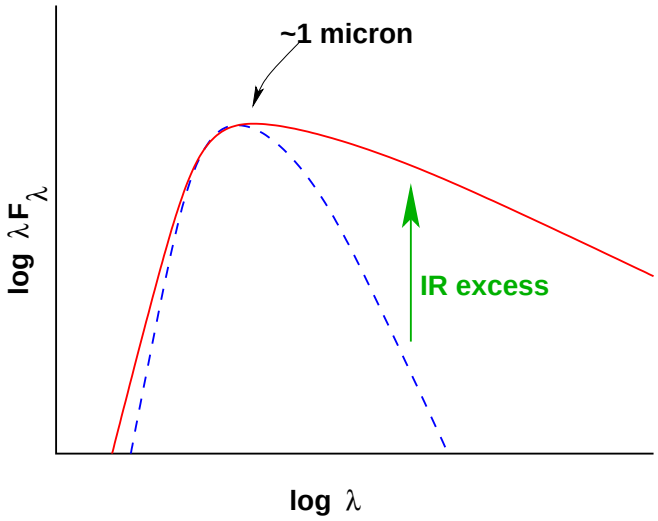

FIG. 8 Schematic depiction of the Spectral Energy Distribution of a young star surrounded by a disk. The presence of a disk is inferred from an infra-red excess (above the expected photospheric value) at wavelengths longward of around 1 μm. An ultra-violet excess is also commonly detected, and this is attributed to gas accretion on to the stellar surface producing hot spots.

- **Class I**: approximately flat or rising SED into mid-IR ($\alpha_{\rm IR} > 0$).

- **Class II**: falling SED into mid-IR ($-1.5 < \alpha_{\rm IR} < 0$). These objects are called 'Classical T Tauri stars'.

- **Class III**: pre-main-sequence stars with little or no excess in the IR. These are the 'Weak lined T Tauri stars' (note that although WTTs are defined via the equivalent width of the Hα line, this is an accretion signature that correlates well with the presence of an IR excess).

This observational classification scheme is theoretically interpreted, in part, as an evolutionary sequence (Adams, Lada & Shu, 1987). In particular, clearly objects in Classes 0 through II eventually lose their disks and become Class III sources. Viewing angle may well, however, play a role in determining whether a given source is observed as a Class I or Class II object.

B. Passive circumstellar disks

An important physical distinction needs to be drawn between *passive* circumstellar disks, which derive most of their luminosity from reprocessed starlight, and *active* disks, which are instead powered by the release of gravitational potential energy as gas flows inward. For a disk with an accretion rate $\dot{M}$, surrounding a star with luminosity $L_\odot$ and radius $R_* = 2R_\odot$, the critical accretion rate below which the accretion energy can be neglected

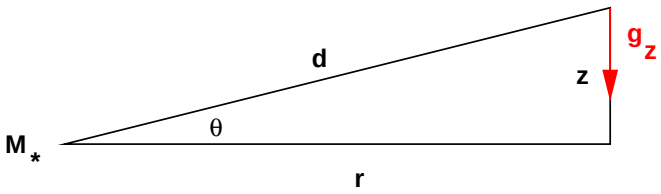

FIG. 9 Geometry for calculation of the vertical hydrostatic equilibrium of a circumstellar disk.

may be estimated as,

$$\frac{1}{4}L_\odot = \frac{GM_*\dot{M}}{2R_*}, \tag{15}$$

where we have anticipated the result, derived below, that a flat disk intercepts one quarter of the stellar flux. Numerically,

$$\dot{M} \approx 3 \times 10^{-8}\ M_\odot \mathrm{yr}^{-1}. \tag{16}$$

Measured accretion rates of Classical T Tauri stars (Gullbring et al., 1998) range from an order of magnitude above this critical rate to two orders of magnitude below, so it is oversimplifying to assume that protoplanetary disks are either always passive or always active. Rather, the thermal structure of disks at early epochs is likely dominated by internal heating due to accretion, while at late times reprocessing dominates.

1. Vertical structure

The vertical structure of a geometrically thin disk (either passive or active) is derived by considering vertical hydrostatic equilibrium (Figure 9),

$$\frac{\mathrm{d}P}{\mathrm{d}z} = -\rho g_z \tag{17}$$

where ρ is the gas density. Ignoring any contribution to the gravitational force from the disk (this is justified provided that the disk is not very massive), the vertical component of gravity seen by a parcel of gas at cylindrical radius r and height above the midplane z is,

$$g_z = \frac{GM_*}{d^2}\sin\theta = \frac{GM_*}{d^3}z. \tag{18}$$

For a thin disk $z \ll r$, so

$$g_z \simeq \Omega^2 z \tag{19}$$

where $\Omega \equiv \sqrt{GM_*/r^3}$ is the Keplerian angular velocity. If we assume for simplicity that the disk is vertically isothermal (this will be a decent approximation for a passive disk, less so for an active disk) then the equation of state is $P = \rho c_s^2$, where c_s is the sound speed. The equation of hydrostatic equilibrium (equation 17) then becomes,

$$c_s^2\frac{\mathrm{d}\rho}{\mathrm{d}z} = -\Omega^2\rho z. \tag{20}$$

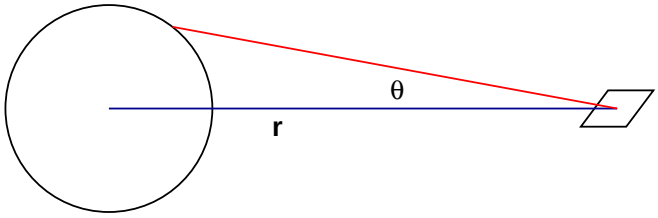

FIG. 10 Geometry for calculating the temperature profile of a flat, passive disk. We consider unit surface area in the disk plane at distance r from a star of radius R_*. The axis of spherical polar co-ordinates is the line between the surface and the center of the star, with $\phi = 0$ in the direction of the stellar pole.

The solution is,

$$\rho = \rho(z=0)e^{-z^2/2h^2} \tag{21}$$

where h, the vertical scale height, is given by,

$$h = \frac{c_s}{\Omega}. \tag{22}$$

Comparing the thickness to the radius,

$$\frac{h}{r} = \frac{c_s}{v_\phi} \tag{23}$$

where v_ϕ is the local orbital velocity. We see that the aspect ratio of the disk h/r is inversely proportional to the Mach number of the flow.

The *shape* of the disk depends upon $h(r)/r$. If we parameterize the radial variation of the sound speed via,

$$c_s \propto r^{-\beta} \tag{24}$$

then the aspect ratio varies as,

$$\frac{h}{r} \propto r^{-\beta+1/2}. \tag{25}$$

The disk will *flare* – i.e. h/r will increase with radius giving the disk a bowl-like shape – if $\beta < 1/2$. This requires a temperature profile $T(r) \propto r^{-1}$ or shallower. As we will show shortly, flaring disks are expected to be the norm, at least relatively close to the star.

2. Radial temperature profile

The physics of the calculation of the radial temperature profile of a passive disk is described in papers by Adams & Shu (1986), Kenyon & Hartmann (1987) and Chiang & Goldreich (1997). We begin by considering the absolute simplest model: a flat thin disk in the equatorial plane that absorbs all incident stellar radiation and re-emits it as a single temperature blackbody. The backwarming of the star by the disk is neglected.

We consider a surface in the plane of the disk at distance r from a star of radius R_*. The star is assumed to

be a sphere of constant brightness I_*. Setting up spherical polar co-ordinates, as shown in Figure 10, the stellar flux passing through this surface is,

$$F = \int I_* \sin\theta \cos\phi d\Omega. \tag{26}$$

We count the flux coming from the top half of the star only (and to be consistent equate that to radiation from only the top surface of the disk), so the limits on the integral are,

$$\begin{aligned} -\pi/2 < \ \phi \ &\leq \pi/2 \\ 0 < \ \theta \ &< \sin^{-1}\left(\frac{R_*}{r}\right). \end{aligned} \tag{27}$$

Substituting $d\Omega = \sin\theta d\theta d\phi$, the integral for the flux is,

$$F = I_* \int_{-\pi/2}^{\pi/2} \cos\phi d\phi \int_0^{\sin^{-1}(R_*/r)} \sin^2\theta d\theta, \tag{28}$$

which evaluates to,

$$F = I_* \left[\sin^{-1}\left(\frac{R_*}{r}\right) - \left(\frac{R_*}{r}\right)\sqrt{1-\left(\frac{R_*}{r}\right)^2}\right]. \tag{29}$$

For a star with effective temperature T_*, the brightness $I_* = (1/\pi)\sigma T_*^4$, with σ the Stefan-Boltzmann constant (Rybicki & Lightman, 1979). Equating F to the one-sided disk emission $\sigma T_{\rm disk}^4$ we obtain a radial temperature profile,

$$\left(\frac{T_{\rm disk}}{T_*}\right)^4 = \frac{1}{\pi}\left[\sin^{-1}\left(\frac{R_*}{r}\right) - \left(\frac{R_*}{r}\right)\sqrt{1-\left(\frac{R_*}{r}\right)^2}\right]. \tag{30}$$

Integrating over radii, we obtain the total disk flux,

$$\begin{aligned} F_{\rm disk} \ &= \ 2 \times \int_{R_*}^{\infty} 2\pi r \sigma T_{\rm disk}^4 dr \\ &= \ \frac{1}{4} F_*. \end{aligned} \tag{31}$$

We conclude that a flat passive disk extending all the way to the stellar equator intercepts a quarter of the stellar flux. The ratio of the observed bolometric luminosity of such a disk to the stellar luminosity will vary with viewing angle, but clearly a flat passive disk is predicted to be less luminous than the star.

The form of the temperature profile given by equation (30) is not very transparent. Expanding the right hand side in a Taylor series, assuming that $(R_*/r) \ll 1$ (i.e. far from the stellar surface), we obtain,

$$T_{\rm disk} \propto r^{-3/4}, \tag{32}$$

as the limiting temperature profile of a thin, flat, passive disk. For fixed molecular weight μ this in turn implies a sound speed profile,

$$c_s \propto r^{-3/8}. \tag{33}$$

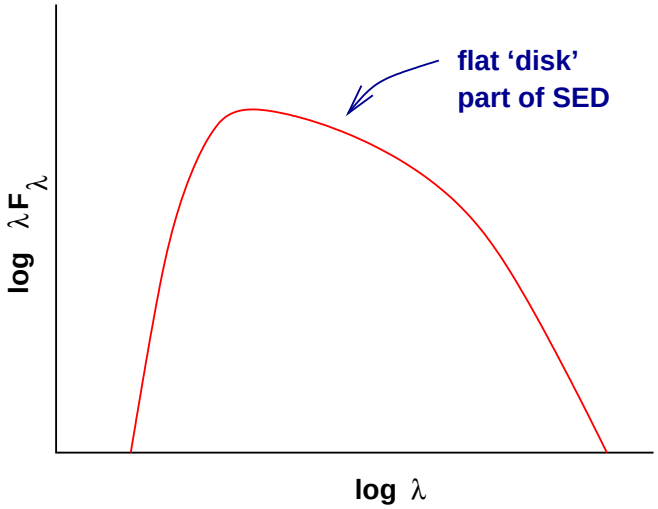

FIG. 11 Schematic disk spectrum. At short wavelengths, we see an exponential cut-off corresponding to the highest temperature annulus in the disk (normally close to or at the inner edge). At long wavelengths, there is a Rayleigh-Jeans tail reflecting the coldest material in the outer disk. At intermediate wavelengths, there is a flatter portion of the spectrum, so that the overall SED resembles a stretched blackbody.

Assuming vertical isothermality, the aspect ratio given by equation (25) is,

$$\frac{h}{r} \propto r^{1/8}, \tag{34}$$

and we predict that the disk ought to flare modestly to larger radii. If the disk does flare, then the outer regions intercept a larger fraction of stellar photons, leading to a higher temperature. As a consequence, a temperature profile $T_{\rm disk} \propto r^{-3/4}$ is probably the steepest profile we would expect to obtain for a passive disk.

3. Spectral energy distribution (SED)

Suppose that each annulus in the disk radiates as a blackbody at the local temperature $T_{\rm disk}(r)$. If the disk extends from $r_{\rm in}$ to $r_{\rm out}$, the disk spectrum is just the sum of these blackbodies weighted by the disk area,

$$F_\lambda \propto \int_{r_{\rm in}}^{r_{\rm out}} 2\pi r B_\lambda[T(r)] dr \tag{35}$$

where B_λ is the Planck function,

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5}\frac{1}{e^{hc/\lambda kT}-1}. \tag{36}$$

The behavior of the spectrum implied by equation (35) is easy to derive. At long wavelengths $\lambda \gg hc/kT(r_{\rm out})$ we recover the Rayleigh-Jeans form,

$$\lambda F_\lambda \propto \lambda^{-3} \tag{37}$$

while at short wavelengths $\lambda \ll hc/kT(r_{\rm in})$ there is an exponential cut-off that matches that of the hottest annulus in the disk,

$$\lambda F_\lambda \propto \lambda^{-4} e^{-hc/\lambda kT(r_{\rm in})}. \quad (38)$$

For intermediate wavelengths,

$$\frac{hc}{kT(r_{\rm in})} \ll \lambda \ll \frac{hc}{kT(r_{\rm out})} \quad (39)$$

the form of the spectrum can be found by substituting,

$$x \equiv \frac{hc}{\lambda kT(r_{\rm in})} \left(\frac{r}{r_{\rm in}}\right)^{3/4} \quad (40)$$

into equation (35). We then have, approximately,

$$F_\lambda \propto \lambda^{-7/3} \int_0^\infty \frac{x^{5/3} dx}{e^x - 1} \propto \lambda^{-7/3} \quad (41)$$

and so

$$\lambda F_\lambda \propto \lambda^{-4/3}. \quad (42)$$

The overall spectrum, shown schematically in Figure 11, is that of a 'stretched' blackbody (Lynden-Bell, 1969).

The SED predicted by this simple model generates an IR-excess, but with a declining SED in the mid-IR. This is too steep to match the observations of even most Class II sources.

4. Sketch of more complete models

Two additional pieces of physics need to be included when computing detailed models of the SEDs of passive disks. First, as already noted above, all reasonable disk models flare toward large r, and as a consequence intercept and reprocess a larger fraction of the stellar flux. At large radii, Kenyon & Hartmann (1987) find that consistent flared disk models approach a temperature profile,

$$T_{\rm disk} \propto r^{-1/2}, \quad (43)$$

which is much flatter than the profile derived previously. Second, the assumption that the emission from the disk can be approximated as a single blackbody is too simple. In fact, dust in the surface layers of the disk radiates at a significantly higher temperature because the dust is more efficient at absorbing short-wavelength stellar radiation than it is at emitting in the IR (Shlosman & Begelman, 1989). Dust particles of size a absorb radiation efficiently for $\lambda < 2\pi a$, but are inefficient absorbers and emitters for $\lambda > 2\pi a$ (i.e. the opacity is a declining function of wavelength). As a result, the disk absorbs stellar radiation close to the surface (where $\tau_{1\mu{\rm m}} \sim 1$), where the optical depth to emission at longer IR wavelengths $\tau_{\rm IR} \ll 1$. The surface emission comes from low optical depth, and is not at the blackbody temperature previously derived. Chiang & Goldreich (1997) showed that a relatively simple disk model made up of,

1. A hot surface dust layer that directly re-radiates half of the stellar flux
2. A cooler disk interior that reprocesses the other half of the stellar flux and re-emits it as thermal radiation

can, when combined with a flaring geometry, reproduce most SEDs quite well. A review of recent disk modeling work is given by Dullemond et al. (2007).

The above considerations are largely sufficient to understand the structure and SEDs of Class II sources. For Class I sources, however, the possible presence of an envelope (usually envisaged to comprise dust and gas that is still infalling toward the star-disk system) also needs to be considered. The reader is directed to Eisner et al. (2005) for one example of how modeling of such systems can be used to try and constrain their physical properties and evolutionary state.

C. Actively accreting disks

The radial force balance in a passive disk includes contributions from gravity, centrifugal force, and radial pressure gradients. The equation reads,

$$\frac{v_\phi^2}{r} = \frac{GM_*}{r^2} + \frac{1}{\rho}\frac{{\rm d}P}{{\rm d}r}, \quad (44)$$

where v_ϕ is the orbital velocity of the gas and P is the pressure. To estimate the magnitude of the pressure gradient term we note that,

$$\begin{aligned} \frac{1}{\rho}\frac{{\rm d}P}{{\rm d}r} &\sim -\frac{1}{\rho}\frac{P}{r} \\ &\sim -\frac{1}{\rho}\frac{\rho c_s^2}{r} \\ &\sim -\frac{GM_*}{r^2}\left(\frac{h}{r}\right)^2, \end{aligned} \quad (45)$$

where for the final step we have made use of the relation $h = c_s/\Omega$. If v_K is the Keplerian velocity at radius r, we then have that,

$$v_\phi^2 = v_K^2 \left[1 - \mathcal{O}\left(\frac{h}{r}\right)^2\right], \quad (46)$$

i.e pressure gradients make a negligible contribution to the rotation curve of gas in a geometrically thin ($h/r \ll 1$) disk[3]. To a good approximation, the specific angular

[3] This is not to say that pressure gradients are unimportant – as we will see later the small difference between v_ϕ and v_K is of critical importance for the dynamics of small rocks within the disk.

momentum of the gas within the disk is just that of a Keplerian orbit,

$$l = r^2\Omega = \sqrt{GM_* r}, \tag{47}$$

which is an **increasing function** of radius. To accrete on to the star, gas in a disk must lose angular momentum, either,

1. Via redistribution of angular momentum within the disk (normally described as being due to 'viscosity', though this is a rather loaded term).

2. Via loss of angular momentum from the star-disk system, for example in a magnetically driven disk wind.

Models in the second class – one well-known example of which is the disk wind solution described by Blandford & Payne (1982) – are, perhaps undeservedly, not widely considered (in part, because in some disk-accreting white dwarf systems observations disfavor wind models, though no such constraints exist for protoplanetary disks). Here, we will derive the equation for the time evolution of the surface density for a thin, viscous disk (Lynden-Bell & Pringle, 1974; Shakura & Sunyaev, 1973). Clear reviews of the fundamentals of accretion disk theory can be found in Pringle (1981) and in Frank, King & Raine (2002).

1. Diffusive evolution equation

Let the disk have surface density $\Sigma(r,t)$ and radial velocity $v_r(r,t)$ (defined such that $v_r < 0$ for inflow). The potential is assumed fixed so that the angular velocity $\Omega = \Omega(r)$ only. In cylindrical co-ordinates, the continuity equation for an axisymmetric flow gives (see e.g. Pringle (1981) for an elementary derivation),

$$r\frac{\partial \Sigma}{\partial t} + \frac{\partial}{\partial r}\left(r\Sigma v_r\right) = 0. \tag{48}$$

Similarly, conservation of angular momentum yields,

$$r\frac{\partial\left(\Sigma r^2\Omega\right)}{\partial t} + \frac{\partial}{\partial r}\left(r\Sigma v_r \cdot r^2\Omega\right) = \frac{1}{2\pi}\frac{\partial G}{\partial r}, \tag{49}$$

where the term on the right-hand side represents the net torque acting on the fluid due to viscous stresses. From fluid dynamics (Pringle, 1981), G is given in terms of the kinematic viscosity ν by the expression,

$$G = 2\pi r \cdot \nu\Sigma r\frac{\mathrm{d}\Omega}{\mathrm{d}r} \cdot r \tag{50}$$

where the right-hand side is the product of the circumference, the viscous force per unit length, and the level arm r. If we substitute for G, eliminate v_r between equation (48) and equation (49), and specialize to a Keplerian potential with $\Omega \propto r^{-3/2}$, we obtain the evolution equation for the surface density of a thin accretion disk in its normal form,

$$\frac{\partial \Sigma}{\partial t} = \frac{3}{r}\frac{\partial}{\partial r}\left[r^{1/2}\frac{\partial}{\partial r}\left(\nu\Sigma r^{1/2}\right)\right]. \tag{51}$$

This partial differential equation for the evolution of the surface density Σ has the form of a diffusion equation. To make that explicit, we change variables to,

$$\begin{aligned} X &\equiv 2r^{1/2} \\ f &\equiv \frac{3}{2}\Sigma X. \end{aligned} \tag{52}$$

For a constant ν, equation (51) then takes the prototypical form for a diffusion equation,

$$\frac{\partial f}{\partial t} = D\frac{\partial^2 f}{\partial X^2} \tag{53}$$

with a diffusion coefficient,

$$D = \frac{12\nu}{X^2}. \tag{54}$$

The characteristic diffusion time scale implied by equation (53) is X^2/D. Converting back to the physical variables, we find that the evolution time scale for disk of scale r with kinematic viscosity ν is,

$$\tau \simeq \frac{r^2}{\nu}. \tag{55}$$

Observations of disk evolution (for example determinations of the time scale for the secular decline in the accretion rate) can therefore be combined with estimates of the disk size to yield an estimate of the effective viscosity in the disk (Hartmann et al., 1998).

2. Solutions

In general, ν is expected to be some function of the local conditions within the disk (surface density, radius, temperature, ionization fraction etc). If ν depends on Σ, then equation (51) becomes a non-linear equation with no analytic solution (except in some special cases), while if there is more a complex dependence on the local conditions then the surface density evolution equation will often need to be solved simultaneously with an evolution equation for the central temperature (Pringle, Verbunt & Wade, 1986). Analytic solutions *are* possible, however, if ν can be written as a power-law in radius (Lynden-Bell & Pringle, 1974), and these suffice to illustrate the essential behavior implied by equation (51).

First, we describe a Green's function solution to equation (51) for the case ν = constant. Suppose that at $t = 0$, all of the gas lies in a thin ring of mass m at radius r_0,

$$\Sigma(r, t=0) = \frac{m}{2\pi r_0}\delta(r - r_0). \tag{56}$$

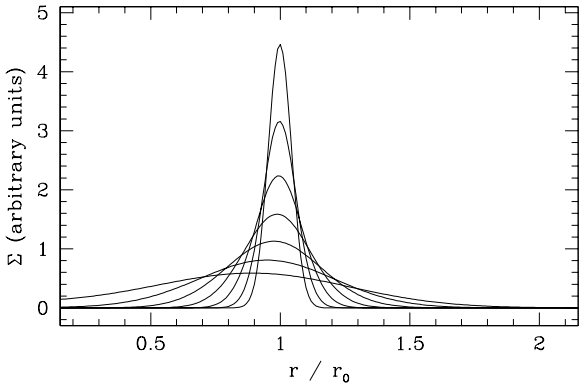

FIG. 12 The Green's function solution to the disk evolution equation with $\nu =$ constant, showing the spreading of a ring of mass initially orbiting at $r = r_0$. From top down the curves show the behavior as a function of the scaled time variable $\tau = 12\nu r_0^{-2} t$, for $\tau = 0.004$, $\tau = 0.008$, $\tau = 0.016$, $\tau = 0.032$, $\tau = 0.064$, $\tau = 0.128$, and $\tau = 0.256$.

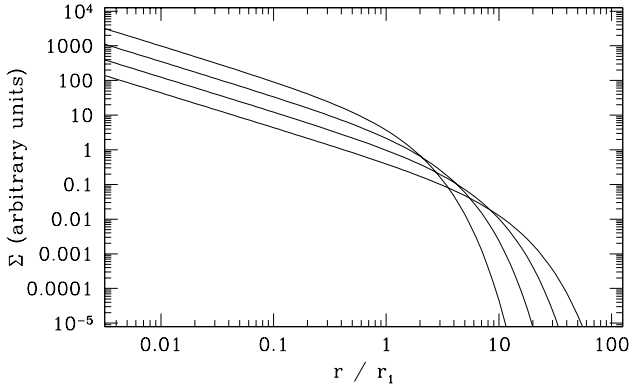

FIG. 13 The self-similar solution to the disk evolution equation is plotted for a viscosity $\nu \propto r$. The initial surface density tracks the profile for a steady-state disk (with $\Sigma \propto r^{-1}$) at small radius, before cutting off exponentially beyond $r = r_1$. The curves show the surface density at the initial value of the scaled time $T = 1$, and at subsequent times $T = 2$, $T = 4$ and $T = 8$.

One can show that the solution is then,

$$\Sigma(x,\tau) = \frac{m}{\pi r_0^2}\frac{1}{\tau}x^{-1/4}e^{-(1+x^2)/\tau}I_{1/4}\left(\frac{2x}{\tau}\right) \qquad (57)$$

where we have written the solution in terms of dimensionless variables $x \equiv r/r_0$, $\tau \equiv 12\nu r_0^{-2}t$, and $I_{1/4}$ is a modified Bessel function of the first kind.

Unless you have a special affinity for Bessel functions, this Green's function solution is not terribly transparent. The evolution it implies is shown in Figure 12. The most important features of the solution are that, as $t \to \infty$,

- The **mass** flows to $r = 0$.
- The **angular momentum**, carried by a negligible fraction of the mass, flows toward $r = \infty$.

This segregation of mass and angular momentum is a generic feature of viscous disk evolution, and is obviously relevant to the angular momentum problem of star formation.

Of greater practical utility is the self-similar solution also derived by Lynden-Bell & Pringle (1974). Consider a disk in which the viscosity can be approximated as a power-law in radius,

$$\nu \propto r^\gamma. \qquad (58)$$

Suppose that the disk at time $t = 0$ has the surface density profile corresponding to a steady-state solution (with this viscosity law) out to $r = r_1$, with an exponential cutoff at larger radii. As we will shortly show, the initial surface density then has the form,

$$\Sigma(t=0) = \frac{C}{3\pi\nu_1\tilde{r}^\gamma}\exp\left[-\tilde{r}^{(2-\gamma)}\right], \qquad (59)$$

where C is a normalization constant, $\tilde{r} \equiv r/r_1$, and $\nu_1 \equiv \nu(r_1)$. The self-similar solution is then,

$$\Sigma(\tilde{r},T) = \frac{C}{3\pi\nu_1\tilde{r}^\gamma}T^{-(5/2-\gamma)/(2-\gamma)}\exp\left[-\frac{\tilde{r}^{(2-\gamma)}}{T}\right], \qquad (60)$$

where,

$$\begin{aligned} T &\equiv \frac{t}{t_s} + 1 \\ t_s &\equiv \frac{1}{3(2-\gamma)^2}\frac{r_1^2}{\nu_1}. \end{aligned} \qquad (61)$$

This solution is plotted in Figure 13. Over time, the disk mass decreases while the characteristic scale of the disk (initially r_1) expands to conserve angular momentum. This solution is quite useful both for studying evolving disks analytically, and for comparing observations of disk masses, accretion rates or radii with theory (Hartmann et al., 1998).

A steady-state solution for the radial dependence of the surface density can be derived by setting $\partial/\partial t = 0$ and integrating the angular momentum conservation equation (49). This yields,

$$\Sigma r^3 \Omega v_r = \nu\Sigma r^3\frac{\mathrm{d}\Omega}{\mathrm{d}r} + \text{constant}. \qquad (62)$$

Noting that the mass accretion rate $\dot{M} = -2\pi r\Sigma v_r$ we have,

$$-\frac{\dot{M}}{2\pi}r^2\Omega = \nu\Sigma r^3\frac{\mathrm{d}\Omega}{\mathrm{d}r} + \text{constant}. \qquad (63)$$

To determine the constant of integration, we note that the torque within the disk vanishes if $\mathrm{d}\Omega/\mathrm{d}r = 0$. At

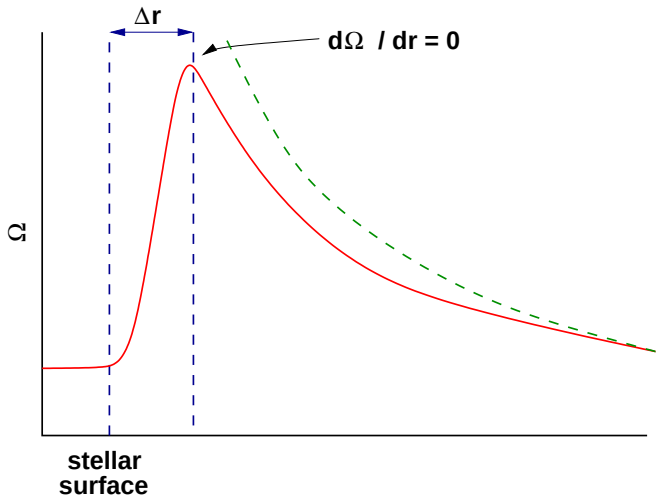

FIG. 14 Schematic depiction of the angular velocity $\Omega(r)$ for a slowly rotating star surrounded by a thin accretion disk that extends to the stellar equator. At large radii in the disk, the angular velocity has the normal Keplerian form $\Omega^{-3/2}$, shown as the dashed green curve. To match smoothly on to the star, the angular velocity must turn over at smaller radii in a transition zone known as the boundary layer. The existence of a boundary layer implies that at some radius $\mathrm{d}\Omega/\mathrm{d}r = 0$, at which point the viscous stress vanishes.

such a location, the constant can be evaluated and is just proportional to the local flux of angular momentum

$$\text{constant} \propto \dot{M} r^2 \Omega. \tag{64}$$

Usually this is determined at the inner boundary. A particularly simple example is the case of a disk that extends to the equator of a slowly rotating star. This case is illustrated in Figure 14. In order for there to be a transition between the Keplerian angular velocity profile in the disk and the much smaller angular velocity at the stellar surface there must be a maximum in Ω at some radius $r_* + \Delta r$. Elementary arguments (Pringle, 1977) – which may fail at the very high accretion rates of FU Orionis objects (Popham et al., 1993) but which are probably reliable otherwise – suggest that $\Delta r \ll r_*$, so that the transition occurs in a narrow *boundary layer* close to the stellar surface[4]. The constant can then be evaluated as,

$$\text{constant} \approx -\frac{\dot{M}}{2\pi} r_*^2 \sqrt{\frac{GM_*}{r_*^3}}, \tag{65}$$

[4] The physics of the boundary layer itself presents interesting complications, since the boundary layer is a region of strong shear that is stable against the magnetorotational instabilities that we will argue later are critical for transporting angular momentum within disks. Pringle (1989), Armitage (2002) and Pessah, Chan & Psaltis (2008) have studied the role of magnetic fields within the boundary layer.

and equation (63) becomes,

$$\nu\Sigma = \frac{\dot{M}}{3\pi}\left(1 - \sqrt{\frac{r_*}{r}}\right). \tag{66}$$

Given a viscosity, this equation defines the steady-state surface density profile for a disk with an accretion rate $\dot{M}$. Away from the boundaries, $\Sigma(r) \propto \nu^{-1}$.

The inner boundary condition which leads to equation (66) is described as a *zero-torque* boundary condition. As noted, zero-torque conditions are physically realized in the case where there is a boundary layer between the star and its disk. This is not, however, the case in most Classical T Tauri stars. Observational evidence suggests (Bouvier et al., 2007) that in accreting T Tauri stars the stellar magnetosphere disrupts the inner accretion disk, leading to a *magnetospheric* mode of accretion in which gas becomes tied to stellar field lines and falls ballistically on to the stellar surface (Königl, 1991). The magnetic coupling between the star and its disk allows for angular momentum exchange, modifies the steady-state surface density profile close to the inner truncation radius, and may allow the star to rotate more slowly than would otherwise be the case (Armitage & Clarke, 1996; Collier Cameron & Campbell, 1993). Whether such 'disk-locking' actually regulates the spin of young stars remains a matter of debate, however, and both theoretical and observational studies have returned somewhat ambiguous results (Herbst & Mundt, 2005; Matt & Pudritz, 2005; Rebull et al., 2006).

3. Temperature profile

Following the approach of Frank, King & Raine (2002), we derive the radial dependence of the effective temperature of an actively accreting disk by considering the net torque on a ring of width Δr. This torque – $(\partial G/\partial r)\Delta r$ – does work at a rate,

$$\Omega \frac{\partial G}{\partial r}\Delta r \equiv \left[\frac{\partial}{\partial r}\left(G\Omega\right) - G\Omega'\right]\Delta r \tag{67}$$

where $\Omega' = \mathrm{d}\Omega/\mathrm{d}r$. Written this way, we note that if we consider the whole disk (by integrating over r) the first term on the right-hand-side is determined solely by the boundary values of $G\Omega$. We therefore identify this term with the *transport* of energy, associated with the viscous torque, through the annulus. The second term, on the other hand, represents the rate of loss of energy to the gas. We assume that this is ultimately converted into heat and radiated, so that the dissipation rate per unit surface area of the disk (allowing that the disk has two sides) is,

$$D(r) = \frac{G\Omega'}{4\pi r} = \frac{9}{8}\nu\Sigma\Omega^2, \tag{68}$$

where we have assumed a Keplerian angular velocity profile. For blackbody emission $D(r) = \sigma T_{\rm disk}^4$. Substituting

for Ω, and for $\nu\Sigma$ using the steady-state solution given by equation (66), we obtain,

$$T_{\rm disk}^4 = \frac{3GM_*\dot{M}}{8\pi\sigma r^3}\left(1-\sqrt{\frac{r_*}{r}}\right). \quad (69)$$

We note that,

- Away from the boundaries ($r \gg r_*$), the temperature profile of an actively accreting disk is $T_{\rm disk} \propto r^{-3/4}$. This has the same form as for a passive disk given by equation (32).
- The temperature profile does *not* depend upon the viscosity. This is an attractive feature of the theory given current uncertainties regarding the origin and efficiency of disk angular momentum transport though, on the flip side, it eliminates many possible routes to learning about the physics underlying ν via observations of steady-disks.

Substituting a representative value for the accretion rate of $\dot{M} = 10^{-7}\ M_\odot {\rm yr}^{-1}$, we obtain for a Solar mass star at 1 AU an effective temperature $T_{\rm disk} = 150$ K. This is the *surface* temperature, as we will show shortly the central temperature is predicted to be substantially higher.

4. Origin of angular momentum transport

Molecular viscosity is negligible in protoplanetary disks. For a gas in which the mean free path is λ, the viscosity

$$\nu \sim \lambda c_s \quad (70)$$

where c_s is the sound speed. In turn, the mean free path is given by $\lambda = 1/n\sigma$, where n is the number density of molecules with cross-section for collision σ. These quantities are readily estimated. For example, consider a protoplanetary disk with $\Sigma = 10^3$ gcm^{-2} and $h/r = 0.05$ at 1 AU. The midplane density is of the order of $n \sim \Sigma/2m_H h \sim 4\times10^{14}$ cm^{-3}, while the sound speed implied by the specified h/r is $c_s \approx 1.5\times10^5$ cms^{-1}. The collision cross-section of a hydrogen molecule is of the order of (Chapman & Cowling, 1970),

$$\sigma \sim 2\times10^{-15}\ {\rm cm}^2, \quad (71)$$

and hence we estimate,

$$\begin{aligned}\lambda &\sim 1\ {\rm cm}\\ \nu &\sim 2\times10^5\ {\rm cm}^2{\rm s}^{-1}.\end{aligned} \quad (72)$$

The implied disk evolution time scale $\tau \simeq r^2/\nu$ then works out to be of the order of 10^{13} yr – at least 10^6 times too slow to account for observed disk evolution.

In a classic paper, Shakura & Sunyaev (1973) noted that turbulence within the disk can provide an effective viscosity that greatly exceeds molecular viscosity. For isotropic turbulence, the maximum scale of turbulent cells within the disk will be of the same order as the vertical scale height h, while the maximum velocity of turbulent motions relative to the mean flow is comparable to the sound speed c_s (any larger velocity would lead to shocks and rapid dissipation of turbulent kinetic energy into heat). Motivated by such considerations, Shakura & Sunyaev (1973) proposed a parameterization,

$$\nu = \alpha c_s h \quad (73)$$

where α is a dimensionless parameter that measures how efficient the turbulence is at creating angular momentum transport. We note at the outset that the existence of turbulence within the disk does not, a priori, guarantee that the outward angular momentum transport necessary to drive accretion will occur.

In the standard theory of so-called 'α-disks', α is treated as a constant. If this is done, it is possible to solve analytically for the approximate vertical structure of an actively accreting disk and derive a scaling for ν as a function of r, Σ and α. Details of this procedure can be found in Frank, King & Raine (2002). Combining the known functional form for ν with the disk evolution equation (51) then yields a full theory for the predicted time dependence of the disk, with the only unknown being the appropriate value for α. This is all very well, but there is no physical reason to assume that α is a constant, and if instead α is regarded as a free *function* then much of the beguiling simplicity of the theory is lost. α-disk models should therefore be regarded as illustrative rather than definitive predictions for the evolution of the disk.

It is straightforward to estimate how large α must be to account for the observed evolution of protoplanetary disks. Suppose, for example, that the evolution time scale at 50 AU is 1 Myr. Then starting from the α-prescription (equation 73), and noting that $c_s \simeq h\Omega$, the evolution time scale becomes,

$$\tau = \frac{r^2}{\nu} = \left(\frac{h}{r}\right)^{-2}\frac{1}{\alpha\Omega}. \quad (74)$$

Substituting for τ and r, and assuming again that $h/r = 0.05$, we obtain an estimate for $\alpha \simeq 0.02$. This is fairly typical – observational attempts to constrain α on large scales in protoplanetary disks (none of which are much more sophisticated than our crude estimate) tend to result in estimates that are around 10^{-2} (Hartmann et al., 1998)[5]. These values are an order of magnitude smaller

[5] An important exception is modeling of the large-amplitude eruptive events known as FU Orionis outbursts (Hartmann & Kenyon, 1995), which, if interpreted as self-regulated thermal instabilities, require small values of α of the order of 10^{-3} or less (Bell & Lin, 1994). My own opinion is that these values are unreasonably small, and that FU Orionis events are instead externally triggered thermal instabilities that originate further out in the disk.

than the values of α derived from the modeling of dwarf nova outbursts that occur in accretion disks around white dwarfs (Cannizzo, 1993).

Despite substantial recent progress, significant uncertainties persist as to the physical origin and properties of angular momentum transport within protoplanetary disks. The Reynolds number of the flow in the disk,

$$\mathrm{Re} \equiv \frac{UL}{\nu} \tag{75}$$

where U is a characteristic velocity and L a characteristic size scale, is extremely large (of the order of 10^{14} using the parameters that we previously estimated when considering the magnitude of molecular viscosity). Terrestrial flows typically develop turbulence above a critical Reynolds number of the order of 10^4, so one's intuition would suggest that disk flows would surely be highly turbulent due to purely hydrodynamic effects. Detailed studies, however, do *not* support this conclusion. We first note that the condition for linear hydrodynamic stability in a differentially rotating fluid (the Rayleigh criterion) is that the specific angular momentum increase outward,

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\Omega\right) > 0. \tag{76}$$

In a Keplerian disk, $r^2\Omega \propto r^{1/2}$, so the flow is always linearly stable.

A vast body of literature has investigated the possibility of non-linear instabilities that might lead to turbulence within accretion disks. To date, there is no compelling evidence that such instabilities exist, and numerical simulations find that hydrodynamic perturbations in a Keplerian disk flow – which can in some circumstances exhibit substantial *transient* growth (Afshordi, Mukhopadhyay & Narayan, 2005; Ioannou & Kakouris, 2001) – ultimately decay (Balbus & Hawley, 2006; Balbus, Hawley & Stone, 1996; Shen, Stone & Gardiner, 2006). Experiments yield a similar result (Ji et al., 2006). Although not all workers in the field would concur, the conclusion one draws from this is that in a *model system* – one in which the disk has simple thermodynamics (e.g. isothermal), is initially in an equilibrium state and is isolated from perturbing influences – hydrodynamic turbulence would not evolve spontaneously.

Real disks are more complicated, and within the complications may lurk loopholes that allow for purely hydrodynamic routes to angular momentum transport. To start with, actively accreting disks dissipate energy near the disk midplane that must be transported vertically through the disk before it can be radiated. If this transport occurs via convection rather than radiative diffusion the result will be convective turbulence. If this turbulence transported angular momentum outward there would be the possibility of an internally self-consistent cycle: convection generates turbulence, turbulence generates a non-zero effective α, and the resulting viscosity leads to inflow which heats the disk sustaining the convection. This idea, which was popular in the 1980s, does not work. Convection acts to transport angular momentum *inward*, and as a result cannot be self-sustaining in the absence of other sources of transport (Ryu & Goodman, 1992; Stone & Balbus, 1996).

Another oft-discussed possibility is that *vortices* (i.e. hydrodynamic structures of the type exemplified by Jupiter's Great Red Spot with non-zero vorticity $\omega \equiv \nabla \times \mathbf{v}$) play a role. Vortices are of extra interest because they might in principle help solve two problems, first by transporting angular momentum and second by trapping dust within their cores and accelerating the formation of larger solid bodies (Barge & Sommeria, 1995). Indeed in strictly two-dimensional disk models vortices are found to be long lived and able to transport angular momentum (Godon & Livio, 1999; Johnson & Gammie, 2005). In three dimensions, however, they are found to be subject to disruptive instabilities (Barranco & Marcus, 2005; Lesur & Papaloizou, 2009; Lithwick, 2009; Shen, Stone & Gardiner, 2006). This probably means that vortices are not important if they arise solely as a consequence of perturbations within the collapsing cloud that formed the disk. For vortices to matter, there must instead be a long-lived instability that continuously generates such structures. Work into such ideas continues.

Self-gravity provides an antidote to this litany of negative results. A sufficiently massive disk is unstable (Toomre, 1964) to the development of trailing spiral arms, which act to transport angular momentum outward. We will discuss the physics underlying this instability later in the context of the disk instability model for giant planet formation, but for now we note that instability occurs when, roughly,

$$\frac{M_{\mathrm{disk}}}{M_*} > \frac{h}{r}. \tag{77}$$

Self-gravity could therefore play a role in protoplanetary disks at early epochs – when the disk may well be massive enough – but will not be important at late times when $M_{\mathrm{disk}} \ll M_*$.

Returning to non-self-gravitating disks, the hydrodynamic stability condition given by equation (76) is dramatically altered in the presence of even a weak magnetic field. Whereas a hydrodynamic flow is stable provided only that the specific angular momentum increase outward, a magnetohydrodynamic (MHD) flow requires that the angular velocity itself be an increasing function of radius,

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(\Omega^2\right) > 0, \tag{78}$$

in order to be stable (Balbus & Hawley, 1991; Chandrasekhar, 1961; Velikhov, 1959)[6]. This condition is *not* satisfied in Keplerian disks. As a consequence, in

[6] The significance of Chandrasekhar's result for the origin of turbulence within the protoplanetary disk was appreciated by Safronov

ideal (zero diffusivity) MHD an arbitrarily weak magnetic field suffices to render a Keplerian disk linearly unstable, with perturbations growing exponentially on a dynamical time scale. A comprehensive review of the physics of this instability – called the magnetorotational (MRI) or Balbus-Hawley instability – is given by Balbus & Hawley (1998). The MRI is a linear instability that leads to self-sustaining turbulence within sufficiently well-ionized accretion disks (Brandenburg et al., 1995; Stone et al., 1996). It transports angular momentum outward, as is required to allow mass to flow inward and liberate gravitational potential energy. The magnitude of the effective α generated by the MRI is not terribly well determined from current simulations, but it certainly very plausible that values of $\alpha \sim 10^{-2}$ could be attained that would match the values inferred empirically for protoplanetary disks.

5. Layered disks

The MRI is widely considered to be the most important (and possibly the only) source of angular momentum transport in accretion disks around white dwarfs, neutron stars and black holes. However in protoplanetary disks an interesting complication arises because the low ionization fraction leads to a finite conductivity. Resistivity (or other departures from ideal MHD due to ambipolar diffusion and the Hall effect) can then potentially damp the MRI and suppress turbulence and resulting angular momentum transport. The linear physics in this regime has been analyzed in numerous papers, including works by Blaes & Balbus (1994), Desch (2004) and Salmeron & Wardle (2005). A recent review by Balbus (2009) provides a good entry into this literature. Here, I summarize a broad picture based on simple physics that was proposed by Gammie (1996) as a model for how the MRI might operate in protoplanetary disks. I stress that, in contrast to the prior discussion of the basic principles of the MRI, this application to the protoplanetary disk remains somewhat speculative.

Following Gammie (1996), we begin by noting that in the presence of resistivity (assumed to be the most important non-ideal MHD effect affecting the growth of the MRI) the magnetic field obeys the usual induction equation,

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) - \nabla \times (\eta \nabla \times \mathbf{B}) \,, \qquad (79)$$

where η is the magnetic diffusivity. In turn, η can be written in terms of the electron fraction $x \equiv n_e/n_H$ via,

$$\eta = 6.5 \times 10^{-3} x^{-1} \text{ cm}^2\text{s}^{-1}. \qquad (80)$$

Our goal is to determine the minimum x for which the MRI will be able to operate despite the damping caused by the diffusivity. To do this, we note that resistivity damps small scales most easily. We therefore consider the largest disk scale $l = h$, and equate the MRI growth time scale (Balbus & Hawley, 1998),

$$\tau_{\rm MRI} \sim \frac{h}{v_A} \qquad (81)$$

where $v_A = \sqrt{B^2/(4\pi\rho)}$ is the Alfvén speed, with the damping time scale,

$$\tau_{\rm damp} \sim \frac{h^2}{\eta}. \qquad (82)$$

This yields a simple criterion for the MRI to operate:

$$\eta < h v_A. \qquad (83)$$

It remains to estimate appropriate values for h and v_A. For a crude estimate, we can guess that at 1 AU $h \sim 10^{12}$ cm and that $v_A \sim c_s \sim 10^5$ cms^{-1} (more accurately, $v_A \sim \alpha^{1/2} c_s$ in MRI turbulence that yields an effective Shakura-Sunyaev α). In that case the limit becomes $\eta < 10^{17}$ cm^{2}s^{-1} which translates into a minimum electron fraction,

$$x > 10^{-13}, \qquad (84)$$

which is more or less the 'right' value derived from more rigorous analyses (Balbus & Hawley, 1998; Gammie, 1996). The most important thing to note is that this is an *extremely* small electron fraction! The linear MRI growth rate is so large that a tiny electron fraction couples the gas to the magnetic field well enough that the MRI can overcome the stabilizing influence of diffusion.

Although only a small degree of ionization is required for the MRI to work, there may be regions in the protoplanetary disk where even $x \sim 10^{-13}$ is not attained. Considering first thermal ionization processes calculations of collisional ionization by Umebayashi (1983) show that ionization of the alkali metals suffices to drive $x > 10^{-13}$. This, however, requires temperatures $T \approx 10^3$ K and above. Only the *very* innermost disk – within a few tenths of an AU of the star – will therefore be able to sustain the MRI as a result of purely thermal ionization.

At larger disk radii the ionization fraction will be determined by a balance between non-thermal ionization processes and recombination. Two sources of ionization are of particular interest,

- Ionization by stellar X-rays. T Tauri stars are observed to be strong X-ray sources (Feigelson et al., 2007), and the harder components of the emission are unquestionably penetrating enough to ionize a fraction of the column through the disk.

(1969), who noted that the MHD stability criterion does not reduce to the Rayleigh criterion as the magnetic field tends toward zero, and that 'for a weak magnetic field the cloud should be less stable than we found earlier in the absence of the field'. Safronov then, however, dismisses the MRI on the (incorrect) grounds that the instability requires that the viscosity and diffusivity are identically zero. The importance of the MRI for accretion disks was only appreciated more than 20 years later by Balbus & Hawley.

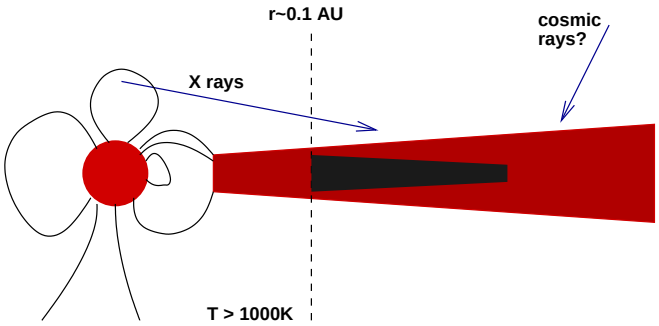

FIG. 15 Schematic illustration of the layered disk model for protoplanetary disks. In this model, the innermost regions of the disk are hot enough that thermal ionization suffices to couple the magnetic field to the gas and allow the MRI to operate. At large radii, some combination of stellar X-rays and cosmic rays penetrate the entire thickness of the disk and provide the necessary ionization. At intermediate radii, it is hypothesized that accretion occurs primarily in an active surface layer of whose column depth is set by the penetration power of the dominant non-thermal ionization process (either cosmic rays, as in the original version of this model proposed by Gammie (1996), or stellar X-rays in some more recent versions). Underlying this layer is a magnetically inactive 'dead zone' in which the MRI is suppressed and the turbulence is either suppressed entirely or, at the least, greatly weakened.

- Ionization by cosmic rays. Cosmic rays have a stopping length that is of the order of $\Sigma_{\rm layer} = 100$ gcm^{-2} (Umebayashi & Nakano, 1981). *If present* they are therefore likely to be more penetrating and important than X-rays. It is possible (or perhaps probable), however, that the disk may be screened from the interstellar cosmic ray flux by the magnetized plasma flowing away from the system in a wind.

If no other sources of ionization exist, then at radii where the disk is simultaneously too cool to be collisionally ionized, and dense enough that the interior is shielded from X-rays (or cosmic rays, if present) we might expect a novel structure in which the disk near the midplane is quiescent (a 'dead zone') and only a surface layer is magnetically active and supporting accretion (Gammie, 1996). Such a *layered* disk model is depicted in Figure 15.

Layered disk models are qualitatively distinct from ordinary fully viscous disks in two critical respects. First, the mass flux through the active layer is set (by analogy to equation 66) by the product of the viscosity and column $\nu_{\rm layer}\Sigma_{\rm layer}$ in the active layer, which is *independent* of the total column density at that radius. The amount of mass that the active layer can support is predicted to be a decreasing function of radius, so gas flowing inward from large radii 'drops out' of the flow and accumulates in the dead zone. Second, layered models predict that the midplane of the disk ought to be almost quiescent at precisely those radii of greatest interest for planet formation. This has important implications for the settling of dust, for the subsequent growth of planetesimals, and for the migration of low mass planets (Matsumura & Pudritz, 2005). There are also possible implications for variability, since the growing dead zone provides a reservoir of gas at small radii which could in principle be heated and activated leading to a burst of accretion (Armitage, Livio & Pringle, 2001; Zhu, Hartmann & Gammie, 2009).

Given these intriguing possibilities, do layered disks actually exist? Currently there are few observational constraints, though theoretically the idea remains plausible. Numerical simulations (Sano & Stone, 2002) have confirmed that when the magnetic Reynolds number (evaluated on the relevant scale for a disk),

$$\mathrm{Re}_M \equiv \frac{v_A^2}{\eta\Omega} \tag{85}$$

falls below a critical value that is in the range of 1 – 30, turbulence driven by the MRI is suppressed. This is broadly consistent (i.e. it yields a similar critical electron fraction) with the simple arguments outlined above. However, very little *energy* would be required to maintain a high enough level of ionization for the MRI to operate, and it has been suggested that enough power from the turbulence could couple into the ionization to sustain an MRI-active disk (Inutsuka & Sano, 2005). Another possibility — which receives support from recent numerical simulations by Turner, Sano & Dziourkevitch (2007) — is that turbulent mixing might be able to transport enough charge from the surface layers to the midplane to allow non-zero transport to occur there. Given these theoretical uncertainties, the question of whether protoplanetary disks have a structure akin to that shown in Figure 15 remains open. A critical question is whether the rate of recombination within the gas is high enough to mop up free electrons efficiently. Recombination depends upon difficult aspects of the chemistry and dust physics within the disk (e.g. how many small dust particles are present close to the disk surface, what is the abundance of metal ions in the gas phase?), and the resulting uncertainties are at least as large as those arising from the nature of the ionization or from the basic physics of the MRI under non-ideal conditions.

6. Disk dispersal

Loss of the gaseous component of protoplanetary disks sets a time limit for the completion of gas giant formation, and will affect the environment for terrestrial planet formation as well. The self-similar solution for the evolution of a viscous disk (equation 60) predicts that the surface density and accretion rate decline as power-laws at late times, and hence that the transition between disk and diskless states should be gradual. Observationally, this is *not* what is observed. Young stars with properties intermediate between CTTS and WTTS (so called 'transition objects') are relatively uncommon, and make up of the order of 10% of the total population of stars that display at least one signature of a circumstellar disk.

From these statistics, one infers that the time scale for clearing the disk is short – of the order of 10^5 yr (Simon & Prato, 1995; Wolk & Walter, 1996).

A clue to the mechanism that may drive disk dispersal was provided by HST observations of low mass stars exposed to the strong ionizing flux produced by massive stars in the core of the Orion Nebula's Trapezium cluster (O'Dell, Wen & Hu, 1993). The images reveal tadpole-shaped nebulae surrounding young stars with circumstellar disks, which are interpreted as the signature of *photoevaporation* and escape of disk gas as a result of illumination by external ionizing radiation (Johnstone, Hollenbach & Bally, 1998). The physics underlying this process is relatively simple (Bally & Scoville, 1982; Hollenbach et al., 1994; Shu, Johnstone & Hollenbach, 1993), and is closely related to the well-studied problem of Compton heated winds from accretion disks in Active Galactic Nuclei (Begelman, McKee & Shields, 1983). Extreme ultraviolet (EUV) photons with $E > 13.6$ eV ionize and heat a surface layer of the disk, raising it to a temperature $T \simeq 10^4$ K characteristic of an HII region. The sound speed in the photoionized gas is $c_s \simeq 10$ kms^{-1}. Outside a critical radius r_g, given by,

$$r_g = \frac{GM_*}{c_s^2} \tag{86}$$

the sound speed in the hot gas exceeds the local Keplerian speed. The gas is then unbound, and flows away from the disk as a thermal wind. For a Solar mass star, r_g as estimated by equation (86) is at 9 AU.

The same basic process can occur regardless of whether the EUV flux arises from an external source, such as a massive star in a cluster, or from the central star itself. In the typical star formation environment (Lada & Lada, 2003), however, most low mass stars receive too low a dose of EUV radiation from external sources to destroy their disks (Adams et al., 2006). Photoevaporation due to radiation from the central star is therefore likely to be necessary for disk dispersal. In this regime, Hollenbach et al. (1994) derived an estimate for the mass loss rate due to photoevaporation,

$$\dot{M}_{\rm wind} \simeq 4 \times 10^{-10} \left(\frac{\Phi}{10^{41}\ {\rm s}^{-1}}\right)^{1/2} \left(\frac{M_*}{M_\odot}\right)^{1/2} M_\odot {\rm yr}^{-1} \tag{87}$$

where Φ is the stellar ionizing flux. Most of the wind mass loss is predicted to originate close to r_g, with a radial dependence of the mass loss given by $\dot{\Sigma} \propto r^{-5/2}$. Numerical hydrodynamic simulations by Font et al. (2004) largely confirm this basic picture, although in detail it is found both that mass is lost for radii $r < r_g$ and that the integrated mass loss is a factor of a few smaller than that predicted by the above equation.

The combination of a photoevaporative wind and viscous disk evolution leads to rapid disk dispersal (Clarke, Gendrin & Sotomayor, 2001). Calculations by Alexander, Clarke & Pringle (2006) suggest a three-stage scenario depicted schematically in Figure 16,

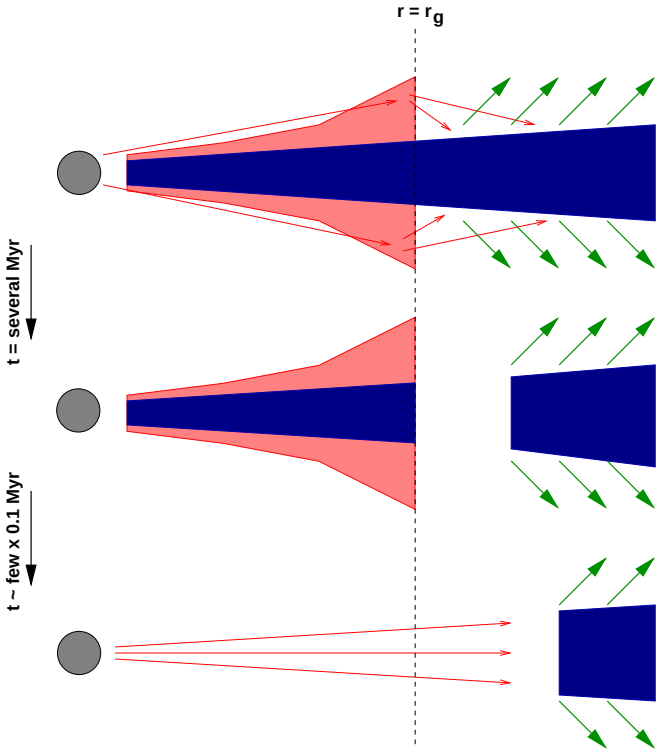

FIG. 16 Schematic depiction of how photoevaporation driven by a central source of UV radiation is predicted to disperse the protoplanetary disk. In the initial phase, UV radiation (shown as the red arrows) ionizes the surface of the disk, producing a vertically extended bound atmosphere for $r < r_g$ and mass loss in a thermal wind for $r > r_g$. The ionizing flux that photoevaporates the outer disk arises primarily from stellar photons scattered by the atmosphere at small radii (the 'diffuse field'). After several Myr, the disk accretion rate drops to a value that is of the same order as the wind mass loss rate. At this point, the wind opens up a gap in the disk close to r_g, cutting off the inner disk from resupply by the disk further out. The inner disk then drains rapidly on to the star – producing an inner hole – and the direct UV flux from the star photoevaporates the outer region.

- Initially $\dot{M} \gg \dot{M}_{\rm wind}$. The wind mass loss has negligible effect on the disk, which evolves in a similar way to an ordinary viscous model. The mass accretion rate and surface density gradually drop on the viscous time scale of the entire disk (determined at large radii), which is of the order of a Myr.

- After a few Myr, the accretion rate drops sufficiently so that $\dot{M} \sim \dot{M}_{\rm wind}$. The wind is then strong enough to dominate the disk surface density evolution near r_g, opening up a gap in the disk and cutting off the inner disk from resupply by gas flowing in from the reservoir at larger radii. The inner disk then drains on to the star on its own (short) viscous time scale, which can be of the order of 10^5 yr or less.

- Once the inner disk has drained, the remaining gas

in the outer disk is directly illuminated by UV radiation from the star (previously, the dominant flux was photons scattered on to the outer disk from a bound atmosphere at smaller radii). This radiation rapidly burns through the outer disk removing all remaining gas.

The primary source of uncertainty in these models is the origin and magnitude of the stellar ionizing flux. There are few constraints on Φ for Solar mass T Tauri stars (Alexander, Clarke & Pringle, 2005), and essentially no information on any scaling with stellar mass.

To date EUV-driven photoevaporation models for disk evolution have received the most theoretical attention. This, however, is primarily because the physics of EUV-ionized gas is particularly easy to calculate. Qualitatively similar flows can be driven by softer FUV radiation (6 eV $< E <$ 13.6 eV), which suffices to dissociate H_2 molecules and drives evaporative flow from the outer disk where the escape velocity is smaller. The detailed physics of such flows – which resemble photodissociation regions rather than HII regions – is harder to calculate because the temperature of the heated gas is determined by a balance between grain photoelectric heating and cooling by both atomic and molecular lines (Adams et al., 2006; Gorti & Hollenbach, 2009). Harder X-ray photons can also be important, with recent work suggesting that X-rays may in fact dominate the total photoevaporative mass loss rate for protoplanetary disks (Ercolano, Clarke & Drake, 2009).

D. The condensation sequence

In an actively accreting disk, there must be a temperature gradient in z in order for energy to be transported from the dense midplane where it is probably liberated to the photosphere where it is radiated (note that for a thin disk with $h/r \ll 1$ gradients in z will dominate over radial gradients, which can consistently be ignored). A simple application of the theory of radiative transport in plane-parallel media (Rybicki & Lightman, 1979) allows us to derive the relation between the central disk temperature $T_{\rm c}$ and the effective disk temperature $T_{\rm disk}$.

To proceed, we define the optical depth to the disk midplane,

$$\tau = \frac{1}{2}\kappa_R \Sigma, \tag{88}$$

where κ_R is the Rosseland mean opacity and Σ is the disk surface density. The vertical density profile of the disk is $\rho(z)$. If the vertical energy transport occurs via radiative diffusion (in some regions convection may also be important), then for $\tau \gg 1$ the vertical energy flux $F(z)$ is given by the equation of radiative diffusion,

$$F(z) = -\frac{16\sigma T^3}{3\kappa_R \rho}\frac{{\rm d}T}{{\rm d}z}. \tag{89}$$

Let us assume for simplicity that *all* the energy dissipation occurs at $z = 0$[7]. In that case $F(z) = \sigma T_{\rm disk}^4$ is a constant with height. Integrating assuming that the opacity is a constant,

$$\begin{aligned} -\frac{16\sigma}{3\kappa_R}\int_{T_{\rm c}}^{T_{\rm disk}} T^3 dT &= \sigma T_{\rm disk}^4 \int_0^z \rho(z')dz' \\ -\frac{16\sigma}{3\kappa_R}\left[\frac{T^4}{4}\right]_{T_{\rm c}}^{T_{\rm disk}} &= T_{\rm disk}^4 \frac{\Sigma}{2}, \end{aligned} \tag{90}$$

where for the final equality we have used the fact that for $\tau \gg 1$ almost all of the disk gas lies below the photosphere. For large τ we expect that $T_{\rm c}^4 \gg T_{\rm disk}^4$, and the equation simplifies to,

$$\frac{T_{\rm c}^4}{T_{\rm disk}^4} \simeq \frac{3}{4}\tau. \tag{91}$$

The implication of this result is that active disks with large optical depths are substantially hotter at the midplane than at the surface. For example, if $\tau = 10^2$ to the thermal radiation emitted by the disk at some radius then $T_{\rm c} \approx 3T_{\rm disk}$. This is important since it is the *central* temperature that determines the sound speed that enters into the viscosity (equation 73), and it is also the central temperature that determines which ices or minerals can be present. Relatively modest levels of accretion can thus affect the thermal structure of the disk substantially.

For both terrestrial planet formation, and gas giant planet formation if it occurs via the core accretion mechanism, the evolution of the trace solid component of the disk is of great interest. The gas that forms the protoplanetary disk will contain interstellar dust grains made up of a mixture of silicates, graphite and polycyclic aromatic hydrocarbons (PAHs). In the interstellar medium, measurements of the wavelength dependence of extinction can be fit by assuming that the dust grains follow a power-law size distribution (Mathis, Rumpl & Nordsieck, 1977),

$$n(a) \propto a^{-3.5} \tag{92}$$

where a is the grain size (assumed to be spherical) and the distribution extends from 0.005 μm to about 1 μm. This distribution is generally assumed to be the starting point for further evolution within the denser conditions prevailing within the disk. In the hottest, inner regions of the disk the central temperature can be high enough to destroy the grains (1000 K to 2000 K, depending on whether the grains are made of carbon or silicate). The resulting absence of dust very close to the star constitutes one of the main arguments against an in situ origin

[7] Note that for disks in which the MRI is active numerical simulations by Miller & Stone (2000) suggest that a significant fraction of the liberated energy is transported to high z and dissipated at much smaller optical depths, possibly forming a hot 'corona'.

TABLE II Condensation temperatures for selected materials

T	Material
1680 K	Al_2O_3
1590 K	$CaTiO_3$
1400 K	$MgAl_2O_4$
1350 K	Mg_2SiO_4, iron alloys
370 K	Fe_3O_4
180 K	water ice
130 K	$NH_3 \cdot H_2O$
40 K – 80 K	methane, methane ices
50 K	argon

for hot Jupiters, but the dust destruction radius is sufficiently close in (normally substantially less than 1 AU) that it rarely impacts either terrestrial or, especially, giant planet formation. It is, however, important observationally, since once the dust is destroyed the remaining gas phase opacity is greatly reduced. There will therefore be an opacity 'hole' in the inner disk even if gas is present there.

If the gas that makes up the protoplanetary disk has a known elemental composition (for example that of the Sun), then it is a well defined problem (for a chemist!) to calculate the *most thermodynamically stable* mix of chemical species at any given pressure and temperature. The abundance of different minerals and ices within the disk will follow this condensation sequence provided that there is sufficient time for chemical reactions to reach equilibrium – this may be a reasonable assumption in the hot inner disk but deviations will occur due to slow chemical reactions in the cool outer disk and radial drift of both gas and solids. The equilibrium mix depends more strongly on temperature than on pressure, so we can roughly map the condensation sequence into a predicted variation of disk composition with radius. Table II, adapted from Lodders (2003), lists characteristic temperatures below which different ices and minerals are predicted to be dominant. Of these, by far the most important is the temperature below which water ice can be present – this is 180 K at a pressure of 10^{-4} bar (though for the conditions in the protoplanetary disk, water ice requires moderately cooler conditions with $T \approx 150$ K). For a Solar mix of elements, the surface density of condensible materials rises dramatically once water ice forms,

$$\Sigma(\text{ices} + \text{rock}) \simeq 4\Sigma(\text{rock}) \tag{93}$$

though the ratio depends upon still uncertain determinations of the exact Solar composition (Lodders, 2003). It is tempting – and extremely plausible – to associate changes in the efficiency or outcome of planet formation (in particular the division between terrestrial and gas giant planets in the Solar System) with the large change in the predicted surface density of solids that occurs at this radius.

The radius in the protoplanetary disk beyond which water ice can be present is called the *snow line*. In the Solar System, water-rich asteroids are found only in the outer asteroid belt (Morbidelli et al., 2000), which suggests that the snow line in the Solar Nebula fell at around 3 AU. Passive protoplanetary disks are predicted to have snow lines at substantially smaller radii – in some cases interior to 1 AU – though accretion rates within the plausible range for Classical T Tauri disks suffice to push the snow line out to the inferred location of 3 AU (Garaud & Lin, 2007; Lecar et al., 2006).

III. PLANET FORMATION

The formation of planets from sub-micron size dust particles requires growth through at least 12 orders of magnitude in spatial scale. It is useful to consider different size regimes in which the interaction between the solid component and the gas is qualitatively distinct:

- **Dust** – small particles ranging from sub-micron to cm in scale. These particles are well-coupled to the gas, but there can be slow drift either vertically or radially. Growth occurs through physical collisions leading to agglomeration.

- **'Rocks'** – objects of meter scale. These particles have increasingly weak coupling to the gas, and so it can be useful to approximate their dynamics as being a combination of Keplerian orbits plus aerodynamic drag. Growth through this size regime is deduced to be rapid but the mechanism remains uncertain.

- **Planetesimals** – bodies of 10 km scale and above. Planetesimals are massive enough that their dynamics is largely decoupled from that of the gas. A population of bodies of this size is often considered as the initial condition for subsequent stage of planet formation, since the evolution of such a population is a well-posed – albeit exceptionally difficult – N-body problem involving primarily purely gravitational forces (though for smaller planetesimals, questions regarding the bodies material strength can also be pertinent).

- **Earth mass** planets or progenitors of the giant planet cores. Once growing planets reach masses of the order of that of Earth, they again become coupled to the gas disk, though this time via gravitational rather than aerodynamic interactions. We will discuss this coupling later in the context of different regimes of planetary *migration*. For terrestrial planet formation it is possible that the formation of Earth mass bodies occurs *after* the gas disk has been dispersed (in which case this coupling is moot), but for growing giant planet cores it is inevitable that interaction will take place.

- **Planetary cores** with masses of the order of 10 $M_\oplus$. At around this mass, there is a transition

from a quasi-hydrostatic core + envelope structure to a regime of rapid gas accretion.

Although it predates the discovery of extrasolar planetary systems, the review by Lissauer (1993) remains an excellent, readable summary of much of the physics that we will address in this section.

A. Planetesimal formation

A spherical particle of radius a, moving relative to the gas at velocity v, experiences an aerodynamic drag force F_D that opposes its motion,

$$F_D = -\frac{1}{2} C_D \cdot \pi a^2 \cdot \rho v^2 \tag{94}$$

where C_D is the *drag coefficient*. The form of the drag coefficient depends upon the size of the particle compared to the mean free path λ of molecules in the gas (Weidenschilling, 1977b; Whipple, 1972). For small particles (typically of cm size or less) for which,

$$a < \frac{9}{4}\lambda \tag{95}$$

the drag coefficient is,

$$C_D = \frac{8}{3}\frac{\bar{v}}{v} \tag{96}$$

where $\bar{v} = (8/\pi)^{1/2} c_s$ is the mean thermal velocity in the gas. This is called the *Epstein regime* of drag. For larger particles the *Stokes* drag law is valid. Defining the Reynolds number via,

$$\mathrm{Re} = \frac{2av}{\nu} \tag{97}$$

where ν is the microscopic (molecular) viscosity in the gas, the drag coefficient can be expressed as a piecewise function,

$$\begin{aligned} C_D &= 24\mathrm{Re}^{-1} && \mathrm{Re} < 1 \\ C_D &= 24\mathrm{Re}^{-0.6} && 1 < \mathrm{Re} < 800 \\ C_D &= 0.44 && \mathrm{Re} > 800. \end{aligned} \tag{98}$$

We will apply these drag laws to consider both the vertical distribution and radial drift of small solid bodies within the gas disk.

1. Dust settling

Dust particles are strongly coupled to the gas via drag forces. For a particle of mass m, the *friction time scale* is defined as,

$$t_{\mathrm{fric}} = \frac{mv}{|F_D|}. \tag{99}$$

It is the time scale on which drag will lead to order unity changes in the relative velocity between the particle and the gas. Writing the particle mass $m = (4/3)\pi a^3 \rho_d$ in terms of the material density ρ_d, the friction time scale has a simple form in the Epstein regime,

$$t_{\mathrm{fric}} = \frac{\rho_d}{\rho}\frac{a}{\bar{v}}. \tag{100}$$

Adopting conditions appropriate to 1 AU within the disk, $\rho = 5 \times 10^{-10}$ gcm^{-3}, $\bar{v} = 2.4 \times 10^5$ cms^{-1} and $\rho_d = 3$ gcm^{-3} we obtain $t_{\mathrm{fric}} \approx 2.5$ s. Small particles are thus very tightly coupled to the gas.

Consider a thin, vertically isothermal gas disk with surface density Σ and scale height $h = c_s/\Omega_K$. The vertical density profile is,

$$\rho(z) = \frac{\Sigma}{h\sqrt{2\pi}} e^{-z^2/2h^2}. \tag{101}$$

To start with, let us ignore the effects of turbulence and assume that the disk is entirely quiescent. In this case the important forces acting on a particle at height z above the midplane are the vertical component of gravity and gas drag, given by,

$$\begin{aligned} |F_{\mathrm{grav}}| &= m\Omega_K^2 z \\ |F_D| &= \frac{4}{3}\pi a^2 \bar{v} \rho v. \end{aligned} \tag{102}$$

Given the strong coupling expected for dust particles terminal velocity will rapidly be attained, so we equate these to obtain the settling speed,

$$v_{\mathrm{settle}} = \left(\frac{\Omega_K^2}{\bar{v}}\right)\frac{\rho_d}{\rho} a z. \tag{103}$$

Settling is more rapid at higher z (where the gas density is lower and the vertical component of gravity stronger), and for larger grains. For example, for micron sized dust particles at $z = h$ at 1 AU the settling velocity is $v_{\mathrm{settle}} \approx 0.1$ cms^{-1} and the settling time scale,

$$t_{\mathrm{settle}} = \frac{z}{|v_{\mathrm{settle}}|} \sim 2 \times 10^5 \text{ yr}. \tag{104}$$

In the absence of turbulence, then, we expect micron sized dust particles to sediment out of the upper layers of the disk on a time scale that is short compared to the disk lifetime, while for particles with sizes < 0.1 μm the time scale is marginal.

Only the density in the equation for the settling time scale is a function of height. Inserting the expression for the vertical density profile the general expression for the settling time scale becomes,

$$t_{\mathrm{settle}} = \frac{2}{\pi}\frac{\Sigma}{\Omega_K \rho_d a} e^{-z^2/2h^2}. \tag{105}$$

The strong z-dependence implies that dust will settle out of the upper regions of the disk rather rapidly in the absence of turbulence. This is of some interest since scattered light images of protoplanetary disks (e.g. Burrows

et al., 1996) are sensitive to dust well away from the midplane.

Assuming that the disk is quiescent is not very realistic – the same turbulence that leads to angular momentum transport is very likely to 'stir up' the dust and prevent it settling to the midplane as easily as our estimate suggests. Working out the influence of turbulence on particle settling is quite tricky, since it involves two subtle issues,

- What is the effective diffusion coefficient in the vertical direction for passive tracer particles injected into the turbulent flow?
- How well coupled are the particles to the gas?

Several recent papers, including Dullemond & Dominik (2004), Johansen & Klahr (2005), Carballido, Stone & Pringle (2005), Turner et al. (2006) and Fromang & Papaloizou (2006), present detailed calculations of the impact of turbulence (primarily that generated by the MRI) on particle settling. For a simple treatment, that is at least qualitatively correct, we can assume that the effective dust diffusion coefficient in z is simply equal to the turbulent viscosity,

$$\nu_d = \nu = \alpha c_s h \tag{106}$$

and that the time scale over which turbulence will stir up a layer of thickness z is given by the usual diffusive expression,

$$t_{\rm stir} \simeq \frac{z^2}{\nu_d}. \tag{107}$$

Clearly $t_{\rm stir}$ increases with z, whereas $t_{\rm settle}$ decreases with height. Equating these expressions, we obtain an estimate of the thickness of the dust layer that will remain well-mixed as a consequence of turbulence. The thickness z is given implicitly via,

$$\frac{h^2}{z^2} e^{-z^2/h^2} = \frac{\pi \rho_d a}{2\alpha\Sigma}. \tag{108}$$

Substituting typical disk and dust properties on the right hand side of this expression, we find that for micron sized particles in a disk with $\alpha = 10^{-2}$ and $\Sigma = 10^3$ gcm^{-2} the right hand side is of the order of 10^{-4}. This means that small dust particles will remain suspended in the disk up to at least several scale heights. On the other hand, for $a = 1$ cm the right hand side is around unity – so large particles will settle to $z \sim h$ even if the disk is fully turbulent.

2. Settling with coagulation

Even if the neglect of turbulence was justified – and it is not – the estimate of the dust settling time in a laminar would be incomplete because it ignores the likelihood that dust particles will collide with one another and grow during the settling process. The settling velocity increases with the particle size, so any such coagulation hastens the collapse of the dust toward the disk midplane.

To estimate how fast particles could grow during sedimentation we appeal to a simple single particle growth model (Dullemond & Dominik, 2005; Safronov, 1969). Imagine that a single "large" particle, of radius a and mass $m = (4/3)\pi a^3 \rho_d$, is settling toward the disk midplane at velocity $v_{\rm settle}$ through a background of much smaller solid particles. By virtue of their small size, the settling of the small particles can be neglected. If every collision leads to coagulation, the large particle grows in mass at a rate that reflects the amount of solid material in the volume swept out by its geometric cross-section,

$$\frac{{\rm d}m}{{\rm d}t} = \pi a^2 |v_{\rm settle}| f \rho(z), \tag{109}$$

where f is the dust to gas ratio in the disk. Substituting for the settling velocity one finds,

$$\frac{{\rm d}m}{{\rm d}t} = \frac{3}{4}\frac{\Omega^2 f}{\bar{v}} z m. \tag{110}$$

Since $z = z(t)$ this Equation cannot generally be integrated immediately[8], but rather must be solved in concert with the equation for the height of the particle above the midplane,

$$\frac{{\rm d}z}{{\rm d}t} = -\frac{\rho_d}{\rho}\frac{a}{\bar{v}}\Omega^2 z. \tag{111}$$

Solutions to these equations provide a very simple model for particle growth and sedimentation in a non-turbulent disk.

Figure 17 shows solutions to equations (110) and (111) for initial particle sizes of 0.01 μm, 0.1 μm and 1 μm. The particles settle from an initial height $z_0 = 5h$ through a disk whose parameters are chosen to be roughly appropriate to a (laminar) Solar Nebula model at 1 AU from the Sun. Both particle and growth and vertical settling are extremely rapid. With the inclusion of coagulation, particles settle to the disk midplane on a time scale of the order of 10^3 yr – more than two orders of magnitude faster than the equivalent time scale in the absence of particle growth. By the time that the particles reach the midplane they have grown to a final size of a few mm, irrespective of their initial radius.

The single particle model described above is very simple, both in its neglect of turbulence and because it assumes that the only reason that particle-particle collisions occur is because the particles have different vertical settling velocities. Other drivers of collisions include Brownian motion, turbulence, and differential *radial* velocities. The basic result, however, is confirmed

[8] Note however that if the particle grows rapidly (i.e. more rapidly than it sediments) then the form of the equation implies exponential growth of m with time.

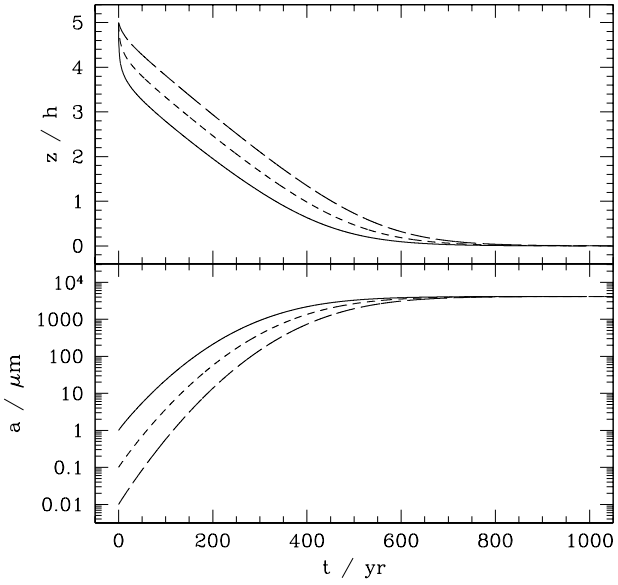

FIG. 17 The settling and growth of a single particle in a laminar (non-turbulent) protoplanetary disk. The model assumes that a single particle (with initial size $a = 1$ μm (solid line), 0.1 μm (dashed line), or 0.01 μm (long dashed line) accretes all smaller particles it encounters as it settles toward the disk midplane. The smaller particles are assumed to be at rest. The upper panel shows the height above the midplane as a function of time, the lower panel the particle radius a. For this example the disk parameters adopted are: orbital radius $r = 1$ AU, scale height $h = 3 \times 10^{11}$ cm, surface density $\Sigma = 10^3$ g cm^{-2}, dust to gas ratio $f = 10^{-2}$, and mean thermal speed $\bar{v} = 10^5$ cm s^{-1}. The dust particle is taken to have a material density $\rho_d = 3$ g cm^{-3} and to start settling from a height $z_0 = 5h$.

by more sophisticated models (Dullemond & Dominik, 2005), which show that if collisions lead to particle adhesion growth from sub-micron scales up to small macroscopic scales (of the order of a mm) occurs rapidly. There are no time scale problems involved with the very earliest phases of particle growth. Indeed, what is more problematic is to understand how the population of small grains – which are unquestionably present given the IR excesses characteristic of Classical T Tauri star – survive to late times. The likely solution to this quandary involves the inclusion of particle *fragmentation* in sufficiently energetic collisions, which allows a broad distribution of particle sizes to survive out to late times. Fragmentation is not likely given collisions at relative velocities of the order of a cm s^{-1} – values typical of settling for micron-sized particles – but becomes more probable for collisions at velocities of a m s^{-1} or higher.

3. Radial drift of particles

Previously we showed (equation 46) that the azimuthal velocity of gas within a geometrically thin disk is close to the Keplerian velocity. That it is not identical, however, turns out to have important consequences for the evolution of small solid bodies within the disk (Weidenschilling, 1977b). We can distinguish two regimes,

- **Small particles** ($a <$ cm) are well-coupled to the gas. To a first approximation we can imagine that they orbit with the *gas* velocity. Since they don't experience the same radial pressure gradient as the gas, however, this means that they feel a net inward force and drift inward at their radial terminal velocity.

- **Rocks** ($a >$ m) are less strongly coupled to the gas. To a first approximation we can imagine that they orbit with the *Keplerian* velocity. This is faster than the gas velocity, so the rocks see a headwind that saps their angular momentum and causes them to spiral in toward the star.

To quantify these effects, we first compute the magnitude of the deviation between the gas and Keplerian orbital velocities. Starting from the radial component of the momentum equation,

$$\frac{v_{\phi,\mathrm{gas}}^2}{r} = \frac{GM_*}{r^2} + \frac{1}{\rho}\frac{\mathrm{d}P}{\mathrm{d}r}, \tag{112}$$

we write the variation of the midplane pressure with radius as a power-law near radius r_0,

$$P = P_0 \left(\frac{r}{r_0}\right)^{-n} \tag{113}$$

where $P_0 = \rho_0 c_s^2$. Substituting, we find,

$$v_{\phi,\mathrm{gas}} = v_K \left(1 - \eta\right)^{1/2} \tag{114}$$

where

$$\eta = n\frac{c_s^2}{v_K^2}. \tag{115}$$

Typically n is positive (i.e. the pressure decreases outward), so the gas orbits slightly slower than the local Keplerian velocity. For example, for a disk of constant $h(r)/r = 0.05$ and surface density profile $\Sigma \propto r^{-1}$ we have $n = 3$ and,

$$v_{\phi,\mathrm{gas}} \simeq 0.996 v_K. \tag{116}$$

The fractional difference between the gas and Keplerian velocities is small indeed! However, at 1 AU even this small fractional difference amounts to a relative velocity of the order of 100 ms^{-1}. Large rocks will then experience a substantial, albeit subsonic, headwind.

The effect of the drag force on the dynamics of particles of arbitrary sizes has been calculated by Weidenschilling (1977b). Here, we adopt the approach of Takeuchi & Lin (2002) and proceed by considering the radial and azimuthal equations of motion for the particle[9],

$$\begin{aligned} \frac{\mathrm{d}v_r}{\mathrm{d}t} &= \frac{v_\phi^2}{r} - \Omega_K^2 r - \frac{1}{t_{\rm fric}}\left(v_r - v_{r,\rm gas}\right) \\ \frac{\mathrm{d}}{\mathrm{d}t}\left(rv_\phi\right) &= -\frac{r}{t_{\rm fric}}\left(v_\phi - v_{\phi,\rm gas}\right). \end{aligned} \tag{117}$$

We simplify the azimuthal equation by noting that the specific angular momentum always remains close to Keplerian (i.e. the particle spirals in through a succession of almost circular, almost Keplerian orbits),

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(rv_\phi\right) \simeq v_r \frac{\mathrm{d}}{\mathrm{d}r}\left(rv_K\right) = \frac{1}{2} v_r v_K. \tag{118}$$

This yields,

$$v_\phi - v_{\phi,\rm gas} \simeq -\frac{1}{2}\frac{t_{\rm fric} v_r v_K}{r}. \tag{119}$$

Turning now to the radial equation, we substitute for Ω_K using equation (114). Retaining only the lowest order terms,

$$\frac{\mathrm{d}v_r}{\mathrm{d}t} = -\eta\frac{v_K^2}{r} + \frac{2v_K}{r}\left(v_\phi - v_{\phi,\rm gas}\right) - \frac{1}{t_{\rm fric}}\left(v_r - v_{r,\rm gas}\right). \tag{120}$$

The $\mathrm{d}v_r/\mathrm{d}t$ term is negligible, and for simplicity we also assume that $v_{r,\rm gas} \ll v_r$, which will be true for those particles experiencing the most rapid orbital decay. Eliminating $(v_\phi - v_{\phi,\rm gas})$ between equations (119) and (120) we obtain,

$$\frac{v_r}{v_K} = \frac{-\eta}{\frac{v_K}{r}t_{\rm fric} + \frac{r}{v_K}t_{\rm fric}^{-1}}. \tag{121}$$

This result can be cast into a more intuitive form by defining a dimensionless stopping time,

$$\tau_{\rm fric} \equiv t_{\rm fric}\Omega_K, \tag{122}$$

in terms of which the particle radial velocity is,

$$\frac{v_r}{v_K} = \frac{-\eta}{\tau_{\rm fric} + \tau_{\rm fric}^{-1}}. \tag{123}$$

The peak radial velocity is attained when $\tau_{\rm fric} = 1$ (i.e. when the friction time scale equals Ω_K^{-1}, and equals $\eta v_K/2$ independent of the disk properties.

Figure 18 plots v_r/v_K as a function of the dimensionless stopping time for a fiducial disk with $h/r = 0.05$.

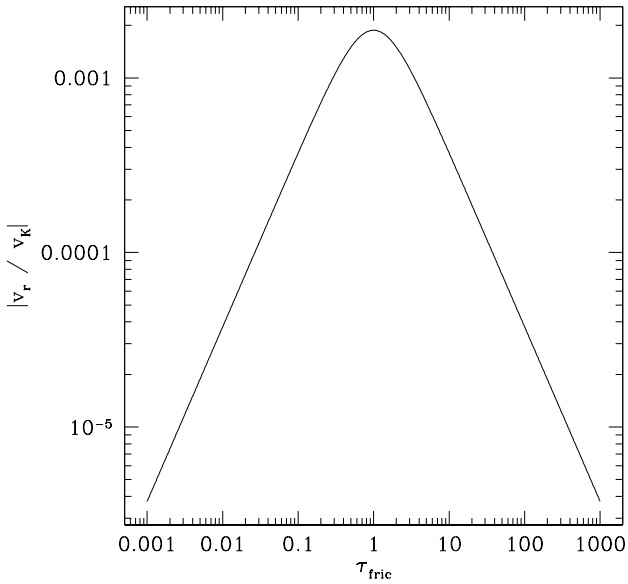

FIG. 18 Radial drift velocity of particles at the midplane of a protoplanetary disk with $h/r = 0.05$, plotted as a function of the dimensionless stopping time $\tau_{\rm fric}$. The radial velocity of the gas has been set to zero. The most rapid inward drift occurs for a physical stopping time Ω_K^{-1}, which for typical disk models translates to a particle size in the 10 cm to m range. At 1 AU, the peak inward velocity is around 60 ms^{-1}, which implies a decay time of less than 100 yr.

Using equations (96) and (98), one can associate a particular $\tau_{\rm fric}$ with a unique particle size a given known conditions in the protoplanetary disk. Generically, one finds that at radii of a few AU the peak inspiral rate is attained for particles with size of the order of 10 cm to a few m. The minimum inspiral time scale at a given orbital radius depends only on η – at 1 AU it is of the order of 100 yr. The inescapable conclusion is that the **radial drift time scale $\ll$ disk lifetime for meter-scale bodies** in the protoplanetary disk.

As we noted earlier, the fact that most of the heavy elements in the Solar System are found in the Sun means that we can tolerate some loss of planetary raw material during planet formation. However, radial drift time scales as short as 100 yr would clearly lead to a catastrophic loss of mass into the star unless, in fact, growth through the meter-scale size regime is very fast. The most important conclusion from this analysis is, therefore, that *planetesimal formation must be a rapid process.* This is a robust inference since it derives directly from the unavoidable existence of a velocity differential between the gas disk and solid bodies orbiting within it.

The radial drift velocities given by equation (123) imply significant radial migration over the lifetime of the disk – not just for particles at the most vulnerable meter-scale size range but also for substantially smaller and larger bodies. This means that we should expect substan-

[9] Although this calculation is straightforward, it's easy to confuse the three different azimuthal velocities that are involved – that of the particle, that of the gas, and the Kepler speed. Be careful!

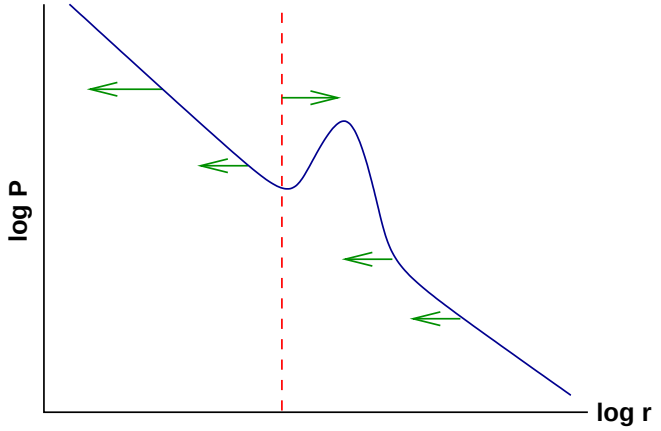

FIG. 19 Illustration of how local pressure maxima within a disk could concentrate solid bodies, forming a ring in this idealized axisymmetric example. Local pressure maxima might arise as a consequence of turbulence within the disk.

tial changes in the local ratio of solids to gas as a function of time and radius in the disk (Takeuchi, Clarke & Lin, 2005). Under some circumstances, radial drift may allow solids to pileup within the inner disk, potentially improving the chances of forming planetesimals there (Youdin & Chiang, 2004).

Several recent papers have examined possible modifications to radial drift that might arise as a consequence of turbulence within the disk (Durisen et al., 2005; Haghighipour & Boss, 2003; Rice et al., 2004). The inward motion of solid bodies embedded within the disk occurs as a consequence of a gas pressure gradient that leads to sub-Keplerian gas orbital velocities. In general, radial drift drives particles toward *pressure maxima*, so that the motion is inevitably inward in a quiescent disk. In a turbulent disk, on the other hand, it may be possible to create local pressure maxima that would act as sites where solids concentrate. The basic idea is illustrated in Figure 19. Such scenarios, although speculative, are potentially interesting since if the hypothesized local pressure maximum occurs on a scale Δr, then the local pressure gradient $\sim P/\Delta r$ exceeds the global gradient $\sim P/r$. The time scale to concentrate solids locally is then faster than the global inspiral time by a factor $\sim (\Delta r/r)^2$.

4. The Goldreich-Ward mechanism

As we have shown, strong circumstantial evidence suggests that the formation of planetesimals must occur on a time scale that is very short compared to the disk lifetime. Two hypothesis have been suggested as to how that occurs,

- Planetesimals may form from pairwise collisional growth of smaller bodies. In this case, the same physical process that allows dust to agglomerate into cm sized objects continues uninterrupted up to the planetesimal size scale. This is an economical hypothesis, and the only difficulty arises due to the uncertain sticking efficiency when cm and meter scale particles collide. Laboratory experiments suggest that the probability of particles sticking and continuing to grow depends upon details of their surface composition (Supulver et al., 1997).

- Planetesimals may form from the gravitational fragmentation of a dense particle sub-disk near the midplane of the gas disk. This suggestion – made by Goldreich & Ward (1973)[10] – is attractive since it forms planetesimals while entirely bypassing the size scales that are most vulnerable to radial drift. However, there are significant theoretical objections which render the simplest versions of the model untenable.

In the following, we discuss the Goldreich-Ward (1973) mechanism together with recent refinements of the concept. The reader should be aware that planetesimal formation is one of the more uncertain aspects of planet formation, and ongoing research continues to study both possibilities.

Figure 20 illustrates the basic idea underlying the Goldreich-Ward (1973) mechanism for planetesimal formation. A combination of vertical settling and radial drift of small solid particles results in the formation of a dense sub-disk within which the solid density exceeds the local gas density (this obviously requires a very thin sub-disk if the local ratio of gas to dust surface density is comparable to the fiducial global value of 100). The solid sub-disk then becomes gravitationally unstable, and fragments into bound clumps of solid particles that subsequently dissipate energy via physical collisions and collapse to form planetesimals.

Gravitational instability requires that the disk be massive (high surface density) and / or dynamically cold (low velocity dispersion). The classic analysis of the conditions for gravitational instability is that of Toomre (1964). Here, we consider the stability of a rotating *fluid* sheet – this is somewhat easier than the collisionless calculation, gives the same answer to a small numerical factor when the gas sound speed is identified with the particle velocity dispersion, and carries over to the instability of a gas disk that we will discuss later. The simplest system to analyze is that of a uniformly rotating sheet – in what follows I follow the notation and approach of Binney & Tremaine (1987).

The setup for the calculation is as shown in Figure 21. We consider a sheet of negligible thickness in the $z = 0$ plane, with constant surface density Σ_0 and angular velocity $\mathbf{\Omega} = \Omega\hat{\mathbf{z}}$. Our aim is to calculate the stability of

[10] Similar considerations are discussed in Safronov (1969), who in turn quotes earlier work by Gurevich & Lebedinskii from as early as 1950.

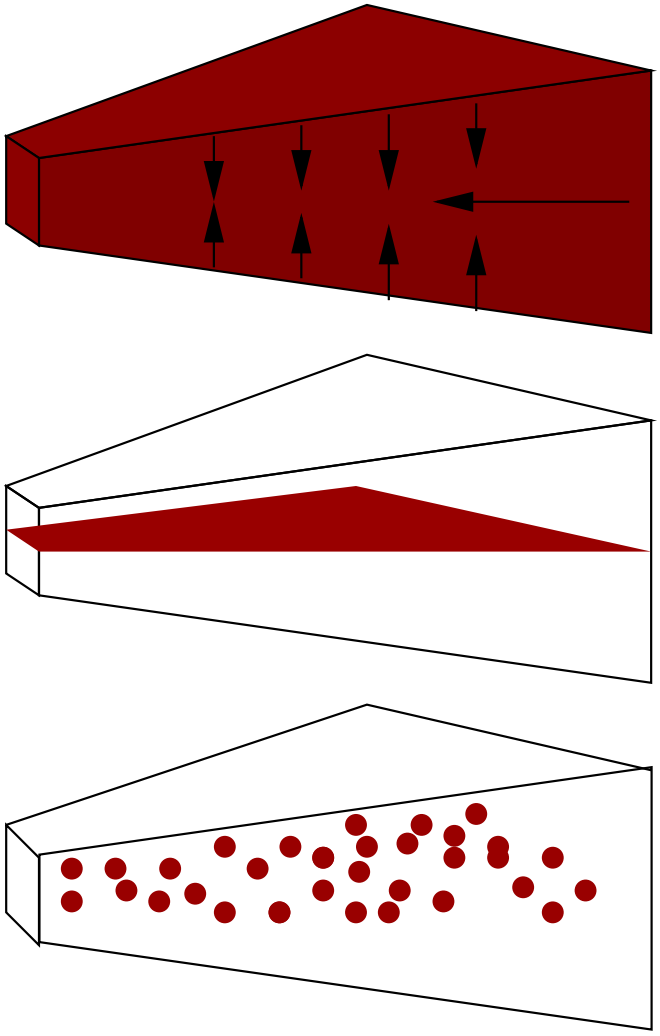

FIG. 20 Illustration of the Goldreich-Ward mechanism for planetesimal formation. A combination of vertical settling and (perhaps) radial drift results in a dust sub-disk whose density exceeds the local gas density. This sub-disk becomes thin enough to be gravitationally unstable, leading to fragmentation into planetesimals.

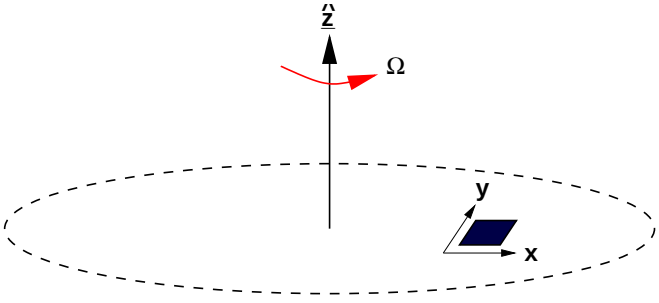

FIG. 21 Geometry for the calculation of the stability of a uniformly rotating sheet.

the sheet to in-plane perturbations. Working in a frame that corotates with the (unperturbed) angular velocity Ω, the fluid equations are,

$$\frac{\partial \Sigma}{\partial t} + \nabla \cdot (\Sigma \mathbf{v}) = 0 \tag{124}$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\,\mathbf{v} = -\frac{\nabla p}{\Sigma} - \nabla\Phi - 2\boldsymbol{\Omega} \times \mathbf{v} + \Omega^2\,(x\hat{\mathbf{x}} + y\hat{\mathbf{y}}) \tag{125}$$

where the momentum equation picks up terms for the Coriolis and centrifugal forces in the rotating frame. These equations apply in the $z = 0$ plane only. The gravitational potential Φ is given by the Poisson equation,

$$\nabla^2\Phi = 4\pi G\Sigma\delta(z) \tag{126}$$

which describes Φ in all space. In these equations, $\mathbf{v} = v_x\hat{\mathbf{x}} + v_y\hat{\mathbf{y}}$ is the velocity *in the rotating frame*, Σ is the surface density, and $p = p(\Sigma)$ is the vertically integrated pressure. The sound speed is defined via,

$$c_s^2 \equiv \left.\frac{\mathrm{d}p}{\mathrm{d}\Sigma}\right|_{\Sigma=\Sigma_0}. \tag{127}$$

In the unperturbed state, $\Sigma = \Sigma_0$, $\Phi = \Phi_0$, $\mathbf{v} = 0$ and $p = p_0 = p(\Sigma_0)$. Substituting these values into the momentum equation yields $\nabla\Phi_0 = \Omega^2(x\hat{\mathbf{x}} + y\hat{\mathbf{y}})$.

We now consider perturbations to the surface density, velocity, pressure and potential,

$$\begin{aligned} \Sigma &= \Sigma_0 + \Sigma_1(x,y,t) \\ \mathbf{v} &= \mathbf{v}_1(x,y,t) \\ p &= p_0 + p_1(x,y,t) \\ \Phi &= \Phi_0 + \Phi_1(x,y,z,t) \end{aligned} \tag{128}$$

where it is assumed that $\Sigma_1 \ll \Sigma_0$ etc. Substituting these expressions into the fluid equations, and retaining only those terms that are linear in the perturbed quantities, we find,

$$\frac{\partial \Sigma_1}{\partial t} + \Sigma_0 \nabla \cdot \mathbf{v}_1 = 0 \tag{129}$$

$$\frac{\partial \mathbf{v}_1}{\partial t} = -\frac{c_s^2}{\Sigma_0}\nabla\Sigma_1 - \nabla\Phi_1 - 2\boldsymbol{\Omega} \times \mathbf{v}_1 \tag{130}$$

$$\nabla^2\Phi_1 = 4\pi G\Sigma_1\delta(z) \tag{131}$$

where we have made use of the fact that since p is only a function of Σ, $\nabla p = (\mathrm{d}p/\mathrm{d}\Sigma)\nabla\Sigma$. Note that these equations *only* involve temporal or spatial derivatives of the perturbed quantities. Since the equations are (by construction) linear, the evolution of an arbitrary perturbation can be decomposed into fourier modes. Assuming a wavevector $\mathbf{k}$ that is parallel to $\hat{\mathbf{x}}$, we therefore write the perturbations in the form,

$$\Sigma_1(x,y,t) = \Sigma_a e^{i(kx-\omega t)} \tag{132}$$

$$\mathbf{v}_1 = (v_{ax}\hat{\mathbf{x}} + v_{ay}\hat{\mathbf{y}})e^{i(kx-\omega t)} \tag{133}$$

$$\Phi_1 = \Phi_a e^{i(kx-\omega t)} \tag{134}$$

where the final expression describes the potential perturbations in the $z = 0$ plane only. Substitution of these expressions into the perturbation equations will reduce them to algebraic expressions, which can be combined to yield the dispersion relation for the system.

First though, we simplify the system by noting that perturbations in Σ are the *source* of perturbations in Φ. We can therefore write Φ_a in terms of Σ_a. To do this, let the general form for Φ_1 (i.e. *not* just at $z = 0$) be,

$$\Phi_1 = \Phi_a e^{i(kx-\omega t)} \times f(z) \tag{135}$$

where $f(z)$ is some function that needs to be determined. Requiring that $\nabla^2\Phi_1 = 0$ for $z \neq 0$, we find,

$$\frac{\mathrm{d}^2 f}{\mathrm{d}z^2} = k^2 f \tag{136}$$

which has a general solution $f = Ae^{-kz} + B^{kz}$, with A and B arbitrary constants. Since Φ_1 must remain finite as $z \to \pm\infty$, the general form of Φ_1 is,

$$\Phi_1 = \Phi_a e^{i(kx-\omega t)-|kz|}. \tag{137}$$

This is valid throughout all space.

To determine Φ_a, we integrate the Poisson equation vertically between $z = -\epsilon$ and $z = +\epsilon$,

$$\int_{-\epsilon}^{+\epsilon} \nabla^2\Phi_1 dz = \int_{-\epsilon}^{+\epsilon} 4\pi G\Sigma_1\delta(z)dz. \tag{138}$$

Mathematically this requires a bit of care, since the integrand on the left hand side is zero everywhere except at $z = 0$. However, noting that $\partial^2\Phi_1/\partial x^2$ and $\partial^2\Phi_1/\partial y^2$ are continuous at $z = 0$, while $\partial^2\Phi_1/\partial z^2$ is *not*, we obtain,

$$\int_{-\epsilon}^{+\epsilon} \nabla^2\Phi_1 dz = \left.\frac{\partial\Phi_1}{\partial z}\right|_{-\epsilon}^{+\epsilon} = \int_{-\epsilon}^{+\epsilon} 4\pi G\Sigma_1\delta(z)dz. \tag{139}$$

Taking the limit as $\epsilon \to 0$,

$$-2|k|\Phi_a = 4\pi G\Sigma_a \tag{140}$$

and,

$$\Phi_1 = -\frac{2\pi G\Sigma_a}{|k|}e^{i(kx-\omega t)-|kz|}. \tag{141}$$

We are now in a position to substitute Σ_1, $\mathbf{v}_1$ and Φ_1 into the remaining equations (continuity plus the x and y components of the momentum equation). The resulting algebraic equations are,

$$\begin{aligned} -i\omega\Sigma_a &= -ik\Sigma_0 v_{ax} \\ -i\omega v_{ax} &= -\frac{c_s^2}{\Sigma_0}ik\Sigma_a + \frac{2\pi G i\Sigma_a k}{|k|} + 2\Omega v_{ay} \quad (142) \\ -i\omega v_{ay} &= -2\Omega v_{ax}. \quad (143) \end{aligned}$$

We seek a *dispersion relation* i.e. a formula for the growth rate $\omega = f(k)$ of modes of different scale k. Eliminating v_{ax} and v_{ay} in turn, we obtain,

$$\omega^2 = c_s^2k^2 - 2\pi G\Sigma_0|k| + 4\Omega^2. \tag{144}$$

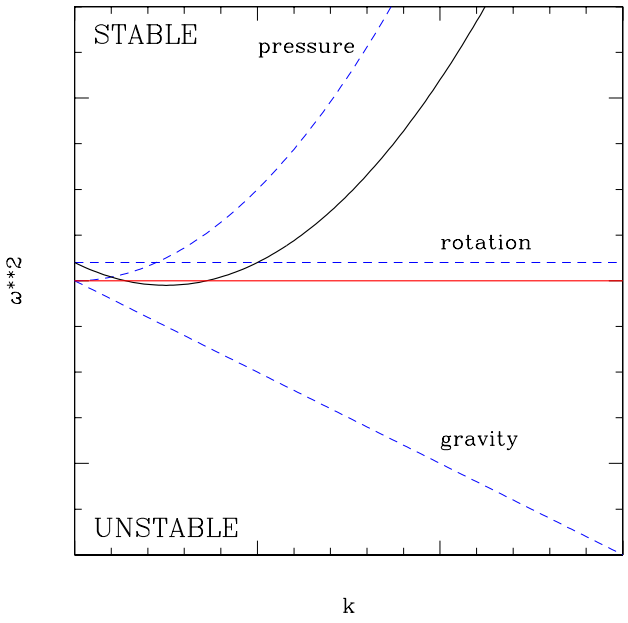

FIG. 22 The dispersion relation (solid black line) for a uniformly rotating sheet, illustrating the contributions from pressure, rotation, and self-gravity (dashed blue lines). The system is unstable if, at any value of the wavenumber k, ω^2 falls below the red line and is negative. Pressure is a stabilizing influence that is most important at large k (small spatial scales), while rotation acts to stabilize the system at small k (large spatial scales).

This is the dispersion relation for a uniformly rotating thin sheet. The scale-dependence of the different terms is shown graphically in Figure 22.

Looking back to the form of the perturbations, we note that the sheet is:

- STABLE if $\omega^2 \geq 0$, since in this case ω is real and the perturbations are oscillatory.
- UNSTABLE if $\omega^2 < 0$, for which case ω is imaginary and perturbations grow exponentially.

The rotational term $(4\Omega^2)$ is stabilizing at all scales, while the pressure term $(c_s^2k^2)$ has a strong stabilizing influence at large k (i.e. small spatial scales). Self-gravity, represented by the $-2\pi G\Sigma_0|k|$ term, has a negative contribution to ω^2 and so destabilizes the sheet.

The condition for marginal stability is that $\omega^2 \geq 0$ at all spatial scales. The most unstable scale $k_{\rm crit}$ can be found by setting $\mathrm{d}\omega^2/\mathrm{d}k = 0$, which yields,

$$k_{\rm crit} = \frac{\pi G\Sigma_0}{c_s^2}. \tag{145}$$

The sheet is marginally stable when $\omega^2(k_{\rm crit}) = 0$, which gives the stability condition as,

$$\frac{c_s\Omega}{G\Sigma_0} = \frac{\pi}{2}. \tag{146}$$

This analysis can be extended in several ways – for example to include differential rotation or global rather than local stability. A generic way of expressing the results of such calculations is to define the Toomre Q parameter,

$$Q \equiv \frac{c_s \Omega}{\pi G \Sigma}. \tag{147}$$

In terms of Q, a disk is unstable[11] to its own self-gravity if $Q < Q_{\rm crit}$, and stable if $Q > Q_{\rm crit}$. Typically $Q_{\rm crit} \simeq 1$ – for the specific system we have investigated it would be 1/2.

We have derived the stability of a fluid disk in uniform rotation. Differential rotation and global effects alter the value of $Q_{\rm crit}$, but do not fundamentally change the result. For a collisionless disk (e.g. one made of stars or small solid particles) a comparable result applies if we replace the sound speed c_s by the one-dimensional velocity dispersion σ.

The most unstable wavelength is,

$$\lambda_{\rm crit} = \frac{2\pi}{k_{\rm crit}} = \frac{2 c_s^2}{G \Sigma_0}. \tag{148}$$

Comparing this to the scale height of the disk $h = c_s/\Omega$, we find that at marginal stability,

$$\frac{\lambda_{\rm crit}}{h} \simeq 2\pi \tag{149}$$

i.e. the instability afflicts small-ish spatial scales within the disk.

Let us apply this analysis to the problem of planetesimal formation. If we ignore radial drift, then at 1 AU $\Sigma_{\rm dust} \sim 10^{-2} \Sigma_{\rm gas}$, or about 10 gcm^{-2} for a minimum mass Solar Nebula model (note that a gas to dust ratio of 100 is a commonly used approximation in protoplanetary disk theory). Setting $Q = \sigma\Omega/(\pi G \Sigma_{\rm dust}) = 1$, and taking $M_* = M_\odot$, we find that instability requires a critical velocity dispersion in the solid component,

$$\sigma \simeq 10 \text{ cms}^{-1}. \tag{150}$$

Since the *gas* sound speed at this radius is of the order of 10^5 cms^{-1}, and the scale heights of the gas and particle disks are respectively proportional to c_s and σ, we see that an extremely thin disk is required before instability will set in!

[11] For a differentially rotating disk, it is easy to verify that stability depends upon the parameter combination $c_s\Omega/(G\Sigma_0)$ via a time scale argument. First derive the time scale for shear to separate two points that are initially Δr apart, and equate this to the collapse time scale under gravity to find the maximum scale on which collapse can occur without being affected by shear. Taking the ratio of this scale to the Jeans scale (the smallest scale on which collapse can occur without being inhibited by pressure gradients) yields the correct functional form of Q.

If instability occurs, the most unstable wavelength is predicted to be,

$$\lambda_{\rm crit} \approx 3 \times 10^8 \text{ cm}. \tag{151}$$

The mass within an unstable patch is then,

$$m \sim \pi \Sigma_{\rm dust} \lambda_{\rm crit}^2 \sim 3 \times 10^{18} \text{ g} \tag{152}$$

which would correspond to a spherical body of size,

$$r = \left(\frac{3m}{4\pi\rho_d} \right)^{1/3} \sim 6 \text{ km} \tag{153}$$

for a material density of $\rho_d = 3$ gcm^{-3}. The collapse time scale at distance $\lambda_{\rm crit}$ from mass m,

$$t_{\rm ff} = \sqrt{\frac{\lambda_{\rm crit}^3}{2Gm}} \tag{154}$$

is very short – less than a year for the parameters adopted above. Even if we allow for the fact that angular momentum will preclude a prompt collapse, the derived time scale for planetesimal formation via gravitational instability remains extremely short – perhaps of the order of 10^3 yr (Goldreich & Ward, 1973).

Formation of planetesimals via the Goldreich-Ward mechanism has several attractive features, most notably the short time scale and complete bypass of the size regime most vulnerable to radial drift. However in its simplest form, the mechanism fails to work. The problem lies in the fact that *even in an intrinsically non-turbulent* gas disk, the formation of a dense solid sub-disk leads to self-generated turbulence and associated vertical stirring *prior* to gravitational instability. As noted above, for gravitational instability to operate we require a thin sub-disk in which, for our choice of parameters,

$$\frac{h_{\rm dust}}{h_{\rm gas}} \sim 10^{-4}. \tag{155}$$

Within this midplane layer, the volume density of solids would *exceed* the density of gas by a factor of the order of 100 – i.e. the extreme thinness of the solid disk inverts the normal gas to dust ratio which favors *gas* by the same factor. Since the gas and dust are well coupled for small particle sizes, within the sub-disk (where the solid component dominates) we expect both the gas and the dust to orbit at the natural velocity for the solid component, which is the Kepler velocity. The gas just above the layer, on the other hand, will rotate slower due to the radial gas pressure gradient. There will therefore be a velocity gradient in the z direction that is of the order of $(h_{\rm gas}/r)^2 v_K / h_{\rm dust}$. This shear will be Kelvin-Helmholtz unstable, leading to turbulence that prevents the layer ever getting thin enough to fragment into planetesimals (Cuzzi, Dobrovolskis & Champney, 1993). The condition for Kelvin-Helmholtz instabilities to develop

(Sekiya, 1998; Youdin & Shu, 2002) is that the *Richardson number*, which measures the competition between vertical shear and buoyancy, is $\mathrm{Ri} < \mathrm{Ri}_{\mathrm{crit}}$, where,

$$\mathrm{Ri} \equiv \frac{N^2}{(\partial v_\phi/\partial z)^2} \tag{156}$$

and N, the Brunt Väisälä frequency, is defined as,

$$N^2 \equiv g_z \frac{\partial \ln \rho}{\partial z}. \tag{157}$$

The standard stability analysis obtains a critical Richardson number $\mathrm{Ri} = 0.25$, but both analytic calculations including the effect of Coriolis forces, and numerical simulations, favor a larger value of around unity (Gomez & Ostriker, 2005; Johansen, Henning & Klahr, 2006).

When the stirring effects of Kelvin-Helmholtz turbulence are included, all authors agree that the Goldreich-Ward path to planetesimal formation fails for a standard gas to dust ratio of 100. It may, however, still be possible to form planetesimals via gravitational instability if some other process is able to increase the local concentration of solids by a large amount (Garaud & Lin, 2004). One idea, for example, is that radial drift might increase the amount of solids in the inner disk at the expense of material further out (Youdin & Chiang, 2004; Youdin & Shu, 2002). Alternatively, photoevaporation might remove the gas leaving a solid-rich residue (Throop & Bally, 2005). Given the empirical inference (in the Solar System) that planetesimals managed to form across a wide range of orbital radii, perhaps the most appealing possibility is that turbulence manages to concentrate particles on very small scales prior to the onset of gravitational instability. As we have already noted, aerodynamic forces result in the concentration of solids at transient pressure maxima, and this effect may be enhanced by the action of *streaming* instabilities which exist in coupled gas-particle systems (Youdin & Goodman, 2005). Johansen et al. (2007) has presented numerical simulations which support such a scenario, while Cuzzi, Hogan & Shariff (2008) have proposed a model which – although different in many details – also appeals to strong particle concentration in turbulence as a prerequisite for self-gravity becoming dynamically important. The details of these models remain to be worked out, but one interesting aspect is that they may imply that the planetesimal formation process skips over not only meter but even kilometer scales. If the first planetesimals were born very large (say 100 km) some signature of this might be evident today in the size distribution of asteroids and other minor bodies in the Solar System (Morbidelli et al., 2009).

B. Growth beyond planetesimals

Once planetesimals have formed, further interaction between the solid and gaseous components of the disk is limited until bodies with sizes $> 10^3$ km form that are large enough to have a *gravitational* coupling to the gas[12]. We will discuss the impact of gravitational coupling ('migration') later in the context of the early evolution of planetary systems. If coupling with the gas disk can be neglected, further growth to form protoplanets or planetary embryos is a well-posed N-body problem in which gravity provides the dominant physics.

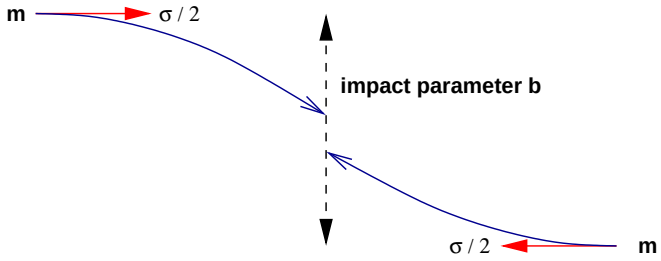

FIG. 23 Setup for calculation of gravitational focusing. Two bodies of mass m, moving on a trajectory with impact parameter b, have a velocity at infinity of $\sigma/2$.

Being well-posed is not the same as easy – if the Earth formed from 5 km radius planetesimals then $N \sim 10^9$. Although N-body simulations with this many particles are certainly feasible (Springel et al., 2005), it is not possible to simulate such large particle numbers for the $\sim 10^8$ orbits required for an *ab initio* calculation of terrestrial planet formation (making matters more difficult, for long duration integrations special numerical techniques are needed to keep integration errors under control). The usual approach is therefore a combination of statistical and N-body methods.

1. Gravitational focusing

For sufficiently small bodies, the effects of gravity can be ignored for the purposes of determining whether they will physically collide. A massive planet, on the other hand, can *gravitationally focus* other bodies toward it, and as a result has a collision cross section that is much larger than its physical cross section.

To evaluate the magnitude of this gravitational focusing, consider two bodies of mass m, moving on a trajectory with impact parameter b, as shown in Figure 23. The relative velocity at infinity is σ. At closest approach, the bodies have separation R_c and velocity $V_{\max}$. Equating energy in the initial (widely separated) and final (closest

[12] Strictly, all that is known for sure is that **aerodynamic** effects are negligible for planetesimals. The gas disk might still couple to planetesimals gravitationally, if turbulence produces surface density fluctuations that can gravitationally scatter planetesimals. There have been a number of recent studies of this process (Johnson, Goodman & Menou, 2006; Laughlin, Steinacker & Adams, 2004; Nelson, 2005; Nelson & Papaloizou, 2004; Yang, Mac Low & Menou, 2009).

approach) states we have,

$$\frac{1}{4}m\sigma^2 = mV_{\rm max}^2 - \frac{Gm^2}{R_c}. \quad (158)$$

Noting that there is no radial component to the velocity at the point of closest approach, angular momentum conservation gives,

$$V_{\rm max} = \frac{1}{2}\frac{b}{R_c}\sigma. \quad (159)$$

If the sum of the physical radii of the bodies is R_s, then for $R_c < R_s$ there will be a physical collision, while larger R_c will result in a harmless flyby[13]. The *largest* value of the impact parameter that will lead to a physical collision is thus,

$$b^2 = R_s^2 + \frac{4GmR_s}{\sigma^2}, \quad (160)$$

which can be expressed in terms of the escape velocity from the point of contact, $v_{\rm esc}^2 = 4Gm/R_s$ as,

$$b^2 = R_s^2\left(1 + \frac{v_{\rm esc}^2}{\sigma^2}\right). \quad (161)$$

The cross section for collisions is then,

$$\Gamma = \pi R_s^2\left(1 + \frac{v_{\rm esc}^2}{\sigma^2}\right), \quad (162)$$

where the term in brackets represents the enhancement to the physical cross section due to gravitational focusing. Clearly a planet growing in a 'cold' planetesimal disk for which $\sigma \ll v_{\rm esc}$ will grow much more rapidly as a consequence of gravitational focusing. As a consequence, determining the velocity dispersion of bodies of different masses during the planet formation process is extremely important.

2. Growth versus fragmentation

When two initially solid bodies physically collide the outcome can be divided broadly into three categories:

- **Accretion**. All or most of the mass of the impactor becomes part of the mass of the final body, which remains solid. Small fragments may be ejected, but overall there is net growth.
- **Shattering**. The impact breaks up the target body into a number of pieces, but these pieces remain part of a single body (perhaps after reaccumulating gravitationally). The structure of the shattered object resembles that of a *rubble pile*.
- **Dispersal**. The impact fragments the target into two or more pieces that do not remain bound.

To delineate the boundaries between these regimes quantitatively, we consider an impactor of mass m colliding with a larger body of mass M at velocity v. We define the specific energy Q of the impact via,

$$Q \equiv \frac{mv^2}{2M}, \quad (163)$$

and postulate, plausibly, that this parameter largely controls the result. The thresholds for the various collision outcomes can then be expressed in terms of Q. Conventionally, we define the threshold for catastrophic disruption Q_D^* as the minimum specific energy needed to disperse the target in two or more pieces, with the largest one having a mass $M/2$. Similarly Q_S^* is the threshold for shattering the body. More work is required to disperse a body than to shatter it, so evidently $Q_D^* > Q_S^*$. It is worth keeping in mind that in detail the outcome of a particular collision will depend upon many factors, including the mass ratio between the target and the impactor, the angle of impact, and the shape and rotation rate of the bodies involved. Quoted values of Q_D^* are often averaged over impact angles, but even when this is done the parameterization of collision outcomes in terms of Q is only an approximation.

The estimated values of Q_D^* for a target of a particular size vary by more than an order of magnitude depending upon the composition of the body, which can broadly be categorized into solid or shattered rock, and solid or porous ice. For any particular type of body, however, two distinct regimes can be identified:

- **Strength dominated regime**. The ability of small bodies to withstand impact without being disrupted depends upon the material strength of the object. In general, the material strength of bodies declines with increasing size, owing to the greater prevalence of defects that lead to cracks. In the strength dominated regime Q_D^* decreases with increasing size.
- **Gravity dominated regime**. Large bodies are held together primarily by gravitational forces. In this regime Q_D^* must at the very least exceed the specific binding energy of the target, which scales with mass M and radius a as $Q_B \propto GM/a \propto \rho_d a^2$. In practice it requires a great deal more than this minimum amount of energy to disrupt the target – so Q_B is *not* a good estimate of Q_D^* – but nonetheless Q_D^* does increase with increasing size.

Although the transition between these regimes is reasonably sharp there is *some* influence of the material properties (in particular the shear strength) on the catastrophic disruption threshold for smaller bodies within the gravity dominated regime.

Values of Q_S^* and Q_D^* can be determined experimentally for small targets (Arakawa, Leliwa-Kopystynski &

[13] This is true for solid bodies – for giant planets or stars tidal effects can lead to significant dissipation of energy even when $R_c > R_s$ (Fabian, Pringle & Rees, 1975).

TABLE III Parameters for the catastrophic disruption threshold fitting formula (equation 164), which describes how Q_D^* scales with the size of the target body. The quoted values were derived by Benz & Asphaug (1999) and Leinhardt & Stewart (2009) using numerical hydrodynamics simulations of collisions, which are supplemented in the strength dominated regime by experimental results.

	v / km s^{-1}	q_s / erg g^{-1}	q_g / erg cm^3 g^{-2}	c	d
Ice (weak)	1.0	1.3×10^6	0.09	-0.40	1.30
Ice (strong)	0.5	7.0×10^7	2.1	-0.45	1.19
Ice (strong)	3.0	1.6×10^7	1.2	-0.39	1.26
Basalt (strong)	3.0	3.5×10^7	0.3	-0.38	1.36
Basalt (strong)	5.0	9.0×10^7	0.5	-0.36	1.36

Maeno, 2002). Experiments are not possible in the gravity dominated regime, but Q_D^* can be estimated theoretically using numerical hydrodynamics (Benz & Asphaug, 1999; Leinhardt & Stewart, 2009) or (for rubble piles) rigid body dynamics simulations (Korycansky & Asphaug, 2006; Leinhardt & Richardson, 2002). The simplest parameterization of the numerical results is as a broken power law that includes terms representing the strength and gravity regimes,

$$Q_D^* = q_s \left(\frac{a}{1 \text{ cm}}\right)^c + q_g \rho_d \left(\frac{a}{1 \text{ cm}}\right)^d . \qquad (164)$$

Often (but not always) Q_D^* is averaged over impact geometry, and q_s, q_g, c and d are all constants whose values are derived by fitting to the results of numerical simulations.

Benz & Asphaug (1999) and Leinhardt & Stewart (2009) determined the values of the fitting parameters in equation (164) from the results of an ensemble of simulations of impacts into icy or rocky targets. Their results are given in Table III and plotted as a function of target size in Figure 24. One observes immediately that the results for a particular target material vary with the impact velocity, and hence that Q_D^* is *not* the sole determinant of the outcome of collisions. There is, however, a clear transition between the strength and gravity dominated regimes, with the weakest bodies being those whose size is comparable to the cross-over point. The most vulnerable bodies are generally those with radii in the 100 m to 1 km range. Just how vulnerable such bodies are to catastrophic disruption depends sensitively on their make-up, and it would be unwise to place too much trust in precise numbers. As a rough guide, however, the weakest icy bodies have minimum $Q_D^* \sim 10^5$ erg g^{-1}, while the strongest conceivable planetesimals (unfractured rocky bodies) have minimum $Q_D^* > 10^6$ erg g^{-1}.

As a reality check, we may note that asteroids in the main belt with $e \simeq 0.1$ would be expected to collide today with typical velocities of the order of 2 km s^{-1}. For a mass ratio $m/M = 0.1$ the specific energy of the collision is then around $Q = 2 \times 10^9$ erg g^{-1}, which from Figure 24

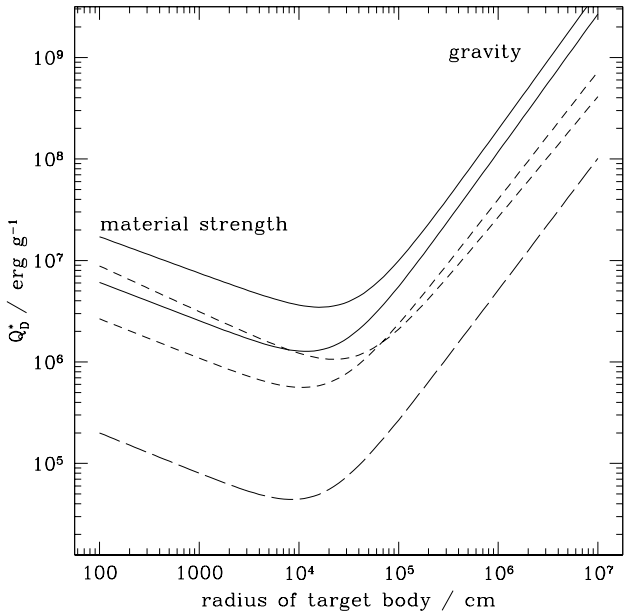

FIG. 24 The specific energy Q_D^* for catastrophic disruption of solid bodies is plotted as a function of the body's radius. The solid and short dashed curves show results obtained using fits to theoretical calculations for impacts into "strong" targets by Benz & Asphaug (1999). The long dashed curve shows the recommended curve for impacts into "weak" targets from Leinhardt & Stewart (2008), derived from a combination of impact experiments and numerical simulations. In detail the solid curves show results for basalt at impact velocities of 5 km s^{-1} (upper curve) and 3 km s^{-1} (lower curve). The short dashed curves show results for water ice at 3 km s^{-1} (the lower curve for small target sizes) and 0.5 km s^{-1} (upper curve for small target sizes). The long dashed curve shows results for normal impacts into weak water ice targets at 1 km s^{-1}.

is sufficient to destroy even quite large solid bodies with $a \simeq 100$ km. This is consistent with the observation of asteroid families, and the interpretation of such families as collisional debris. Evidently the random velocities that characterize collisions must have been *much* smaller during the epoch of planet formation if we are to successfully build large planets out of initially km-scale planetesimals.

3. Shear versus dispersion dominated encounters

A more subtle distinction that nevertheless plays a crucial role in planet formation is whether encounters between bodies can be described via 2-body dynamics — in which only the gravity of the two objects themselves matters — or whether the tidal influence of the Sun also needs to be considered (3-body dynamics). Goldreich, Lithwick & Sari (2004) have recently summarized in simple terms why the distinction between 2 and 3-body dynamics matters at different stages of the planet formation process. We consider a 3-body system consisting

of a large body (a 'planet') with mass M, a small body of negligible mass (described as a test particle), and the Sun, and define the *Hill radius* r_H as the radius within which the gravity of the planet dominates (in astrophysical contexts, the same concept is referred to as the 'Roche lobe'). Roughly, this is obtained by equating the angular velocity for an orbit at distance r_H from the planet with the angular velocity of the planet around the star. We find,

$$r_H = \left(\frac{M_p}{3M_*}\right)^{1/3} a \tag{165}$$

where the factor 3 is included for consistency with more detailed derivations. For circular orbits, collisions are forbidden for an orbital separation Δa between the small body and the planet such that $\Delta a \lesssim r_H$ (c.f. the Trojan asteroids in the Solar System). If we define a characteristic velocity at the Hill radius,

$$v_H \equiv \sqrt{\frac{GM_p}{r_H}} \tag{166}$$

then for,

- $\sigma > v_H$ 2-body dynamics describes collisions quite well. This regime is called **dispersion dominated**.
- $\sigma < v_H$ 3-body effects are important. This regime is called **shear dominated**.

When $\sigma < v_H$ and we are shear dominated, the collision rate is modified compared to expectations based on 2-body dynamics.

4. Growth rates

We now proceed to derive an estimate for how fast a planet will grow due to accretion of planetesimals. We assume that the growing body, of mass M, radius R_s, and surface escape speed $v_{\rm esc}$ is embedded within a 'swarm' of planetesimals with local surface density Σ_p, velocity dispersion σ, and scale height h_p. The volume density of the planetesimal swarm is,

$$\rho_{sw} = \frac{\Sigma_p}{2h_p}. \tag{167}$$

Then if 3-body effects can be ignored, the large body grows at a rate,

$$\frac{{\rm d}M}{{\rm d}t} = \rho_{sw}\sigma\pi R_s^2\left(1+\frac{v_{\rm esc}^2}{\sigma^2}\right). \tag{168}$$

This can be simplified since $h_p \sim \sigma/\Omega$ and hence ρ_{sw} is inversely proportional to σ. We find,

$$\frac{{\rm d}M}{{\rm d}t} = \frac{1}{2}\Sigma_p\Omega\pi R_s^2\left(1+\frac{v_{\rm esc}^2}{\sigma^2}\right) \tag{169}$$

where the numerical prefactor, which has not been derived accurately here, depends upon the assumed velocity distribution of the planetesimals. For an isotropic distribution the prefactor is $\sqrt{3}/2$ (Lissauer, 1993).

This simple result is important. We note that:

- The velocities of the planetesimals enter only via the gravitational focusing term, which can however by very large.
- The rate of mass growth scales linearly with Σ_p — we expect faster growth in disks that have more mass in planetesimals (due to a higher gas mass and / or a higher ratio of solids to gas).
- Other things being equal, growth will be slower at large radii, due to lower Σ_p and smaller Ω.

Complexity arises because as a planet grows, it stars to influence both the velocity dispersion and, eventually, the surface density of the planetesimal swarm in its vicinity.

Two simple solutions of the growth equation give an idea of the possibilities present in more sophisticated models. First, assume that the gravitational focusing term F_g is constant. In this regime,

$$\frac{{\rm d}M}{{\rm d}t} \propto R_s^2 \propto M^{2/3} \tag{170}$$

which has solution,

$$R_s \propto t. \tag{171}$$

The radius of the planet grows at a linear rate. Writing the planet mass $M = (4/3)\pi R_s^3\rho_{\rm planet}$, where $\rho_{\rm planet}$ is the planet density,

$$\frac{{\rm d}R_s}{{\rm d}t} = \frac{\Sigma_p\Omega}{8\rho_{\rm planet}}F_g. \tag{172}$$

If we assume that at the orbital radius of Jupiter $\Sigma_p = 10$ gcm^{-2}, then for $\rho_{\rm planet} = 3$ gcm^{-3},

$$\frac{{\rm d}R_s}{{\rm d}t} \simeq 0.2F_g \text{ cm yr}^{-1}. \tag{173}$$

This initial growth rate is slow, which implies that to form the cores of the giant planets in a reasonable time, large gravitational focusing factors are needed. For example, to reach 1000 km in 10^5 yr, we require $F_g \sim 5000$. The need for large gravitational enhancements to the collision rate is even more severe for the ice giants, but substantially easier in the terrestrial planet region.

Since empirically F_g *must* be large, a second useful limit to consider is the case where $F_g \gg 1$. If we assume that σ is constant (i.e. consider the regime where the growing planet has not yet managed to dominate the dynamical excitation of the planetesimal swarm) then,

$$\begin{aligned} F_g &= \left(1+\frac{v_{\rm esc}^2}{\sigma^2}\right) \\ &\simeq \frac{v_{\rm esc}^2}{\sigma^2} \\ &\propto \frac{M}{R_s}. \end{aligned} \tag{174}$$

The growth equation (169) gives,

$$\frac{\mathrm{d}M}{\mathrm{d}t} \propto MR_s \tag{175}$$

with solution,

$$M = \frac{1}{(M_0^{-1/3} - kt)^3}, \tag{176}$$

where M_0 is the initial mass at time $t = 0$ and k is a constant. In this regime the increasing gravitational focusing factor means that $M \to \infty$ in a finite time, allowing much more rapid growth.

5. Isolation mass

As noted above, rapid growth requires that σ remain low — i.e. that the planetesimals remain on roughly circular orbits. This means that there is a finite supply of planetesimals that have orbits that pass close enough to a growing planet to collide — once these have all been consumed growth is bound to slow. The mass at which this slowdown occurs is described as the *isolation mass* $M_{\rm iso}$.

To estimate the isolation mass, we note that a planet grows by accreting planetesimals within a 'feeding zone'. The size of the feeding zone $\Delta a_{\rm max}$ is set by the maximum distance over which the planet's gravity is able to perturb planetesimal orbits sufficiently to allow collisions, so it will scale with the Hill radius. Writing

$$\Delta a_{\rm max} = C r_H \tag{177}$$

with C a constant of order unity, we have that the mass of planetesimals within the feeding zone is,

$$2\pi a \cdot 2\Delta a_{\rm max} \cdot \Sigma_p \propto M^{1/3}. \tag{178}$$

Note the 1/3 power of the planet mass, which arises from the mass dependence of the Hill radius. As a planet grows, its feeding zone expands, but the mass of new planetesimals within the expanded feeding zone rises more slowly than linearly. We thus obtain the isolation mass by setting the planet mass equal to the mass of the planetesimals in the feeding zone of the original disk,

$$M_{\rm iso} = 4\pi a \cdot C \left(\frac{M_{\rm iso}}{3M_*}\right)^{1/3} a \cdot \Sigma_p \tag{179}$$

which gives,

$$M_{\rm iso} = \frac{8}{\sqrt{3}} \pi^{3/2} C^{3/2} M_*^{-1/2} \Sigma_p^{3/2} a^3. \tag{180}$$

Evaluating this expression in the terrestrial planet region, taking $a = 1$ AU, $\Sigma_p = 10$ gcm^{-2}, $M_* = M_\odot$ and $C = 2\sqrt{3}$ (Lissauer, 1993), we obtain,

$$M_{\rm iso} \simeq 0.07\ M_\oplus. \tag{181}$$

Isolation is therefore likely to occur late in the formation of the terrestrial planets. Repeating the estimate for the conditions appropriate to the formation of Jupiter's core, using $\Sigma_p = 10$ gcm^{-2} as adopted by Pollack et al. (1996)[14], gives,

$$M_{\rm iso} \simeq 9\ M_\oplus. \tag{182}$$

This estimate is comparable to, or larger than, the current best determinations for the mass of the Jovian core (Guillot, 2005). Full isolation may or may not be relevant to the formation of Jupiter, depending upon the adopted disk model.

6. Coagulation equation

One might legitimately question whether the assumption that the mass distribution of growing bodies can be neatly divided into two groups — planetesimals and growing planetary embryos — is any good. The quantitative approach to describing the evolution of a arbitrary size distribution is based on the *coagulation equation* (Smoluchowski, 1916). This allows us to drop the two groups approximation though at the expense of an enormous increase in complexity.

To write the coagulation equation in its simplest form[15], assume that the masses of bodies are integral multiples of some small mass m_1. At time t there are n_k bodies of mass $m_k = km_1$. The coagulation equation in discrete form is,

$$\frac{\mathrm{d}n_k}{\mathrm{d}t} = \frac{1}{2} \sum_{i+j=k} A_{ij} n_i n_j - n_k \sum_{i=1}^{\infty} A_{ki} n_i \tag{183}$$

where A_{ij} is the rate of mergers between bodies of mass m_i and m_j. The first term on the right-hand side of the equation describes the increase in the number of bodies of mass m_k due to collisions of *all possible pairs* of bodies whose masses m_i and m_j sum to m_k. The second term describes the decrease due to bodies of mass m_k being incorporated into even larger bodies. The possibility of fragmentation is here neglected. In this formulation of the problem of planetary growth, all of the physics — such as gravitational focusing — enters via the rate coefficients A_{ij}.

Equation (183), or variants of it, has been used extensively to study planet formation (Inaba et al., 2001;

---

[14] Note that this is a factor of several enhanced above the minimum mass Solar Nebula value.

[15] It is also possible to write the coagulation equation as an integro-differential equation for a continuous mass function $n(m,t)$ (Safronov, 1969), or as a discrete equation where bodies are binned into arbitrary mass intervals (typically logarithmic). Kenyon & Luu (1998) provide a clear description of how the coagulation equation may be formulated and solved in the more general case.

Kenyon & Luu, 1998; Safronov, 1969; Wetherill & Stewart, 1993), either on its own or in combination with direct N-body simulations (Bromley & Kenyon, 2006). Generally the coagulation equation needs to be supplemented with additional equations that describe the evolution of the velocity dispersion as a function of mass, as described for example in Kenyon & Luu (1998). Because of the fact that all i, j such that $m_i + m_j = m_k$ contribute to the evolution of n_k, even the coagulation equation on its own is not a simple equation to deal with, and few analytic solutions are known. One (over)-simple case for which an analytic solution exists is for the case when,

$$A_{ij} = \alpha \tag{184}$$

with α a constant. Then, if the initial state of the system comprises n_1 bodies all of mass m_1, the solution to equation (183) is,

$$\begin{aligned} n_k &= n_1 f^2 (1-f)^{k-1} \\ f &\equiv \frac{1}{1+\frac{1}{2}\alpha n_1 t}. \end{aligned} \tag{185}$$

This solution is shown as Figure 25. The mass spectrum remains smooth and well-behaved as growth proceeds, and with increasing time the characteristic mass increases linearly while maintaining a fixed shape.

More generally, solutions to the coagulation equation fall into two classes (e.g. Lee, 2000):

- Solutions that exhibit *orderly growth*, in which the mass distribution evolves smoothly with time toward higher mean masses. The analytic solution given above for the case A_{ij} = constant is an example of this type of evolution. Another analytic example is $A_{ij} \propto (m_i + m_j)$.
- Solutions that show *runaway growth*. In this case the mass distribution develops a power-law tail toward high masses — physically this corresponds to one or a handful of bodies growing rapidly at the expense of all the others. The long-term validity of the coagulation equation once runaway growth occurs is evidently limited. An analytic example occurs for a rate coefficient $A_{ij} \propto m_i m_j$.

Looking back to equation (169), we note that the rate coefficient is expected to scale as $A \propto R_s^2 \propto m^{2/3}$ in the regime where gravitational focusing is unimportant, and $A \propto R_s^2 v_{\rm esc}^2 \propto m^{4/3}$ once gravitational focusing is dominant. By comparison with the aforementioned analytic solutions, we expect that the initial growth of planetesimals will occur in the orderly regime, while runaway growth may occur once the largest bodies are massive enough for gravitational focusing to become significant.

7. Overview of terrestrial planet formation

We conclude the discussion of terrestrial planet formation by summarizing briefly the main stages of the process:

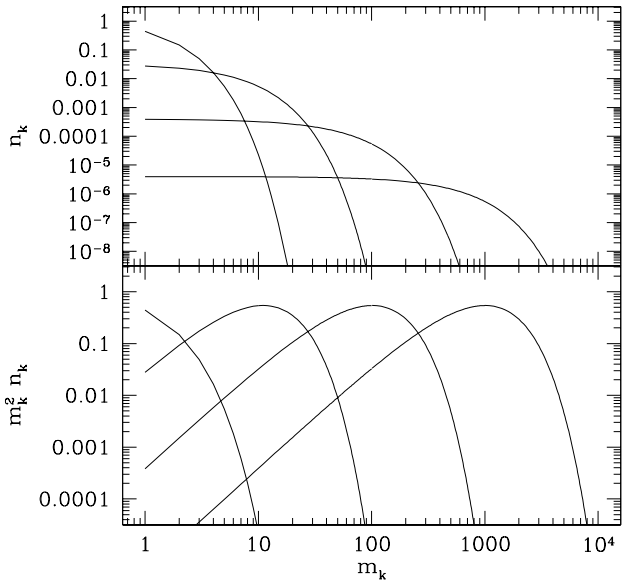

FIG. 25 Illustrative analytic solution to the coagulation equation for the simple case in which $A_{ij} = \alpha$, with α a constant. Initially all bodies have mass m_1. The solution is plotted for scaled times $t' \equiv \alpha n_1 t$ equal to 1, 10, 100 and 10^3. The upper panel shows the number of bodies n_k of each mass (the vertical scale is arbitrary), while the lower panel shows how the mass distribution evolves. This solution is an example of orderly growth — as time progresses the mean mass steadily increases while the shape of the mass spectrum remains fixed.

1. Dust particles agglomerate to form, eventually, planetesimals. Initially this almost certainly occurs via pairwise collisions, though how (or whether) this process can continues to work for cm to meter scale particles remains somewhat murky. Gravitational instability may allow a bypass of these tricky sizes.

2. Growth beyond planetesimals occurs via direct collisions, with an increasing role for gravitational focusing as masses become larger. Dynamical friction keeps the velocity dispersion of the most massive bodies low. A phase of runaway growth occurs in which a few bodies grow rapidly at the expense of the rest.

3. Runaway growth ceases once the largest bodies become massive enough to stir up the planetesimals in their vicinity. A phase of *oligarchic growth* ensues, in which the largest objects grow more slowly than they did during runaway growth, but still more rapidly than small bodies (Kokubo & Ida, 1998; Thommes, Duncan & Levison, 2003). Growth continues in this mode until the isolation mass is approached, at which point growth slows further.

4. Further evolution occurs as a result of collisions

between the initially relatively isolated planetary embryos left over after oligarchic growth. The embryos are perturbed onto crossing orbits due to the influence of the giant planets and mutual secular resonances (Chambers & Wetherill, 1998). The final assembly of the terrestrial planets takes around 100 Myr, with the predicted configuration varying depending upon the assumed surface density of planetesimals and existence (or not) of giant planets (Kokubo, Kominami & Ida, 2006; Levison & Agnor, 2003; Raymond, Quinn & Lunine, 2005). In the Solar System, one of the final impacts on the Earth is widely considered to have given rise to the ejection of enough mass into orbit to subsequently form the Moon (Canup, 2004).

The dominant uncertainties in theoretical models for terrestrial planet formation are arguably found during stage 1 — the formation of planetesimals. It is also true that most simulations of the late stages of terrestrial planet formation lead to planetary properties (such as the eccentricity, and the mass of Mars compared to the other terrestrial planets) that differ somewhat from those observed in the Solar System. Thus, although there is general confidence that the basic physics of terrestrial planet formation is understood, it is clear that current models do not include all of the ingredients needed to accurately match Solar System constraints (Raymond et al., 2009).

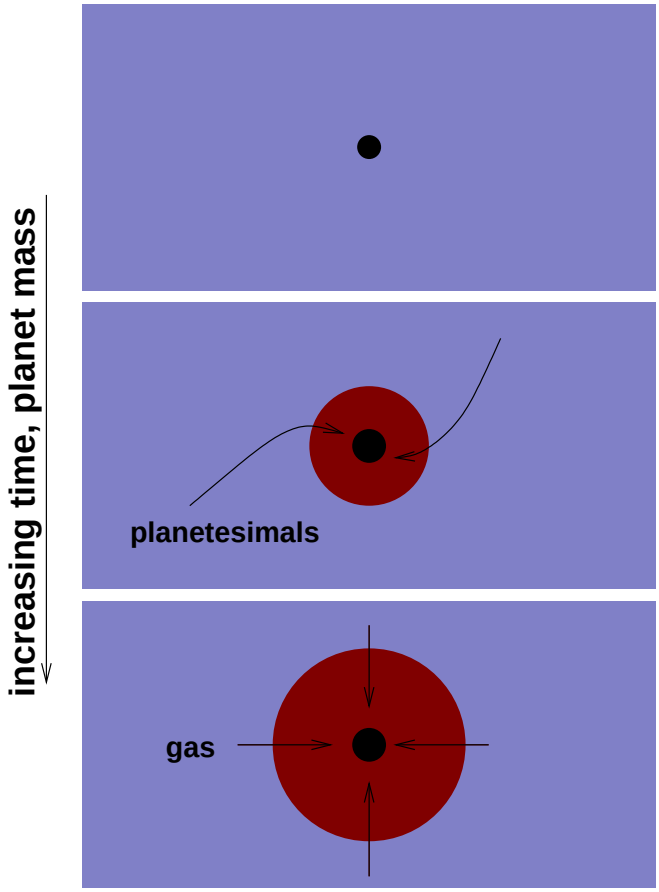

FIG. 26 Illustration of the main stages of the core accretion model for giant planet formation.

C. Gas giant formation

Two theoretical models vie to explain the formation of gas giant planets. The *core accretion model* (Bodenheimer & Pollack, 1986; Mizuno, 1980), developed in its most refined form by Pollack et al. (1996), postulates that the envelopes of gas giants are accreted subsequent to the formation of a large core, which is itself assembled in a manner analogous to terrestrial planet formation. Core accretion is the dominant theory for massive planet formation. The *gravitational instability* model, on the other hand, is based on the idea that a massive protoplanetary disk might collapse directly to form massive planets (Cameron, 1978; Kuiper, 1951). Boss (1997) is the most recent advocate of this idea, which has come under renewed theoretical scrutiny with the discovery of many extrasolar planets with masses much larger than that of Jupiter.

In this Section, we review the physics of these theories in turn. We also discuss the observational constraints on the different theories, which include inferences as to the core masses of the gas giants in the Solar System, the host metallicity / planet frequency correlation for extrasolar planetary systems, and — indirectly — comparison of the theoretically derived time scales with observations of protoplanetary disk lifetimes. This is a critical issue, since gas giants must form prior to the dispersal of the gas disk. Any successful model of massive planet formation must grow such bodies within at most 5-10 Myr (Haisch, Lada & Lada, 2001).

1. Core accretion model

The main stages in the formation of a gas giant via core accretion are illustrated in Figure 26. A core of rock and / or ice forms via the same mechanisms that we have previously outlined for terrestrial planet formation. Initially, there is either no atmosphere at all (because the potential is too shallow to hold on to a bound atmosphere), or any gas is dynamically insignificant. However, as the core grows, eventually it becomes massive enough to hold on to a significant envelope. At first, the envelope is able to maintain hydrostatic equilibrium. The core continues to grow via accretion of planetesimals, and the gravitational potential energy liberated as these planetesimals rain down on the core provides the main source of luminosity. This growth continues until the core reaches a *critical mass*. Once the critical mass is reached, the envelope can no longer be maintained in hydrostatic equilibrium. The envelope contracts on its own Kelvin-Helmholtz time scale, and a phase of rapid gas accretion occurs. This process continues until (a) the

planet becomes massive enough to open up a gap in the protoplanetary disk, thereby slowing down the rate of gas supply, or (b) the gas disk itself is dispersed.

The novel aspect of the core accretion model is the existence of a critical core mass. Mizuno (1980) used numerical models to demonstrate the existence of a maximum core mass, and showed that it depends only weakly on the local properties of the *gas* within the protoplanetary disk. A clear exposition of this type of calculation is given in, for example, Papaloizou & Terquem (1999). Here, following Stevenson (1982), we show that a toy model in which energy transport is due solely to radiative diffusion displays the key property of a critical core mass.

Consider a core of mass $M_{\rm core}$ and radius $R_{\rm core}$, surrounded by a gaseous envelope of mass $M_{\rm env}$. The total mass of the planet,

$$M_t = M_{\rm core} + M_{\rm env}. \tag{186}$$

The envelope extends from $R_{\rm core}$ to some outer radius $R_{\rm out}$, which marks the boundary between the gas bound to the planet and the gas in the protoplanetary disk. $R_{\rm out}$ may be determined by thermal effects (in which case $R_{\rm out} \sim GM_t/c_s^2$, with c_s the disk sound speed) or by tidal considerations (giving an outer radius of r_H), whichever is the smaller. If the envelope is of low mass, then the largest contribution to the luminosity is from accretion of planetesimals onto the core. This yields a luminosity,

$$L = \frac{GM_{\rm core}\dot{M}_{\rm core}}{R_{\rm core}} \tag{187}$$

which is constant through the envelope.

If we assume that radiative diffusion dominates the energy transport, then the structure of the envelope is determined by the equations of hydrostatic equilibrium and radiative diffusion,

$$\frac{{\rm d}P}{{\rm d}r} = -\frac{GM(r)}{r^2}\rho \tag{188}$$

$$\frac{L}{4\pi r^2} = -\frac{16}{3}\frac{\sigma T^3}{\kappa_R \rho}\frac{{\rm d}T}{{\rm d}r} \tag{189}$$

where σ is the Stefan-Boltzmann constant and κ_R the Rosseland mean opacity (assumed constant). Eliminating the density between these equations we find that,

$$\frac{{\rm d}T}{{\rm d}P} = \frac{3\kappa_R L}{64\pi\sigma G M T^3}. \tag{190}$$

We now integrate this equation inward from the outer boundary, making the approximation that $M(r) \approx M_t$ and taking L and κ_R to be constants,

$$\int_{T_{\rm disk}}^{T} T^3 dT = \frac{3\kappa_R L}{64\pi\sigma G M_t}\int_{P_{\rm disk}}^{P} dP. \tag{191}$$

Once we are well inside the planet we expect that $T^4 \gg T_{\rm disk}^4$ and that $P \gg P_{\rm disk}$, so the integral yields, approximately,

$$T^4 \simeq \frac{3}{16\pi}\frac{\kappa_R L}{\sigma G M_t}P. \tag{192}$$

Substituting P in this equation with an ideal gas equation of state,

$$P = \frac{k_B}{\mu m_p}\rho T, \tag{193}$$

we eliminate T^3 in favor of the expression involving ${\rm d}T/{\rm d}r$ and integrate once more with respect to radius to obtain,

$$T \simeq \left(\frac{\mu m_p}{k_B}\right)\frac{GM_t}{4r} \tag{194}$$

$$\rho \simeq \frac{64\pi\sigma}{3\kappa_R L}\left(\frac{\mu m_p G M_t}{4k_B}\right)^4 \frac{1}{r^3}. \tag{195}$$

Having derived the density profile the mass of the envelope follows immediately,

$$\begin{aligned} M_{\rm env} &= \int_{R_{\rm core}}^{R_{\rm out}} 4\pi r^2 \rho(r) dr \\ &= \frac{256\pi^2\sigma}{3\kappa_R L}\left(\frac{\mu m_p G M_t}{4k_B}\right)^4 \ln\left(\frac{R_{\rm out}}{R_{\rm core}}\right). \end{aligned} \tag{196}$$

The right-hand-side of this equation has a strong dependence on the total planet mass M_t and a weaker dependence on the core mass $M_{\rm core}$ via the expression for the luminosity,

$$L = \frac{GM_{\rm core}\dot{M}_{\rm core}}{R_{\rm core}} \propto M_{\rm core}^{2/3}\dot{M}_{\rm core}. \tag{197}$$

In principle there are further dependencies to consider since $R_{\rm out}$ is a function of M_t and $R_{\rm core}$ is a function of $M_{\rm core}$, but these enter only logarithmically and can be safely ignored. Noting that,

$$M_{\rm core} = M_t - M_{\rm env}, \tag{198}$$

we find that,

$$M_{\rm core} = M_t - \left(\frac{C}{\kappa_R \dot{M}_{\rm core}}\right)\frac{M_t^4}{M_{\rm core}^{2/3}}, \tag{199}$$

where we have shown explicitly the dependence on the envelope opacity and planetesimal accretion rate but have swept all the remaining constants (and near-constants) into a single constant C.

Solutions to equation (199) are plotted as Figure 27. One sees that for fixed $\dot{M}_{\rm core}$, there exists a maximum or critical core mass $M_{\rm crit}$ beyond which no solution is possible. The physical interpretation of this result — whose origin is not terribly clear even within this simple model

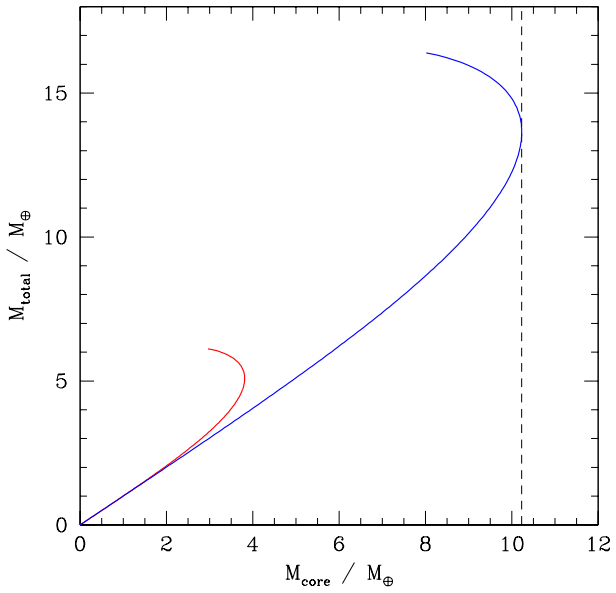

FIG. 27 Solutions to equation (199) for the core mass $M_{\rm core}$ and total mass $M_{\rm total}$. The blue curve is for a higher planetesimal accretion rate than for the red curve. The critical core mass is shown as the vertical dashed line. One should not take solutions to this toy model very seriously, but the numbers have been fixed here to correspond roughly to the values obtained from real calculations.

— is that if one tries to build a planet with a core mass above the critical mass hydrostatic equilibrium cannot be achieved in the envelope. Rather the envelope will contract, and further gas will fall in as fast as gravitational potential energy can be radiated.

This toy model should not be taken too seriously. However, it does illustrate the most important result from more detailed calculations — namely that the critical mass increases with larger $\dot{M}_{\rm core}$ and with enhanced opacity. Ikoma, Nakazawa & Emori (2000) derive an approximate fit to numerical results,

$$\frac{M_{\rm crit}}{M_\oplus} \approx 12 \left(\frac{\dot{M}_{\rm core}}{10^{-6}\ M_\oplus {\rm yr}^{-1}} \right)^{1/4} \left(\frac{\kappa_R}{1\ {\rm cm}^2 {\rm g}^{-1}} \right)^{1/4} \quad (200)$$

where the power-law indices are uncertain by around ± 0.05. The weak dependence of the critical core mass on the planetesimal accretion rate means that, within a particular core accretion model, we can always speed up the approach to runaway gas accretion simply by increasing the assumed surface density of planetesimals in the vicinity of the growing core. Contrary to what is sometimes implied, there is no intrinsic difficulty in building planets quickly via core accretion. However, faster growth occurs *at the expense of* a larger final core mass. As we will shortly note, this tradeoff is of concern since estimates of the core mass of Jupiter are smaller than the values obtained in the classic calculations of core accretion by Pollack et al. (1996).

Although they appear very detailed, extant calculations of planet growth via core accretion should probably be regarded as illustrative rather than definitive. Two sources of uncertainty are particularly worrying:

- **What is the magnitude of the opacity?** Although κ_R enters equation (200) as rather a weak power, its magnitude is highly uncertain. Hubickyj, Bodenheimer & Lissauer (2005) have computed new core accretion models in which the opacity is reduced from the interstellar value by a factor of 50. This allows for much more rapid formation of gas giants than was obtained in prior models. What matters most is the opacity in the upper regions of the envelope. There is some motivation for considering smaller values of the opacity — for example in work by Podolak (2003) — but it remains unclear how accurately we can determine what the appropriate value to use is.

- **The neglect of Type I migration of growing cores**. Theoretical work, which we will discuss more fully in a subsequent Section, suggests that planets or planetary cores with masses exceeding $M_\oplus$ are highly vulnerable to radial migration as a consequence of gravitational torques exerted by the gas disk. This effect is not included in the calculations of Pollack et al. (1996) or Hubickyj, Bodenheimer & Lissauer (2005). Papaloizou & Terquem (1999) and Alibert et al. (2005) have studied the effect of steady inward migration on core formation, and have showed that it makes a large change to the time scale and outcome of the process. Matters could be different again if the migration process is instead unsteady (Rice & Armitage, 2003).

To summarize, the broad outlines of how core accretion works are well established, but further work is needed to delineate under what circumstances (i.e. for what values of the surface density, disk lifetime, migration rates and envelope opacity) it results in successful formation of a massive planet.

2. Gravitational instability model

A sufficiently massive and / or cold gas disk is gravitationally unstable[16]. If — and this is a big if — gravitational instability leads to *fragmentation* this can lead to

[16] The terminology used to discuss this process is potentially confusing. I will use the term *gravitational instability* to refer to disks in which the self-gravity of the gas is significant enough to alter the structure or evolution of the disk. *Fragmentation* refers to the case where gravitational instability leads to the breakup of the disk into bound objects.

massive planet formation (Cameron, 1978; Kuiper, 1951). Durisen et al. (2007) provides a recent review of the status of the gravitational instability model for giant planet formation.

We have already derived the necessary conditions for gravitational instability to occur. We need the Toomre Q parameter to be low enough, specifically,

$$Q \equiv \frac{c_s \Omega}{\pi G \Sigma} < Q_{\rm crit} \simeq 1 \tag{201}$$

where c_s is the sound speed in a gas disk of local surface density Σ and the disk mass is assumed to be small enough that the distinction between the orbital and epicyclic frequencies is of little import. If we consider a disk with $h/r = 0.05$ at 10 AU around a Solar mass star, then the relation $h/r = c_s/v_\phi$ yields a sound speed $c_s \simeq 0.5$ kms^{-1}. To attain $Q = 1$, we then require a surface density,

$$\Sigma \approx 1.5 \times 10^3 \ {\rm gcm}^2. \tag{202}$$

This is much larger than estimates based, for example, on the minimum mass Solar Nebula, from which we conclude robustly that gravitational instability is most likely to occur at an early epoch when the disk mass is still high. Recalling that the characteristic wavelength for gravitational instability is $\lambda_{\rm crit} = 2c_s^2/(G\Sigma)$, we find that the mass of objects formed if such a disk fragmented would be,

$$M_p \sim \pi \Sigma \lambda_{\rm crit}^2 \sim \frac{4\pi c_s^4}{G^2 \Sigma} \sim 5 M_J \tag{203}$$

where M_J is the mass of Jupiter. These order of magnitude estimates suffice to indicate that gravitational instability followed by fragmentation could form gas giants.

It is also straightforward to derive where in the disk gravitational instability is most likely to occur. Noting that in a steady-state accretion disk $\nu\Sigma = \dot{M}/(3\pi)$, we use the α prescription (Shakura & Sunyaev, 1973) and obtain,

$$Q \propto \frac{c_s^3}{\dot{M}}. \tag{204}$$

The sound speed in a protoplanetary disk decreases outward, so a steady-state disk becomes less stable at large radii. Indeed, unless the temperature becomes so low that external irradiation (not that from the central star) dominates the heating, a steady-state disk *will* become gravitational unstable provided only that it is large enough.

To derive sufficient conditions for fragmentation, we need to go beyond these elementary considerations and ask what happens to a massive disk as instability is approached. The critical point is that as Q is reduced, *non-axisymmetric* instabilities set in which do not necessarily lead to fragmentation. Rather, the instabilities generate spiral arms (Laughlin & Bodenheimer, 1994) which both transport angular momentum and lead to dissipation and heating. The dissipation in particular results in heating of the disk, which raises the sound speed and makes fragmentation less likely. On a longer time scale, angular momentum transport also leads to lower surface density and, again, enhanced stability (Lin & Pringle, 1990).

Given these consideration, when will a disk fragment? Gammie (2001) used both analytic arguments and local numerical simulations to identify the *cooling time* as the control parameter determining whether a gravitationally unstable disk will fragment. For an annulus of the disk we can define the equivalent of the Kelvin-Helmholtz time scale for a star,

$$t_{\rm cool} = \frac{U}{2\sigma T_{\rm disk}^4} \tag{205}$$

where U is the thermal energy content of the disk per unit surface area. Then for an ideal gas equation of state with $\gamma = 5/3$ the boundary for fragmentation is:

- $t_{\rm cool} \lesssim 3\Omega^{-1}$ — the disk fragments.
- $t_{\rm cool} \gtrsim 3\Omega^{-1}$ — disk reaches a steady state in which heating due to dissipation of gravitational turbulence balances cooling.

This condition is intuitively reasonable. Spiral arms resulting from disk self-gravity compress patches of gas within the disk on a time scale that is to order of magnitude Ω^{-1}. If cooling occurs on a time scale that is shorter that Ω^{-1}, the heating due to adiabatic compression can be radiated away, and in the absence of extra pressure collapse is likely. It is also worth noting that although the above condition was derived locally, global simulations show that it provides a good approximation to the stability of protoplanetary disks more generally (Rice et al., 2003b). One can also express the fragmentation boundary in terms of a *maximum stress* that a self-gravitating disk can sustain without fragmenting (Gammie, 2001). Writing this in terms of an effective α parameter, $\alpha_{\rm max} \simeq 0.06$ (Rice, Lodato & Armitage, 2005).

None of the above is the subject of much theoretical doubt. Whether a massive protoplanetary disk can fragment into massive planets depends upon its cooling time. What remains controversial is whether the cooling time scale can, in fact, ever be short enough. Analytic arguments suggest that attaining a short enough cooling time scale is difficult (Rafikov, 2005), especially at small orbital radii, and that the most likely scenario for fragmentation would involve very massive planets (or brown dwarfs) forming at radii of the order of 50 or 100 AU. Simulations yield a confused picture. Work by Boley et al. (2006) suggests that model protoplanetary disks fail to fragment, while Boss (2005) and Mayer et al. (2007) obtain fragmentation. The origin of these divergent results remains unclear.

The architecture of the Solar System's giant planets provides qualified support for the core accretion model. The time scale for core accretion increases with orbital radius, which is qualitatively consistent with the general trend of planetary properties in the outer Solar System — Jupiter is closest to Solar composition (reflecting a fully formed gas giant), while Saturn and the ice giants are relatively gas poor. Perhaps these outermost planets formed as the gas disk was in the process of being dispersed. Explaining the origin of Uranus and Neptune as a consequence of disk fragmentation is not easy. Moreover the core accretion time scale for the formation of Jupiter — about 8 Myr in the Pollack et al. (1996) calculation — is reasonable for plausible assumptions. Applying the model to extrasolar planetary systems, we would expect that a greater surface density of planetesimals would lead to faster core growth and an increased chance of reaching runaway prior to disk dispersal. This is consistent with the observed correlation of planet frequency with host metallicity (Fischer & Valenti, 2005; Ida & Lin, 2004). It is currently unclear whether this correlation — which appears to reflect the formation process rather than a metallicity dependence of the migration rate (Livio & Pringle, 2003) — could also be consistent with disk fragmentation.

Solar System observations also raise doubts about core accretion. The time scale to form Neptune, in particular, is prohibitively long. This result is now normally interpreted as an indication that Uranus and Neptune may not have formed in situ, and as such cannot be used to argue against core accretion. It means, however, that the ice giants are poor laboratories for testing core accretion. Potentially more seriously, measurements of the gravitational multipole moments of Jupiter (from the *Galileo* orbiter) can be combined with interior structure models to yield constraints on the core mass. Until recently there was an unambiguous discrepancy between the resulting constraints and the predictions of core accretion. Guillot (2005), for example, obtained an upper limit on the core mass of Jupiter of $10M_\oplus$ for the most optimistic choice of equation of state (optimistic in the sense of yielding the weakest constraints). For some equations of state the constraint on the core mass was below $5M_\oplus$[17]. This was evidently smaller than predictions based on the simplest models of core accretion (Pollack et al., 1996), and in complete agreement with the zero core prediction of disk instability.

This may sound like a clear strike against core accretion, but in fact matters are not so simple. First, as we have already noted fiducial core accretion models are based on particular choices of uncertain parameters and as such should not be regarded as definitive. Currently, it seems reasonable to believe that smaller core masses — perhaps as low as $5M_\oplus$ — could be consistent with plausible variants of the basic core accretion model (Alibert et al., 2005b; Hubickyj, Bodenheimer & Lissauer, 2005). Second, the 'observational' constraint on the Jovian core mass is highly indirect. It requires the adoption both of an equation of state and of an assumed interior structure model whose parameters are then optimized by comparing the model against the actual measured data. There has been substantial recent progress in determining the high pressure equation of state via *ab initio* methods, but alas this has not eliminated the uncertainty in the quantity we are interested in – Jupiter's core mass. Recently Militzer et al. (2008) have presented Jovian models that include a substantial ($14 - 18\ M_\oplus$) core, while Nettelmann et al. (2008) have computed similarly state-of-the-art models that support the Guillot (2005) conclusion that any core must be small. The differences appear to result primarily from different assumptions made by the two groups as to the number of distinct layers within the interior of Jupiter.

One may hope that a combination of additional data and further modeling will eventually result in a clear measurement of the Jovian core mass. If Jupiter is genuinely devoid of a core, then that would spell serious trouble for core accretion. Similarly the discovery of massive planets at very large orbital radii — where disk instability is most likely and the time scale for core accretion very large — would support fragmentation models, though it may be hard to rule out the possibility that any such planets formed closer to the star and migrated outward (Veras & Armitage, 2004).

Although we have phrased this discussion in terms of either core accretion or disk fragmentation providing a mechanism for massive planet formation, it of course remains possible that both could be viable formation channels. If so, the most likely scenario would see core accretion forming lower mass planets at small orbital radii, while gravitational instability would yield very massive planets typically further out. The existence of two channels could be inferred, for example, by looking for *different* metallicity distributions of stars hosting high and low mass planets (Rice et al., 2003c), or, even more speculatively, by searching for different eccentricity distributions. Marcy et al. (2005), Rice et al. (2003b) and Ribas & Miralda-Escude (2007) have noted that the existence of two populations could be consistent with existing observational data, though the statistical significance of any differences between the properties of high and low mass planet systems is currently weak, being limited by the small number of known high mass planets and is weak.

IV. EVOLUTION OF PLANETARY SYSTEMS

The story is not over once planets have managed to form. Theoretical models, which are now strongly sup-

[17] The same exercise yielded a core mass for Saturn of 10-20$M_\oplus$, in good accord with the expectations of core accretion

ported by observations of the Solar System and of extrasolar planetary systems, suggest at least three mechanisms that can lead to substantial post-formation orbital evolution:

- **Interaction between planets and the gaseous protoplanetary disk**. This leads to orbital *migration* (Goldreich & Tremaine, 1980) as a consequence of angular momentum exchange between the planet and the gas disk, and can be important for both terrestrial-mass planets and gas giants while the gas disk is still present. Gas disk migration provides the standard theoretical explanation for the existence of hot Jupiters (Lin, Bodenheimer & Richardson, 1996).
- **Interaction between planets and a remnant planetesimal disk**. Planets, especially giant planets, can also exchange angular momentum by interacting with and ejecting planetesimals left over from the planet formation process. This mechanism appears likely to have caused orbital migration of at least the ice giants, and possibly also Saturn, during the early history of the Solar System (Levison et al., 2007).
- **Interaction within an initially unstable system of two or more massive planets**. There is no guarantee that the architecture of a newly formed planetary system will be stable over the long run. Instabilities can lead to planet-planet scattering, which usually results in the ejection of the lower mass planets, leaving the survivors on eccentric orbits. This could be the origin of the typically eccentric orbits seen in extrasolar planetary systems (Lin & Ida, 1997; Rasio & Ford, 1996; Weidenschilling & Marzari, 1996).

In this Section I discuss each of these mechanisms in turn. The focus here is exclusively on *dynamical evolution* of newly formed planetary systems. Of course the *surface* properties of planets also evolve with time, even in the absence of orbital evolution, due to changes in the stellar luminosity and geological processes. Considerations of this kind, which are crucial to determining the habitability of planets, are discussed for example in Kasting, Whitmire & Reynolds (1993).

A. Gas disk migration

The most detailed calculations of the rate of angular momentum exchange between a planet and a gas disk are based on summing the torques exerted at discrete *resonances* within the disk. This calculation, which was introduced in the planetary context in a paper by Goldreich & Tremaine (1979), is too lengthy and technical to reproduce here. The reader interested in the gory details is advised to start with the most refined recent calculation by Tanaka, Takeuchi & Ward (2002), which includes pointers to the earlier literature (it is important to note that there have been significant improvements to the original calculation of Goldreich & Tremaine (1979), which are particularly important for the case of low mass planets). Here we summarize the conditions for resonances to exist, and discuss the effect of the torques on the planet and on the disk in the limits of high and low planet masses.

1. Conditions for resonance

We consider a planet orbiting a star on a circular orbit with angular frequency Ω_p. A standard exercise in dynamics (e.g. Binney & Tremaine 1987) yields the conditions for resonances. A *corotation resonance* exists for radii in the disk where the angular frequency Ω,

$$\Omega = \Omega_p. \tag{206}$$

Lindblad resonances exist when,

$$m(\Omega - \Omega_p) = \pm\kappa_0 \tag{207}$$

where m is an integer and κ_0, the *epicyclic frequency*, is defined as,

$$\kappa_0 \equiv \left(\frac{\mathrm{d}^2\Phi_0}{\mathrm{d}r^2} + 3\Omega^2\right) \tag{208}$$

with Φ_0 the stellar gravitational potential. For a Keplerian potential $\kappa_0 = \Omega$. If we approximate the angular velocity of gas in the disk by the Keplerian angular velocity, the Lindlad resonances are located at,

$$r_L = \left(1 \pm \frac{1}{m}\right)^{2/3} r_p \tag{209}$$

where r_p is the planet orbital radius. The locations of some of the low order (small m) resonances are shown in Figure 28. For an orbiting test particle, the resonances are locations where the planet can be a strong perturbation to the motion. For a gas disk, angular momentum exchange between the planet and the gas disk occurs at resonant locations.

As long as we are only concerned with planets on circular orbits, the above discussion tells us all that we need to know about the important resonances. For a planet on an eccentric orbit[18], however, matters become much

[18] Note that it is also possible for the *disk* to be eccentric. Indeed, the interaction between the disk and a massive planet (or brown dwarf) on an eccentric orbit is likely to result in non-zero disk eccentricity (Papaloizou, Nelson & Masset, 2001). In principle, the protoplanetary disk could also develop eccentricity even in the absence of external forcing. Latter & Ogilvie (2006), and references therein, discuss this possibility, and show that the rate (and sign) of eccentricity evolution depends upon the magnitude and nature of the turbulent stress within the disk. A three-dimensional treatment is found to be necessary to capture the relevant physics.

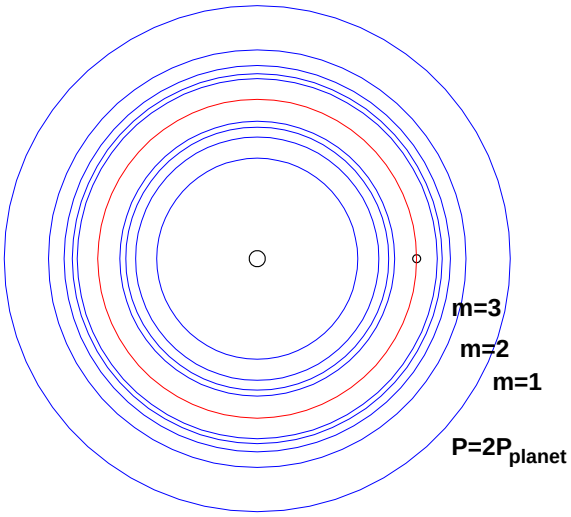

FIG. 28 Nominal locations of the corotation (red) and Lindblad resonances (blue) for a planet on a circular orbit. Only the low order Lindblad resonances are depicted — there are many more closer to the planet.

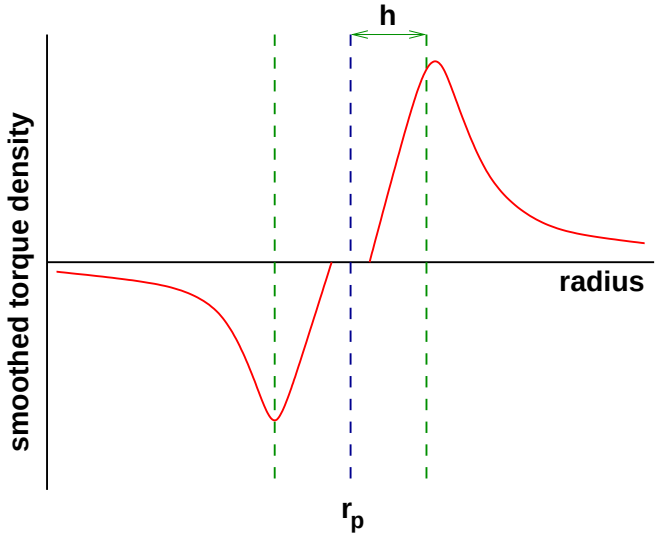

FIG. 29 Schematic illustration of the smoothed torque density due to angular momentum exchange between a planet and a gas disk at the location of Lindblad resonances, after Ward (1997). The peak torque occurs at $r \approx r_p \pm h$. The disk gains angular momentum from the planet as a result of the interaction for $r > r_p$, and loses angular momentum for $r < r_p$. The interaction is almost invariably asymmetric, such that when integrated over the entire disk the planet loses angular momentum and migrates inward.

more complicated (Artymowicz & Lubow, 1994). In this case the potential perturbation due to the planet (which now does *not* rotate at constant angular velocity) can be decomposed into independent components that rotate at constant pattern speed,

$$\Omega_{\rm pattern} = \left(\frac{l}{m}\right)\Omega_p \tag{210}$$

with l an integer – i.e there is now forcing at the harmonics of the mean motion angular frequency. The conditions for corotation and Lindblad resonances remain unaltered with the replacement of the mean angular frequency with the pattern speed,

$$\Omega = \Omega_{\rm pattern} \tag{211}$$
$$m(\Omega - \Omega_{\rm pattern}) = \pm\kappa_0. \tag{212}$$

Aside from the general proliferation of resonances implied by these equations, the important point to note is that for an eccentric planet there are *non-coorbital* corotation resonances.

2. Gravitational torques at resonances

For a planet on a circular orbit embedded within a geometrically thin protoplanetary disk, angular momentum exchange between the planet and the gas occurs at the location of the resonances defined by equation (206) and (207). The sense of the exchange is that,

- The planet **gains angular momentum** from interacting with the gas disk at the interior Lindblad resonances ($r_L < r_p$). This tends to drive the planet outward. The gas loses angular momentum, and moves inward.
- The planet **loses angular momentum** from interacting with the gas disk at exterior Lindblad resonances ($r_L > r_p$). This tends to drive the planet toward the star. The gas gains angular momentum, and moves outward.

Notice that the gravitational interaction of a planet with a gas disk tends — somewhat counter-intuitively — to *repel* gas from the vicinity of the planet's orbit.

The flux of angular momentum exchanged at each Lindblad resonance can be written as,

$$T_{LR}(m) \propto \Sigma M_p^2 f_c(\xi) \tag{213}$$

where Σ is the gas density and M_p the planet mass. That the torque should scale with the *square* of the planet mass is intuitively reasonable — the perturbation to the disk surface density scales as the planet mass in the linear regime so the torque scales as M_p^2. The factor $f_c(\xi)$ is the *torque cutoff function* (Artymowicz, 1993), which encodes the fact that resonances very close to the planet contribute little to the net torque. The torque cutoff function peaks at,

$$\xi \equiv m\left(\frac{c_s}{r\Omega}\right)_p \simeq 1 \tag{214}$$

i.e. at a radial location $r \simeq r_p \pm h$, where h is the disk scale height (this result immediately implies that a three-dimensional treatment is necessary for the dominant resonances if the planet is completely embedded within a gas disk, as is the case for low mass planets). The strength of the torque exerted at each resonance is essentially independent of properties of the disk such as the disk viscosity, though of course the viscosity still matters inasmuch as it controls the value of the unperturbed disk surface density Σ.

Figure 29 illustrates the differential torque exerted on the disk by the planet, after smoothing over the Lindblad resonances (Ward, 1997). The flux of angular momentum is initially deposited in the disk as waves, which propagate radially before dissipating. The details of this dissipation matter little for the net rate of angular momentum exchange.

Angular momentum transfer at the corotation resonance requires additional consideration. In a two-dimensional disk, the rate of angular momentum deposition at corotation is proportional to (Goldreich & Tremaine, 1979; Tanaka, Takeuchi & Ward, 2002),

$$T_{CR} \propto \frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{\Sigma}{B}\right) \tag{215}$$

where B is the Oort parameter,

$$B(r) = \Omega + \frac{r}{2}\frac{\mathrm{d}\Omega}{\mathrm{d}r}. \tag{216}$$

This implies that in a two-dimensional disk, the corotation torque vanishes identically in the moderately interesting case of a disk with a surface density profile $\Sigma \propto r^{-3/2}$. This result does *not apply* to three-dimensional disks (Tanaka, Takeuchi & Ward, 2002), however, so corotation torques do need to be considered whenever there is gas present close to the planet's location.

3. Type I migration

For low mass planets (generically $M_p \sim M_\oplus$, though the exact mass depends upon the disk properties) the angular momentum flux injected into the disk as a consequence of the planet-disk interaction is negligible when compared to the viscous transport of angular momentum. As a result, the gas surface density profile $\Sigma(r)$ remains approximately unperturbed, gas is present at the location of each of the resonances, and the net torque on the planet is obtained by summing up the torque exerted at each resonance. Schematically,

$$T_{\rm planet} = \sum_{ILR} T_{LR} + \sum_{OLR} T_{LR} + T_{CR} \tag{217}$$

where the planet gains angular momentum from the inner Lindblad resonances (ILR) and loses angular momentum to the outer Lindblad resonances (OLR). Changes to the planet's orbit as a result of this net torque are called **Type I migration** (Ward, 1997).

As noted above (equation 213) the torque exerted at each resonance scales as the planet mass squared. If the azimuthally averaged surface density profile of the gas disk remains unperturbed, it follows that the total torque will also scale as M_p^2 and the migration time scale,

$$\tau_I \propto \frac{M_p}{T_{\rm planet}} \propto M_p^{-1}. \tag{218}$$

Type I migration is therefore most rapid for the largest body for which the assumption that the gas disk remains unaffected by the planet remains valid.

Actually evaluating the sum sketched out in equation (217) is not easy, and there is no simple physical argument that I am aware of that gives even the *sign* of the net torque on the planet. However invariably it is found that the Lindblad resonances exterior to the planet are more powerful than those interior, so that the net torque due to Lindblad resonances leads to *inward* migration. The torque at corotation is of comparable magnitude to the net Lindblad torque, but does not reverse the sense of migration (Tanaka, Takeuchi & Ward, 2002; Ward, 1997). Note that one might think (for example by looking at the surface density dependence of the torque in equation 213) that the sense of migration ought to depend upon the surface density gradient — i.e. that a steep surface density profile should strengthen the inner resonances relative to the outer ones and hence drive outward migration. This *is not true*. Pressure gradients (which depend upon the radial dependence of the surface density and temperature) alter the angular velocity in the disk from its Keplerian value (equation 114), and thereby shift the radial location of resonances from their nominal positions. A steep surface density profile implies a large pressure gradient, so that all the resonances move slightly inward. This weakens the inner Lindblad resonance relative to the outer ones, in such a way that the final dependence on the surface density profile is surprisingly weak (Ward, 1997).

Tanaka, Takeuchi & Ward (2002) compute the net torque on a planet in a three-dimensional but isothermal gas disk, including the effect of both Lindblad and corotation resonances. For a disk in which,

$$\Sigma(r) \propto r^{-\gamma} \tag{219}$$

they obtain a migration time scale,

$$\begin{aligned} \tau_I &\equiv -\frac{r_p}{\dot{r}_p} \\ &= (2.7+1.1\gamma)^{-1}\frac{M_*}{M_p}\frac{M_*}{\Sigma_p r_p^2}\left(\frac{c_s}{r_p\Omega_p}\right)^2 \Omega_p^{-1}, \end{aligned} \tag{220}$$

where Σ_p, c_s and Ω_p are respectively the gas surface density, gas sound speed, and angular velocity at the location of a planet orbiting at distance r_p from a star of mass M_*. As expected based on the simple considerations discussed previously, the migration rate ($\propto \tau_I^{-1}$)

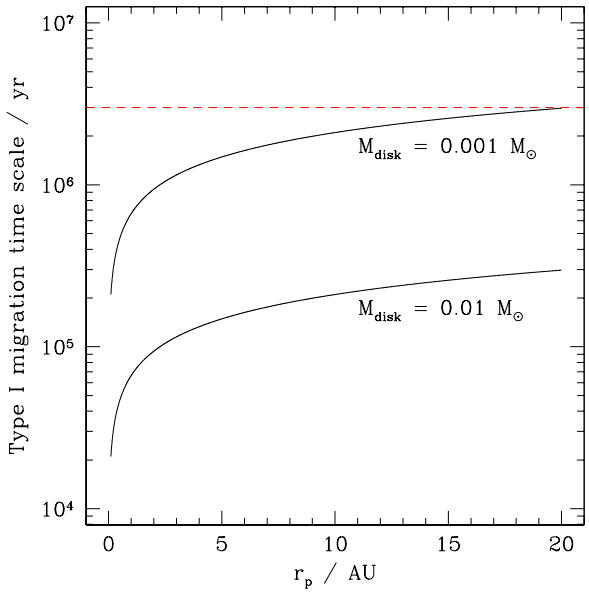

FIG. 30 The inward Type I migration time scale for a $5M_\oplus$ core as a function of orbital radius, calculated using the three-dimensional isothermal disk formula of Tanaka, Takeuchi & Ward (2002). The lower curve assumes a disk with $\Sigma \propto r^{-1}$, $h/r = 0.05$, and a total mass of $0.01M_\odot$ within 30 AU. The upper curve shows the migration time scale in a similar disk with a mass of only $0.001M_\oplus$ — the absolute minimum needed to form a Jupiter mass planet. The red dashed line illustrates a typical estimate for the lifetime of the gas disk.

scales linearly with both the planet mass and the local disk mass. The time scale becomes shorter for cooler, thinner disks — provided that the interaction remains in the Type I regime — since for such disks more resonances close to the planet contribute to the net torque.

Figure 30 shows the predicted migration time scale as a function of radius for a $5M_\oplus$ core in a disk with $h/r = 0.05$ and $\Sigma \propto r^{-1}$. A core of 5 Earth masses is considered since recent numerical simulations by Masset, D'Angelo & Kley (2006) suggest that non-linear effects set in at about this mass and reduce the net torque as compared to the linear prediction. Two disk masses are plotted, one in which the disk mass interior to 30 AU is $10^{-2}M_\odot$, and one in which the disk mass is $10^{-3}M_\odot$. As is obvious from the figure, the migration time scale from radii close to the snow line is a small fraction of the disk lifetime (Haisch, Lada & Lada, 2001) for the more massive disk model. One concludes that — unless the torque calculation is missing essential physics that qualitatively changes the answer — Type I migration is likely to be an essential element of giant planet formation via core accretion. Only if the disk mass is very low (almost the absolute minimum needed to form a gas giant at all) can the effects of Type I migration be reduced. It may be that achieving successful planet formation via core accretion requires waiting until the gas disk is weak enough

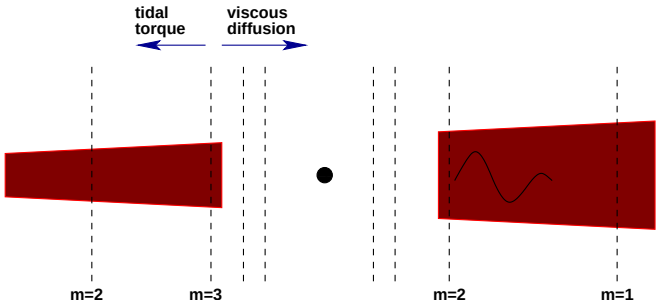

FIG. 31 Illustration of the viscous condition for gap opening. A gap can open when the time scale for opening a gap of width Δr due to tidal torques becomes shorter than the time scale on which viscous diffusion can refill the gap.

to slow Type I migration to a manageable rate. Simple models of this kind have been presented by Thommes & Murray (2006).

4. Type II migration

For sufficiently large planet masses, the angular momentum flux from the planet locally dominates the viscous flux. As a consequence, gas is repelled from high-m resonances. The surface density drops near $r = r_p$, forming a *gap* — an annular region in which the surface density is smaller than its unperturbed value.

Two conditions are necessary for gap formation. First, the Hill sphere (or Roche radius) of the planet needs to be comparable to the thickness of the gas disk,

$$r_H \equiv \left(\frac{M_p}{3M_*}\right)^{1/3} r \gtrsim h \tag{221}$$

which requires a mass ratio $q \equiv M_p/M_*$,

$$q \gtrsim 3\left(\frac{h}{r}\right)_p^3. \tag{222}$$

This condition is satisfied for typical protoplanetary disk parameters for $q \sim 4 \times 10^{-4}$ — i.e. for planet masses somewhere between that of Saturn and Jupiter.

A second condition for gap opening arises due to the viscous considerations depicted in Figure 31. To open a gap, we require that the tidal torques must be able to remove gas from the gap region faster than viscosity can fill the gap back in (Goldreich & Tremaine, 1980; Lin & Papaloizou, 1980; Papaloizou & Lin, 1984). There are various ways to estimate the critical mass above which this condition is satisfied. Following Takeuchi, Miyama & Lin (1996), we note that the time scale for viscous diffusion to close a gap of width Δr is just,

$$t_{\rm close} \sim \frac{(\Delta r)^2}{\nu} \tag{223}$$

where $\nu = \alpha c_s h$ is the disk viscosity. The time scale to open a gap as a result of the tidal torque at an m-th

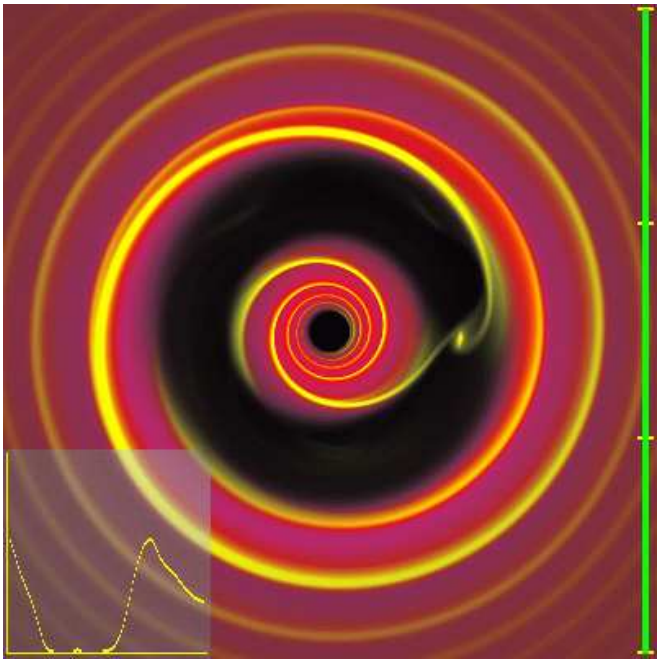

FIG. 32 Simulation of the planet-disk interaction in the Type II regime in which the planet is sufficiently massive to open a gap in the gas disk. Note the presence of streams of gas that penetrate the gap region. The inset plot shows the azimuthally averaged disk surface density. A movie showing the interaction as a function of mass is available at http://jilawww.colorado.edu/∼pja/planet_migration.html.

order Lindblad resonance is,

$$t_{\rm open} \sim \frac{1}{m^2 q^2 \Omega_p} \left(\frac{\Delta r}{r_p}\right)^2. \tag{224}$$

Setting $t_{\rm open} = t_{\rm close}$, and taking $m = r_p\Omega_p/c_s$ (since, as noted above, this value of m is where the torque cutoff function peaks), we obtain,

$$q \gtrsim \left(\frac{c_s}{r_p\Omega_p}\right)^2 \alpha^{1/2}. \tag{225}$$

For typical disk parameters ($h/r = 0.05$, $\alpha = 10^{-2}$), the viscous condition for gap opening is satisfied for q modestly larger than 10^{-4}. Combined with the thermal condition outlined above, we conclude that Jupiter mass planets ought to be massive enough to open a gap within the disk, whereas Saturn mass planets are close to the critical mass required for gap opening. Figure 32 from Armitage & Rice (2005), shows results from a two-dimensional simulation of the planet-disk interaction in the Type II regime. Both the gap, and the presence of a prominent spiral wave excited within the gas disk, are obvious.

5. The Type II migration rate

Once a planet becomes massive enough to open a gap, orbital evolution is predicted to occur on the same local time scale as the protoplanetary disk. The radial velocity of gas in the disk is,

$$v_r = -\frac{\dot{M}}{2\pi r\Sigma}, \tag{226}$$

which for a steady disk away from the boundaries can be written as,

$$v_r = -\frac{3}{2}\frac{\nu}{r}. \tag{227}$$

If the planet enforces a rigid tidal barrier at the outer edge of the gap, then evolution of the disk will force the orbit to shrink at a rate $\dot{r}_p \simeq v_r$, provided that the local disk mass exceeds the planet mass, i.e. that $\pi r_p^2\Sigma \gtrsim M_p$. This implies a nominal Type II migration time scale, valid for *disk dominated migration* only,

$$\tau_0 = \frac{2}{3\alpha}\left(\frac{h}{r}\right)_p^{-2} \Omega_p^{-1}. \tag{228}$$

For $h/r = 0.05$ and $\alpha = 10^{-2}$, the migration time scale at 5 AU is of the order of 0.5 Myr.

In practice, the assumption that the local disk mass exceeds that of the planet often fails. For example, a plausible model of the protoplanetary disk with a mass of 0.01 $M_\odot$ within 30 AU has a surface density profile,

$$\Sigma = 470\left(\frac{r}{1\ {\rm AU}}\right)^{-1}\ {\rm gcm}^{-2}. \tag{229}$$

The condition that $\pi r_p^2\Sigma = M_p$ gives an estimate of the radius within which disk domination ceases of,

$$r = 6\left(\frac{M_p}{M_J}\right)\ {\rm AU}. \tag{230}$$

Interior to this radius, the planet acts as a slowly moving barrier which impedes the inflow of disk gas. Gas piles up behind the barrier – increasing the torque – but this process does not continue without limit because the interaction also deposits angular momentum into the disk, causing it to expand (Pringle, 1991). The end result is to slow migration compared to the nominal rate quoted above. Given a disk model, and in particular a specification of how the angular momentum transport efficiency depends upon the radius and surface density within the disk, the extent of the suppression can be calculated (Ivanov, Papaloizou & Polnarev, 1999; Syer & Clarke, 1995). To illustrate the idea, imagine that we have a disk in which the surface density can be written as a power-law in accretion rate and radius,

$$\Sigma \propto \dot{M}^a r^b, \tag{231}$$

Syer & Clarke (1995) define a measure of the degree of disk dominance,

$$B \equiv \frac{4\pi r_p^2 \Sigma}{M_p}. \quad (232)$$

Then for $B < 1$ (the planet dominated case appropriate to small radii) the actual Type II migration rate is (Syer & Clarke, 1995),

$$\tau_{II} = \tau_0 B^{-a/(1+a)}. \quad (233)$$

Note that with this definition of B, disk dominance extends inward a factor of a few further than would be predicted based on the simple estimate given above.

In practice evaluating how the surface density depends upon the accretion rate – and thereby determining the a which enters into the suppression term – is rather difficult, since it requires us to place perhaps unwarranted faith in aspects of disk models which are immune from observational tests. Proceeding anyway, for the disk models of Bell et al. (1997) we find that $a \simeq 0.5$ at 1 AU for $\dot{M} \sim 10^{-8}\ M_\odot \mathrm{yr}^{-1}$. At this radius the model with $\alpha = 10^{-2}$ has a surface density of about 200 gcm^{-2}. For a Jupiter mass planet we then have $B = 0.3$ and $\tau_{II} = 1.5\tau_0$. This is a modest suppression of the Type II rate, but the effect becomes larger at smaller radii (or for more massive planets). It slows the inward drift of massive planets, and allows a greater chance for them to be stranded at sub-AU scales as the gas disk is dissipated. Detailed models suggest that the distribution of massive extrasolar planets is consistent with those planets forming at larger radii, before becoming stranded during the migration process due to the dispersal of the gas disk (Armitage, 2007; Armitage et al., 2002).

These estimates of the Type II migration velocity assume that once a gap has been opened, the planet maintains an impermeable tidal barrier to gas inflow. In fact, simulations show that planets are able to accrete gas via tidal streams that bridge the gap (Lubow, Siebert & Artymowicz, 1999). The effect is particularly pronounced for planets only just massive enough to open a gap in the first place. If the accreted gas does not have the same specific angular momentum as the planet, this constitutes an additional accretion torque in addition to the resonant torque.

6. Stochastic migration

To a first approximation the efficiency of angular momentum transport has little impact on the predicted Type I migration rate. This assumes, however, that the disk is laminar. More realistically, angular momentum transport itself derives from turbulence, which is accompanied by a spatially and temporally varying pattern of density fluctuations in the protoplanetary disk. These fluctuations will exert *random* torques on planets of any mass embedded within the disk, in much the same way as transient spiral features in the Galactic disk act to increase the velocity dispersion of stellar populations (Carlberg & Sellwood, 1985). If we assume that the random torques are uncorrelated with the presence of a planet, then the random torques' linear scaling with planet mass will dominate over the usual Type I torque (scaling as M_p^2) for sufficiently low masses. The turbulence will then act to increase the velocity dispersion of collisionless bodies, or, in the presence of damping, to drive a random walk in the semi-major axis of low mass planets.

To go beyond such generalities, and in particular to estimate the crossover mass between stochastic and Type I migration, we need to specify the source of turbulence in the protoplanetary disk. MHD disk turbulence driven by the magnetorotational instability (Balbus & Hawley, 1998) provides a well-understood source of outward angular momentum transport in sufficiently well-ionized disks, and has been used as a model system for studying stochastic migration by several authors (Laughlin, Steinacker & Adams, 2004; Nelson, 2005; Nelson & Papaloizou, 2004; Yang, Mac Low & Menou, 2009). Density fluctuations in MHD disk turbulence have a typical coherence time of the order of the orbital period, and as a consequence are able to exchange angular momentum with an embedded planet across a range of disk radii (not only at narrow resonances). The study by Nelson & Papaloizou (2004) was based on both global ideal MHD disk simulations, with an aspect ratio of $h/r = 0.07$, and local shearing box calculations. The global runs realized an effective Shakura-Sunyaev $\alpha = 7 \times 10^{-3}$, which if replicated in a real disk would be consistent with observational measures of T Tauri disk evolution (Hartmann et al., 1998). For all masses considered in the range $3\ M_\oplus \leq M_p \leq 30\ M_\oplus$, the *instantaneous* torque on the planet from the MHD turbulent disk exhibited large fluctuations in both magnitude and sign. Averaging over ≈ 20 orbital periods, the mean torque showed signs of converging to the Type I rate, although the rate of convergence was slow, especially for the lowest mass planets in the global runs. These properties are generally in accord with other studies of the variability of MHD disk turbulence (Hawley, 2001; Winters, Balbus & Hawley, 2003). Very roughly, the Nelson & Papaloizou (2004) simulations suggest that up to $M_p \sim 10\ M_\oplus$ the random walk component dominates steady Type I drift over time scales that substantially exceed the orbital period.

How important stochastic (or diffusive) migration is for planet formation depends, first and foremost, on the strength and nature of the disk turbulence. Many existing studies are based on the properties of turbulence simulated under ideal MHD conditions, which as we noted earlier do not apply to protoplanetary disks. The magnitude of stochastic migration would certainly be reduced – though not eliminated – within disks that have a layered structure (Oishi, Mac Low & Menou, 2007). Strong stochastic migration would pump the mean eccentricity (and perhaps inclination) of planetesimals, reducing the magnitude of gravitational focusing and potentially lead-

ing to a greater likelihood of disruptive collisions (Ida, Guillot & Morbidelli, 2008). For low mass planets the impact of stochastic migration would be to modify their survival prospects compared to what would be expected from ordinary Type I migration (Johnson, Goodman & Menou, 2006). Finally, *if* stochastic migration still matters for rather more massive bodies ($M_p \gtrsim M_\oplus$) then it will affect the formation time scale of giant planet cores (Rice & Armitage, 2003).

7. Eccentricity evolution during migration

Most massive extrasolar planets are observed to be on significantly eccentric orbits. Since orbital migration due to planet-disk interactions is likely to have occurred in these systems, it is of interest to ask whether the same process — gravitational interactions between the gas disk and an orbiting planet in the Type II regime — also leads to excitation of eccentricity. No-one knows for sure.

The basic considerations relevant to this problem were set out in Goldreich & Tremaine (1980). Eccentricity growth (or decay) depends upon the relative strength of:

- External Lindblad resonances, which act to excite eccentricity.
- Non-co-orbital corotation resonances, which act to damp eccentricity. As noted above, the only corotation resonance that exists for a planet on a circular orbit is co-orbital, so a finite eccentricity is necessary for these resonances to be present.

Unfortunately, the effects leading to damping and excitation of eccentricity are finely balanced, making robust assessment of the sign of the eccentricity evolution difficult. The simplest analytic estimates favor damping, but only modest saturation of the corotation resonances would be needed to tilt the balance in favor of excitation (Goldreich & Sari, 2003; Masset & Ogilvie, 2004; Ogilvie & Lubow, 2003). Numerically, there is general agreement that substellar objects of brown dwarf mass and above suffer substantial eccentricity growth when embedded within a gas disk (Artymowicz et al., 1991; Papaloizou, Nelson & Masset, 2001), while more modest excitation has been reported for Jovian mass planets by D'Angelo, Lubow & Bate (2006). These results, along with related analytic considerations discussed by Moorhead & Adams (2008), suggest that although some eccentricity excitation may attend Type II migration it is unlikely that the high e tail of the observed exoplanet distribution derives solely from this process.

8. Observational evidence for inner holes

No current observational facility has the ability to directly image the annular gaps or inner holes that are predicted to result from the interaction of a planet with the protoplanetary disk, though *ALMA* has a shot at being able to do so (Wolf & D'Angelo, 2005). There are, however, a number of T Tauri stars whose spectral energy distributions (SEDs) exhibit robust excesses in the mid-IR (indicative of gas and dust disks at AU scales) *without* matching excesses in the near-IR (Sicilia-Aguilar et al., 2006). Well-known examples of such *inner hole* or *transition* sources include GM Aur (Calvet et al., 2005), TW Hya (Eisner, Chiang & Hillenbrand, 2006) and CoKu Tau/4 (Forrest et al., 2004). By definition, these sources lack optically thick inner disks, from which one deduces that small grains are absent close to the star, though disks are unquestionably present at larger radii.

What is going on in inner hole sources? Some may genuinely be stars caught in the act of dispersing their disks — perhaps as a result of the photoevaporative mechanism discussed earlier in these notes. Others, however, may be 'normal' Classical T Tauri stars around which an orbiting planet has created a tidal barrier to the inflow of gas and dust, thereby creating an inner hole. Theoretical studies suggest that models that invoke the presence of planets can fit the observed SEDs (Quillen et al., 2004; Rice et al., 2003d), though it is unlikely that this interpretation is unique or that it can be proved beyond a reasonable doubt without spatially resolved observations. Interestingly, some of the inner hole sources exhibit significant levels of *gas* accretion (indicating that the hole is only in the dust component), and these objects — which are inconsistent with a photoevaporation model — are probably the best candidates for harboring planets (Alexander & Armitage, 2007).

B. Planetesimal disk migration

It is unlikely that the formation of gas and ice giant planets consumes the entire inventory of planetesimals in their vicinity. The interaction of any remnant planetesimals with planets, after the dispersal of the gas disk, can result in orbital migration of the planets.

Here, we follow the simple discussion of Malhotra (1995)[19]. If we consider a single planetesimal of mass δm interacting with a planet of mass M_p at orbital radius a there are two possible outcomes,

- The planetesimal may be scattered outward — possibly sufficiently to be ejected — in which case the planet moves in by angular momentum conservation. Up to numerical factors,

$$\frac{\delta a}{a} \simeq -\frac{\delta m}{M_p}. \tag{234}$$

[19] The treatment here is deliberately over-simplified. The reader interested in exploring more realistic analytic and numerical models is advised to consult Ida et al. (2000) and Kirsh et al. (2009), and references therein.

- The scattering is inward, in which case $\delta a/a \simeq +\delta m/M_p$

Evidently for significant migration to occur we require that the total mass in planetesimals be comparable to the planet mass,

$$\sum \delta m \sim M_p. \tag{235}$$

This is a similar result to that obtained in the case of gas disk migration, though for planetesimals the restriction is more severe since while a low mass gas disk can drive migration — albeit at a slower pace — ejected planetesimals are permanently removed from the system and cannot influence the planet further. We also note that for a single massive planet embedded within a sea of planetesimals, inward and outward scatterings will at least partially balance, leading to little net change in orbital radius.

The foregoing discussion suggests that planetesimal migration might be a negligible effect. However, Fernandez & Ip (1984) showed that the architecture of the outer Solar System favors substantial *outward* migration of the ice giants. The key point is that Jupiter is able to eject planetesimals from the Solar System more easily that the other giant planets. Jupiter itself therefore tends to move inward by a relatively small amount due to the ejection of debris at initially larger orbital radii. The other outer planets scatter bodies inward, to locations from which they are removed by Jupiter. This depletion reduces the number of outward scatterings, and as a consequence the outer planets (minus Jupiter) migrate outward.

1. Solar System evidence

Malhotra (1993) and Malhotra (1995) considered the effect of the outward migration of Neptune on the origin of Pluto and dynamically similar Kuiper Belt Objects. The idea is that as Neptune migrated outward, Pluto and smaller KBOs were captured into mean motion resonances. The eccentricities of captured bodies then increase as Neptune continues to move out. For a particle locked into a $j : j+1$ resonance, the eccentricity is (Malhotra, 1995)

$$e^2 = e_0^2 + \frac{1}{j+1} \ln \left(\frac{a_{\rm Neptune}}{a_{\rm Neptune,init}} \right) \tag{236}$$

where e_0 is the eccentricity on capture into the resonance, $a_{\rm Neptune,init}$ is the semi-major axis of Neptune when the particle was captured, and $a_{\rm Neptune}$ is the final semimajor axis. For example, if Pluto, then at 33 AU, was captured into 3:2 resonance with Neptune when the latter was at 25 AU, then migration within the resonance out to Neptune's current location at 30.2 AU matches Pluto's current eccentricity of $e \approx 0.25$.

This explanation for the origin of Pluto's peculiar orbit is attractive, but even more persuasive evidence for Neptune's migration comes from the existence of a large population of KBOs in 3:2 resonance (and smaller numbers in other major resonances) with Neptune. This population stands out in even the raw plot of a vs e for KBOs shown in Figure 33. In more detail, Murray-Clay & Chiang (2005) and Hahn & Malhotra (2005) have shown that the distribution of KBOs in resonance with Neptune (not just the 3:2 resonance) is broadly consistent with, and constrains the time scale of, outward migration of Neptune. Overall, the Solar System evidence seems entirely consistent with the hypothesis that substantial migration of Neptune captured a substantial disk of planetesimals and swept them into resonant configurations akin to that of Pluto.

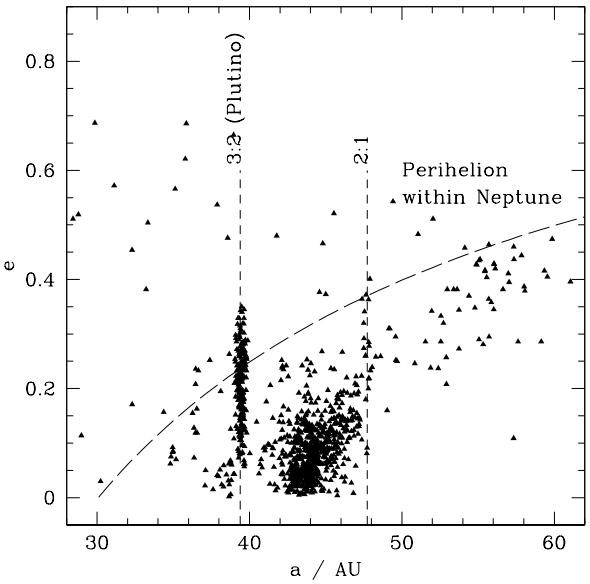

FIG. 33 The semimajor axes and eccentricities of known (as of 2008) transneptunian bodies. The vertical lines shows the location of the 3:2 and 2:1 resonances with Neptune, the dashed line shows the minimum eccentricity needed for a body to cross Neptune's orbit.

2. The Nice model

As noted above, the evidence in favor of *Neptune* having migrated outward is strong and relatively direct. Tsiganis et al. (2005) have proposed a significant extension of these ideas — dubbed the *Nice model*[20] — in which *all* of the outer planets started off in a much more compact configuration (see also Thommes, Duncan & Levison, 1999). A specific example of the initial conditions envisaged in the Nice model (say, at the epoch when the gas disk is dispersed) is outlined in Table III. The basic setup is that Jupiter is initially modestly further from the

[20] After the French city.

TABLE IV Initial conditions of the Nice model

	$a_{\rm initial}$/AU	$a_{\rm now}$/AU
Jupiter	5.45	5.2
Saturn	8.5	9.6
Uranus	11-17	19.3
Neptune	11-17	30.2
Planetesimals	30-50$M_\oplus$ at $a < 30$ AU	

Sun that it is now, while the rest of the giant planets are squeezed within about 15 AU of the Sun. Surrounding the planets is a massive disk of planetesimals that provide the impetus for outward migration of all the planets except Jupiter.

An N-body calculation started with the initial conditions of the Nice model results in outward migration of Saturn, Uranus and Neptune, as in earlier models. The critical new element of the Nice model is the hypothesis that Saturn is initially *interior* to the 2:1 resonance with Jupiter. When Saturn crosses this resonance the eccentricity of the planets is increased dramatically, and the ice giants penetrate the outer planetesimal disk. This leads to a brief phase of very rapid scattering, which Gomes et al. (2005) associate with the Late Heavy Bombardment on the Moon (Hartmann et al., 2000; Strom et al., 2005). If this association is correct, then lunar chronology pins the epoch of resonance crossing at about 700 Myr after the formation of the Solar System — i.e. surprisingly late in the history of the Solar System. Morbidelli et al. (2005) show that Jupiter's Trojan asteroids could have been captured into their now stable orbits at around the same time.

To date, the Nice model is the closest thing planetary science has to a 'standard model' that aspires to provide a full account of the early history of the outer Solar System. Although – given the necessarily sparse data available 4.5 Gyr later – it is hard to assess the uniqueness of such a model, it appears to have a great deal of promise as an economical explanation for a diverse array of Solar System phenomena.

C. Planet-planet scattering

While the gas disk is present, gas damping can potentially protect a multiple planet system against the development of crossing orbits from planet-planet gravitational interactions (at least if interactions with the gas disk actually damp eccentricity, which as noted above is somewhat uncertain). Once the gas is gone, gravity can go to work on what may be an unstable planetary system and change the orbital radii and eccentricities of the planets. This process — gravitational scattering — is probably the most widely invoked mechanism to explain the large eccentricities of many extrasolar giant planets.

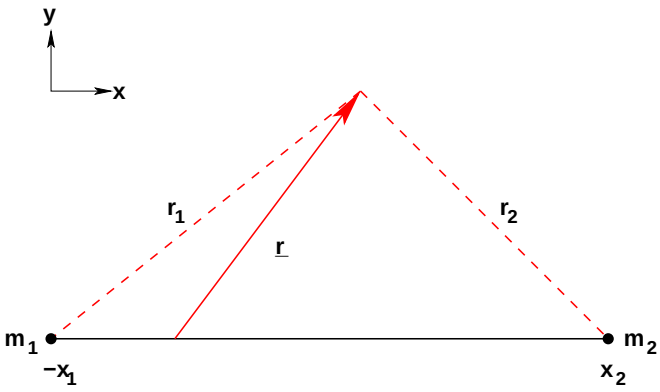

FIG. 34 Co-ordinate system for the restricted three body problem. We work in a co-rotating Cartesian co-ordinate system centered on the center of mass in which the star and planet are located at $(-x_1, 0)$ and $(x_2, 0)$ respectively. The test particle is at position $\mathbf{r}$.

1. Hill stability

Let us begin with some analytic considerations. The general N-body problem of the motion of N point masses interacting under Newtonian gravity is analytically insoluble for $N > 2$. Here, we start by considering a special case of $N = 3$ in which two bodies, of arbitrary mass, have a circular orbit, while a third body of negligible mass orbits in the known gravitational field of the massive objects. This problem — called the *circular restricted 3-body problem* — still defies analytic solution, but it is possible to place useful limits on the motion of the third body (often described as a 'test particle'). The circular restricted 3-body problem is a reasonable approximation to several situations of great practical interest, including the motion of asteroids in the vicinity of Jupiter, and the evolution of planetesimals near a growing planet. A good description of the problem can be found in Murray & Dermott (1999), whose treatment we largely mirror here. The more general 3-body problem is discussed (in both the planetary and multiple star contexts) in a book by Valtonen & Karttunen (2006).

As shown in Figure 34, we consider a binary system in which the massive bodies have mass m_1 and m_2 respectively. We work in a corotating co-ordinate system centered on the center of mass. The orbital plane is (x, y) in Cartesian co-ordinates, and the test particle is located at position $\mathbf{r}$.

If the angular velocity of the binary is Ω, the equations of motion for the test particle are,

$$\ddot{\mathbf{r}} = -\nabla\Phi - 2\left(\mathbf{\Omega}\times\dot{\mathbf{r}}\right) - \mathbf{\Omega}\times\left(\mathbf{\Omega}\times\mathbf{r}\right) \quad (237)$$

$$\Phi = -\frac{Gm_1}{r_1} - \frac{Gm_2}{r_2}. \quad (238)$$

Expressed in components, we have,

$$\begin{aligned}
\ddot{x} - 2\Omega\dot{y} - \Omega^2 x &= -G\left[\frac{m_1(x+x_1)}{r_1^3} + \frac{m_2(x-x_2)}{r_2^3}\right] \\
\ddot{y} + 2\Omega\dot{x} - \Omega^2 y &= -G\left[\frac{m_1}{r_1^3} + \frac{m_2}{r_2^3}\right] y \\
\ddot{z} &= -G\left[\frac{m_1}{r_1^3} + \frac{m_2}{r_2^3}\right] z. \qquad (239)
\end{aligned}$$

The acceleration due to the centrifugal force can be subsumed into a pseudo-potential. Defining,

$$U \equiv \frac{\Omega^2}{2}\left(x^2 + y^2\right) + \frac{Gm_1}{r_1} + \frac{Gm_2}{r_2} \qquad (240)$$

we obtain,

$$\begin{aligned}
\ddot{x} - 2\Omega\dot{y} &= \frac{\partial U}{\partial x} \\
\ddot{y} + 2\Omega\dot{x} &= \frac{\partial U}{\partial y} \\
\ddot{z} &= \frac{\partial U}{\partial z}. \qquad (241)
\end{aligned}$$

Digressing briefly, we note that U is (up to an arbitrary minus sign) the 'Roche potential'. Two stars, or a star plus a planet, that rotate synchronously while on circular orbits occupy Roche equipotentials. If their size is comparable to the size of the *Roche lobe* — defined by the critical figure-of-eight shaped equipotential that passes through the inner Lagrange point L_1 — then the bodies suffer significant tidal distortion. A useful approximation for the radius $R_{\rm RL}$ of a sphere with the same volume as the Roche lobe was provided by Eggleton (1983). For a binary with mass ratio $q \equiv m_2/m_1$ and separation a,

$$\frac{R_{\rm RL}}{a} \simeq \frac{0.49q^{2/3}}{0.6q^{2/3} + \ln(1+q^{1/3})}. \qquad (242)$$

This equation can be used to assess, for example, how close hot Jupiters are to overflowing their Roche lobes. For a Jupiter mass planet with $q = 10^{-3}$,

$$R_{\rm RL} \simeq 0.048a. \qquad (243)$$

A planet with the same radius as Jupiter (7.14×10^9 cm) would then overflow its Roche lobe interior to $a = 0.01$ AU. A very short period hot Jupiter, such as OGLE-TR-56b (Torres et al., 2004) with a period of 1.2 days, has a semi-major axis that is about 0.0225 AU. So this planet, and more securely other hot Jupiters that orbit modestly further out, is safe against mass transfer, though not by a large margin.

Returning to the general equations (241), we eliminate the Coriolis terms by multiplying through by $\dot{x}$, $\dot{y}$ and $\dot{z}$ and adding. We then obtain,

$$\begin{aligned}
\dot{x}\ddot{x} + \dot{y}\ddot{y} + \dot{z}\ddot{z} &= \dot{x}\frac{\partial U}{\partial x} + \dot{y}\frac{\partial U}{\partial y} + \dot{z}\frac{\partial U}{\partial z} \\
\frac{\rm d}{{\rm d}t}\left(\frac{1}{2}\dot{x}^2 + \frac{1}{2}\dot{y}^2 + \frac{1}{2}\dot{z}^2\right) &= \frac{{\rm d}U}{{\rm d}t} \\
\dot{x}^2 + \dot{y}^2 + \dot{z}^2 &= 2U - C_J \\
C_J &= 2U - v^2 \qquad (244)
\end{aligned}$$

where v is the velocity and C_J, called the *Jacobi constant*, is the arbitrary constant of integration. C_J is an energy-like quantity that is a conserved quantity in the circular restricted 3-body problem.

The existence of this integral of motion is important because it places limits on the range of motion possible for the test particle. For a particle with a given initial position and velocity, we can use equation (244) to compute C_J, and hence to specify *zero-velocity surfaces*, defined via,

$$2U = C_J, \qquad (245)$$

which the particle can never cross. If the volume enclosed by one of the zero-velocity surface is finite, then a particle initially within that region is guaranteed to remain there for all time. This concept is known as *Hill stability*.

The topology of the zero-velocity surfaces in the restricted three-body problem varies according to the value

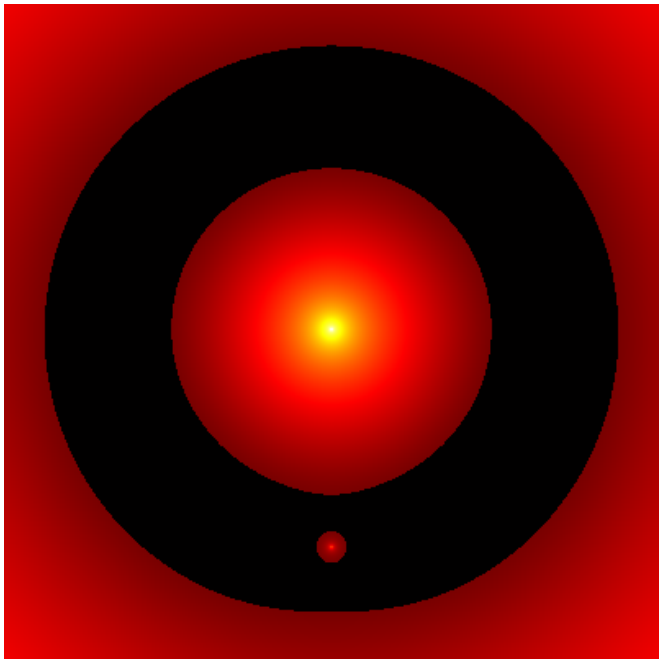

FIG. 35 Forbidden zones (dark regions) in an example of the restricted 3-body problem. For this particular choice of the Jacobi constant C_J, particles can orbit the star at small radii; the planet in a tight orbit; or the star-planet binary as a whole. The existence of zero-velocity surfaces, however, means that particles cannot be exchanged between these regions.

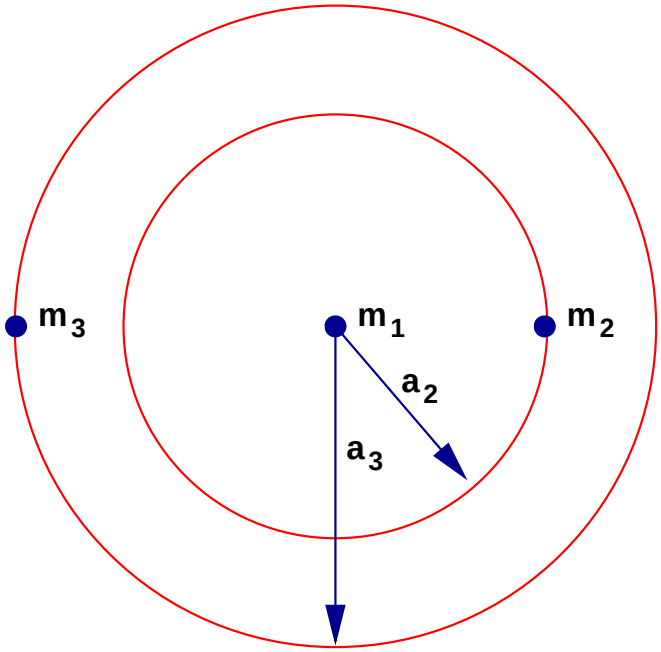

FIG. 36 Setup for the stability calculation of a two planet system in which both of the planets are on circular orbits. Unlike in the case of the Hill problem, here we strictly require that $m_2 \ll m_1$ and $m_3 \ll m_1$.

of C_J. An example is shown in Figure 35. In this instance the zero-velocity surfaces define three disjoint regions in the (x, y) plane, one corresponding to orbits around the star, one corresponding to orbits around the planet, and one corresponding to orbits around the star-planet binary. A particle in any one of these states is stuck there — it cannot cross the forbidden zone between the different regions to move into a different state.

2. Scattering and exoplanet eccentricities

The test particle analysis discussed above can, somewhat surprisingly, be extended to the much tougher problem of the stability of two planets orbiting a star. Consider the situation shown in Figure 36, in which planets of mass m_2 and m_3 orbit a star of mass m_1 in circular orbits with semi-major axes a_2 and a_3 respectively. The stability of the system evidently must depend upon the relative, rather than the absolute, spacing between the orbits. Accordingly we write,

$$a_3 = a_2(1 + \Delta) \tag{246}$$

with Δ being a dimensionless measure of the orbital separation between the planets. We further define $\mu_2 = m_2/m_1$ and $\mu_3 = m_3/m_1$. Then for μ_2,$\mu_3 \ll 1$, Gladman (1993), drawing on earlier results derived by Marchal & Bozis (1982) and others, showed that the system is guaranteed to be stable provided that the separation Δ exceeds a critical separation Δ_c given by,

$$\Delta_c \simeq 2.40\,(\mu_2 + \mu_3)^{1/3}\,. \tag{247}$$

Note that analytic results leave open the question of whether systems with $\Delta < \Delta_c$ are actually unstable, all we know is that $\Delta > \Delta_c$ is sufficient for stability. This condition reduces to the test particle result if $\mu_3 \to 0$, as of course it should[21]. As an example, if we compute the critical separation for planets of the mass of Jupiter and Saturn, we obtain $\Delta_c \simeq 0.26$. The actual separation of Jupiter and Saturn in these units is $\Delta \simeq 0.83$, so an isolated planetary system in which Jupiter and Saturn were on circular orbits would assuredly be stable for all time.

What about more complex systems? It is possible to include non-zero eccentricities into this analysis, but *not* more planets. For a multiple planet system one might plausibly reason that the system will be unstable if any pair substantially violates the critical two-planet separation for Hill stability. It is also true that the system will generally become more stable as the separations increase. However, no absolute stability bound is known for any planetary system with $N > 3$.

If a two-planet system is unstable, the possible outcomes of the instability can be divided into four classes:

1. The separation evolves (increases) until the system achieves a state that is stable over the long term.
2. One planet is ejected, while the other remains bound, generally with $e \neq 0$.
3. The planets physically collide.
4. One planet impacts the star.

The last two channels are not possible in a model 3-body problem, in which the planets are represented by point masses, but can occur (especially planet-planet collisions) in real systems.

The idea that gravitational scattering and planetary ejections might account for the eccentricity of extrasolar planets was proposed as soon as it became clear that extrasolar planets were not typically on circular orbits (Lin & Ida, 1997; Rasio & Ford, 1996; Weidenschilling & Marzari, 1996). Quantitative study of such models requires large-scale N-body integrations, first to derive the statistical distribution of outcomes of any given scenario (since the systems are typically chaotic, nothing can be said about any single run), and second to map out the large parameter space that results when one considers different numbers of planets with different initial separations, masses and so forth.

[21] Note, however, that the analysis for the restricted three-body problem applies for an arbitrary mass ratio of the massive bodies, whereas the result for two planets requires that *both* be much less massive than the star.

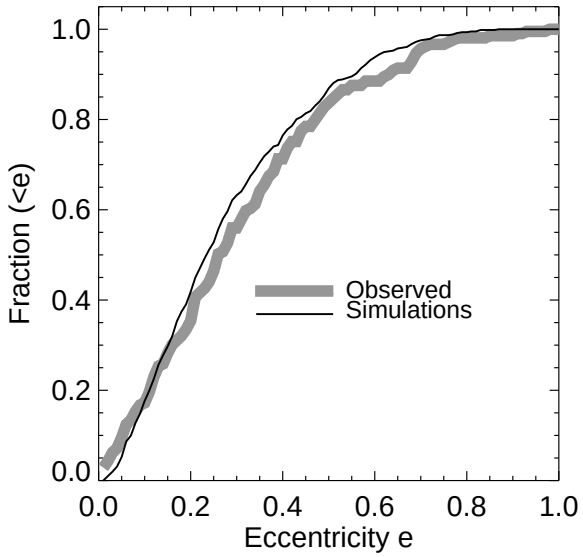

FIG. 37 The cumulative eccentricity distribution of known extrasolar planets is compared to the predicted distribution that results from scattering in three planet systems. The simulation results were derived by numerically evolving an ensemble of unstable planetary systems made up of three planets whose masses were drawn from the observed mass function for extrasolar planets in the range $M_{\rm Sat} < M_p < 3\ M_J$. Based on simulations by Raymond et al. (2008).

Ford, Havlickova & Rasio (2001) presented a comprehensive study of the dynamics of equal mass two planet systems. The planets were set up on circular orbits close to the stability boundary, and allowed to evolve under purely N-body forces until the system relaxed to a stable state. They found that the predicted fraction of collisions increases sharply for small orbital radii and / or larger planetary radii. For pairs of Jupiter mass and Jupiter radius planets initially located at 5 AU, the most common outcome is two planets (65%), followed by ejections (35%), with collisions (10%) a distant third. If the same pair of planets starts at 1 AU, however, collisions occur roughly 30% of the time. This conclusion is important for studies of extrasolar planet eccentricity, because collisions yield relatively low eccentricities for the merged planet. Indeed, Ford, Havlickova & Rasio (2001) found that equal mass planet scattering failed to match the observed distribution of eccentricities. However, subsequent calculations that relaxed the equal mass assumption showed that two planet systems in which the planets have a realistic range of masses *can* yield agreement with observations (Ford, Rasio & Yu, 2003).

There is only a rather small range of orbital separations which allows a two planet system to be unstable over the long term (greater than around 10^5 yr, which is roughly the dispersal time for the gas disk), while not being violently unstable. This observation means that it is easier to set up an internally self-consistent scattering model with three or more planets, since a wider range of such systems eventually lead to interesting dynamics. Models starting with three or more planets have also been studied in some detail (Adams & Laughlin, 2003; Marzari & Weidenschilling, 2002; Terquem & Papaloizou, 2002). Recent comprehensive studies, such as those by Chatterjee et al. (2008) and Jurić & Tremaine (2008), find that scattering models yield a good quantitative match to the observed distribution of extrasolar planet eccentricities. Just how good this agreement is is illustrated in Figure 37, which shows how the final simulated eccentricities compare to the data (Raymond et al., 2008). Although this match does not prove that planet-planet scattering is the sole (or even the main) source of exoplanet eccentricities, it is sufficient to establish this model as the leading candidate for explaining the broad distribution of exoplanet eccentricity.

D. Predictions of migration theories

In summary, there is persuasive circumstantial evidence for the action of at least three separate processes that lead to the early evolution of planetary systems:

- **Gas disk migration** in the Type II regime appears to be necessary to explain the existence of hot Jupiters. An argument can be made that more modest levels of migration ought to be common, since at least some gas must necessarily be present at the epoch when giant planets form.

- **Planetesimal disk migration** provides a persuasive explanation for the origin of Pluto's odd orbit together with some of the more detailed properties of the Kuiper Belt. One can make a case that this process too ought to be common in the outer reached of planetary systems. Gas giant formation almost certainly becomes more difficult further out in the disk, so it is quite plausible that the zone where gas and ice giants manage to form is often surrounded by a disk of planetesimal debris that has been unable to grow to large sizes.

- **Planet-planet scattering** works well as an explanation for the eccentricity distribution of giant extrasolar planets. There is *no* straightforward independent argument that the unstable initial conditions needed for such models to work are generically realized in nature, but the empirical evidence seems to suggest that they are.

A number of qualitatively different tests of these theoretical ideas are possible in the near future. Planet-planet scattering models, for example, predict that the surviving planets ought to have a distribution of inclination as well as eccentricity, which can be measured (relative to the stellar equatorial plane) for those objects observed in transit (Rossiter, 1924; Winn et al., 2009). With a large sample of measured inclinations it ought

to be possible to determine whether the predictions of pure planet-planet scattering models are confirmed, or whether the final inclinations are instead also affected by other processes (such as damping from a residual gas disk, or Kozai pumping of inclination by more distant companions). The *combined* action of multiple evolutionary mechanisms may also give rise to new classes of planetary systems. When multiple planets interact with a gas disk, for example, convergent orbital migration can result in the planets becoming trapped into resonant orbital configurations that resemble those seen in some known extrasolar planetary systems (Lee & Peale, 2002). At larger orbital radii (than those currently probed by observations of exoplanets) it seems likely that we ought to see planetary systems whose dynamics has been affected by both planet-planet scattering and planetesimal disk migration. N-body simulations suggest that the signature of this combination is a transition from generally eccentric to nearly circular planetary orbits as the mass of the planetary system is reduced (Raymond, Armitage & Gorelick, 2009). If true, the near-circular orbits of the giant planets in the Solar System might in fact be typical of the architecture of relatively low-mass systems at large orbital radii. For higher mass systems the same simulations predict a high abundance of resonant configurations, including resonant chains that would be planetary analogs of the Laplace resonance in the Jovian satellite system.

Acknowledgements

These notes are based on a graduate course given at the University of Colorado, Boulder, in Fall 2004, 2006 and 2008, and on lectures given at the XIII Ciclo de Cursos Especiais given at the Observatório Nacional in Rio de Janeiro in Fall 2008. My thanks to the students in those classes for stimulating my interest in this subject, and to Dimitri Veras for reading a draft of the initial (2006) edition of these notes. In preparing this revision I've been greatly assisted by discussions with Eric Feigelson and Dave Stevenson. My work on planet formation and accretion physics has been generously supported by NASA under the auspices of the Origins of Solar Systems, Astrophysics Theory and Beyond Einstein Foundation Science programs, and by the National Science Foundation.

References

Adams, F. C., Lada, C. J., & Shu, F. H. 1987, *ApJ*, 312, 788

Adams, F. C., & Laughlin, G. 2003, *Icarus*, 163, 290

Adams, F. C., Proszkow, E. M., Fatuzzo, M., & Myers, P. C. 2006, *ApJ*, 641, 504

Adams, F. C., & Shu, F. H. 1986, *ApJ*, 308, 836

Afshordi, N., Mukhopadhyay, B., & Narayan, R. 2005, *ApJ*, 629, 373

Agol, E., Steffen, J., Sari, R., & Clarkson, W. 2005, *MNRAS*, 359, 567

Alexander, R. D., & Armitage, P. J. 2007, *MNRAS*, 375, 500

Alexander, R. D., Clarke, C. J., & Pringle, J. E. 2005, *MNRAS*, 358, 283

Alexander, R. D., Clarke, C. J., & Pringle, J. E. 2006, *MNRAS*, 369, 229

Alibert, Y., Mordasini, C., Benz, W., & Winisdoerffer, C. 2005, *A&A*, 434, 343

Alibert, Y., Mousis, O., Mordasini, C., & Benz, W. 2005, *ApJ*, 626, L57

Arakawa, M., Leliwa-Kopystynski, J., & Maeno, N. 2002, *Icarus*, 158, 516

Armitage, P. J. 2002, *MNRAS*, 330, 895

Armitage, P. J. 2007, *ApJ*, 665, 1381

Armitage, P. J., & Clarke, C. J. 1996, *MNRAS*, 280, 458

Armitage, P. J., Livio, M., Lubow, S. H., & Pringle, J. E. 2002, *MNRAS*, 334, 248

Armitage, P. J., Livio, M., & Pringle, J. E. 2001, *MNRAS*, 324, 705

Armitage, P. J., & Rice, W. K. M. 2005, to appear in *A Decade Of Extrasolar Planets Around Normal Stars*, STScI May Symposium 2005, astro-ph/0507492

Artymowicz, P. 1993, *ApJ*, 419, 155

Artymowicz, P., Clarke, C. J., Lubow, S. H., & Pringle, J. E. 1991, *ApJ*, 370, L35

Artymowicz, P., & Lubow, S. H. 1994, *ApJ*, 421, 651

Baglin, A., Auvergne, M., Barge, P., Buey, J.-T., Catala, C., Michel, E., Weiss, W., & COROT Team 2002, in *Proceedings of the First Eddington Workshop on Stellar Structure and Habitable Planet Finding*, ed. B. Battrick, ESA SP-485, Noordwijk: ESA Publications Division, p. 17

Bally, J., & Scoville, N. Z. 1982, *ApJ*, 255, 497

Balbus, S. A. 2009, *Physical Processes in Circumstellar Disks Around Young Stars*, ed. P. Garcia, University of Chicago Press: Chicago (http://arxiv.org/abs/0906.0854)

Balbus, S. A., & Hawley, J. F. 1991, *ApJ*, 376, 214

Balbus, S. A., & Hawley, J. F. 1998, *Reviews of Modern Physics*, 70, 1

Balbus, S. A., & Hawley, J. F. 2006, *ApJ*, 652, 1020

Balbus, S. A., Hawley, J. F., & Stone, J. M. 1996, *ApJ*, 467, 76

Barge, P., & Sommeria, J. 1995, *A&A*, 295, L1

Barranco, J. A., & Marcus, P. S. 2005, *ApJ*, 623, 1157

Beaulieu, J.-P., et al. 2006, *Nature*, 439, 437

Begelman, M. C., McKee, C. F., & Shields, G. A. 1983, *ApJ*, 271, 70

Bell, K. R., Cassen, P. M., Klahr, H. H., & Henning, Th. 1997, *ApJ*, 486, 372

Bell, K. R., & Lin, D. N. C. 1994, *ApJ*, 427, 987

Benz, W., & Asphaug, E. 1999, *Icarus*, 142, 5

Binney, J., & Tremaine, S. 1987, *Galactic Dynamics*, Princeton University Press, p. 310

Blaes, O. M., & Balbus, S. A. 1994, *ApJ*, 421, 163

Blandford, R. D., & Payne, D. G. 1982, *MNRAS*, 199, 883

Bodenheimer, P., & Pollack, J. B. 1986, *Icarus*, 67, 391

Boley, A. C., Mejia, A. C., Durisen, R. H., Cai, K., Pickett, M. K., & D'Alessio, P. 2006, *ApJ*, 651, 517

Borucki, W. J., et al. 2003, in *Future EUV/UV and Visible Space Astrophysics Missions and Instrumentation*, eds J. C. Blades & O. H. W. Siegmund, Proceedings of the SPIE,

Volume 4854, p. 129
Boss, A. P. 1997, *Science*, 276, 1836
Boss, A. P. 2005, *ApJ*, 629, 535
Bouvier, J., Alencar, S. H. P., Harries, T. J., Johns-Krull, C. M., & Romanova, M. M. 2007, *Protostars and Planets V*, eds B. Reipurth, D. Jewitt, and K. Keil, University of Arizona Press, Tucson, astro-ph/0603498
Brandenburg, A.,, Nordlund, A., Stein, R. F., & Torkelsson, U. 1995, *ApJ*, 446, 741
Bromley, B. C., & Kenyon, S. J. 2006, *AJ*, 131, 2737
Brown, M. E., Trujillo, C., & Rabinowitz, D. 2004, *ApJ*, 617, 645
Burrows, C. J., et al. 1996, *ApJ*, 473, 437
Burrows, A., Hubeny, I, Budaj, J., & Hubbard, W. B. 2007, *ApJ*, 661, 502
Butler, R. P., Marcy, G. W., Williams, E., McCarthy, C., Dosanjh, P., & Vogt, S. S. 1996, *PASP*, 108, 500
Butler, R. P., et al. 2006, *ApJ*, 646, 505
Calvet, N., et al. 2005, *ApJ*, 630, L185
Cameron, A. G. W. 1978, *Moon and the Planets*, 18, 5
Cannizzo, J. K. 1993, *ApJ*, 419, 318
Canup, R. M. 2004, *ARA&A*, 42, 441
Canup, R. M., & Ward, W. R. 2002, *AJ*, 124, 3404
Carballido, A., Stone, J. M., & Pringle, J. E. 2005, *MNRAS*, 358, 1055
Carlberg, R. G., & Sellwood, J. A. 1985. *ApJ*, 292, 79
Chambers, J. E., & Wetherill, G. W. 1998, *Icarus*, 136, 304
Chandrasekhar, S. 1961, *Hydrodynamic and hydromagnetic stability*, International Series of Monographs on Physics, Oxford: Clarendon
Chapman, S., & Cowling, T. G. 1970, *The mathematical theory of non-uniform gases*, Cambridge University Press
Charbonneau, D., Brown, T. M., Burrows, A., & Laughlin, G. 2007, *Protostars and Planets V*, eds B. Reipurth, D. Jewitt, and K. Keil, University of Arizona Press, Tucson
Chatterjee, S., Ford, E. B., Matsumura, S., & Rasio, F. A. 2008, *ApJ*, 686, 580
Chiang, E. I., & Goldreich, P. 1997, *ApJ*, 490, 368
Chiang, E., Lithwick, Y., Murray-Clay, R., Buie, M., Grundy, W., & Holman, M. 2007, *Protostars and Planets V*, eds B. Reipurth, D. Jewitt, and K. Keil, University of Arizona Press, Tucson, astro-ph/0601654
Clarke, C. J., Gendrin, A., & Sotomayor, M. 2001, *MNRAS*, 328, 485
Collier Cameron, A., & Campbell, C. G. 1993, *A&A*, 274, 309
Cumming, A., et al. 2008, *PASP*, 120, 531
Cuzzi, J. N., Dobrovolskis, A. R., & Champney, J. M. 1993, *Icarus*, 106, 102
Cuzzi, J. N., Hogan, R. C., & Shariff, K. 2008, *ApJ*, 687, 1432
D'Angelo, G., Lubow, S. H., & Bate, M. R. 2006, *ApJ*, 652, 1698
Desch, S. J. 2004, *ApJ*, 608, 509
Dullemond, C. P., & Dominik, C. 2005, *A&A*, 434, 971
Dullemond, C. P., Hollenbach, D., Kamp, I., & D'Alessio, P. 2007, *Protostars and Planets V*, eds B. Reipurth, D. Jewitt, and K. Keil, University of Arizona Press, Tucson, astro-ph/0602619
Dullemond, C. P., & Dominik, C. 2004, *A&A*, 421, 1075
Durisen, R. H., Boss, A. P., Mayer, L., Nelson, A. F., Quinn, T., & Rice W. K. M. 2007, *Protostars and Planets V*, eds B. Reipurth, D. Jewitt, and K. Keil, University of Arizona Press, Tucson, astro-ph/0603179
Durisen, R. H., Cai, K., Mejía, A. C., & Pickett, M. K. 2005, *Icarus*, 173, 417
Eggleton, P. P. 1983, *ApJ*, 268, 368
Eisner, J. A., Chiang, E. I., & Hillenbrand, L. A. 2006, *ApJ*, 637, L133
Eisner, J. A., Hillenbrand, L. A., Carpenter, J. M., & Wolf, S. 2005, *ApJ*, 635, 396
Ercolano, B., Clarke, C. J., & Drake, J. J. 2009, *ApJ*, 699, 1639
Fabian, A. C., Pringle, J. E., & Rees, M. J. 1975, *MNRAS*, 172, 15
Feigelson, E., Townsley, L., Güdel, M., & Stassun, K. 2007, *Protostars and Planets V*, eds B. Reipurth, D. Jewitt, and K. Keil, University of Arizona Press, Tucson, p. 313
Fernandez, J. A., & Ip, W.-H. 1984, *Icarus*, 58, 109
Fischer, D. A., & Valenti, J. 2005, *ApJ*, 622, 1102
Font, A. S., McCarthy, I. G., Johnstone, D., & Ballantyne, D. R. 2004, *ApJ*, 607, 890
Ford, E. B., Havlickova, M., & Rasio, F. A. 2001, *Icarus*, 150, 303
Ford, E. B., Rasio, F. A., & Yu, K. 2003, in *Scientific Frontiers in Research on Extrasolar Planets*, ASP Conference Series, Vol 294, eds D. Deming & S. Seager. ASP, San Francisco, p. 181
Forrest, W. J., et al. 2004, *ApJS*, 154, 443
Frank, J., King, A., & Raine, D. J. 2002, *Accretion Power in Astrophysics*, (3rd edition, Cambridge University Press)
Fraser, W. C., & Kavelaars, J. J. 2009, *AJ*, 137, 71
Fromang, S., & Papaloizou, J. 2006, *A&A*, 452, 751
Gammie, C. F. 1996, *ApJ*, 457, 355
Gammie, C. F. 2001, *ApJ*, 553, 174
Garaud, P., & Lin, D. N. C. 2004, *ApJ*, 608, 1050
Garaud, P., & Lin, D. N. C. 2007, *ApJ*, 654, 606
Gladman, B. 1993, *Icarus*, 106, 247
Godon, P., & Livio, M. 1999, *ApJ*, 523, 350
Goldreich, P. 1964, *MNRAS*, 130, 159
Goldreich, P., Lithwick, Y., & Sari, R. 2004, *ARA&A*, 42, 549
Goldreich, P., & Sari, R. 2003, *ApJ*, 585, 1024
Goldreich, P., & Tremaine, S. 1979, *ApJ*, 233, 857
Goldreich, P., & Tremaine, S. 1980, *ApJ*, 241, 425
Goldreich, P., & Ward, W. R. 1973, *ApJ*, 183, 1051
Gomes, R., Levison, H. F., Tsiganis, K., & Morbidelli, A. 2005, *Nature*, 435, 466
Gómez, G. C., & Ostriker, E. C. 2005, *ApJ*, 630, 1093
Gorti, U., & Hollenbach, D. 2009, *ApJ*, 690, 1539
Greenberg, R., Hartmann, W. K., Chapman, C. R., & Wacker, J. F. 1978, *Icarus*, 35, 1
Guillot, T. 2005, *Annual Review of Earth and Planetary Sciences*, 33, 493
Gullbring, E., Hartmann, L., Briceno, C., & Calvet, N. 1998, *ApJ*, 492, 323
Haghighipour, N., & Boss, A. P. 2003, *ApJ*, 598, 1301
Hahn, J. M., & Malhotra, R. 2005, *AJ*, 130, 2392
Haisch, K. E., Lada, E. A., & Lada, C. J. 2001, *ApJ*, 553, L153
Hartmann, L., Calvet, N., Gullbring, E., & D'Alessio, P. 1998, *ApJ*, 495, 385
Hartmann, L., & Kenyon, S. J. 1995, *ARA&A*, 34, 207
Hartmann, W. K., Ryder, G., Dones, L., & Grinspoon, D. 2000, in *Origin of the Earth and Moon*, edited by R.M. Canup and K. Righter, University of Arizona Press (Tus-

con), p. 493
Hawley, J. F. 2001, *ApJ*, 554, 534
Hayashi, C. 1981, *Progress of Theoretical Physics Supplement*, 70, 35
Herbst, W., Eisloeffel, J., Mundt, R., & Scholz, A. 2007, *Protostars and Planets V*, eds B. Reipurth, D. Jewitt, and K. Keil, University of Arizona Press, Tucson, astro-ph/0603673
Herbst, W., & Mundt, R. 2005, *ApJ*, 633, 967
Hollenbach, D., Johnstone, D., Lizano, S., & Shu, F. 1994, *ApJ*, 428, 654
Holman, M. J., & Murray, N. W. 2005, *Science*, 307, 1288
Hubickyj, O., Bodenheimer, P., & Lissauer, J. J. 2005, *Icarus*, 179, 415
Ida, S., Bryden, G., Lin, D. N. C., & Tanaka, H. 2000, *ApJ*, 534, 428
Ida, S., Guillot, T., & Morbidelli, A. 2008, *ApJ*, 686, 1292
Ida, S., Lin, D. N. C. 2004, *ApJ*, 616, 567
Ikoma, M., Nakazawa, K., & Emori, H. 2000, *ApJ*, 537, 1013
Ioannou, P. J., & Kakouris, A. 2001, *ApJ*, 550, 931
Inaba, S., Tanaka, H., Nakazawa, K., Wetherill, G. W., & Kokubo, E. 2001, *Icarus*, 149, 235
Inutsuka, S., & Sano, T. 2005, *ApJ*, 628, L155
Ivanov, P. B., Papaloizou, J. C. B., & Polnarev, A. G. 1999, *MNRAS*, 307, 79
Jewitt, D., & Luu, J. 1993, *Nature*, 362, 730
Ji, H., Burin, M., Schartman, E., & Goodman, J. 2006, *Nature*, 444, 343
Johansen, A., Henning, T., & Klahr, H. 2006, *ApJ*, 643, 1219
Johansen, A., & Klahr, H. 2005, *ApJ*, 634, 1353
Johansen, A., Oishi, J. S., Mac Low, M.-M., Klahr, H., Henning, T., & Youdin, A. 2007, *Nature*, 448, 1022
Johnson, B. M., & Gammie, C. F. 2005, *ApJ*, 635, 149
Johnson, E. T., Goodman, J., & Menou, K. *ApJ*, 647, 1413
Johnstone, D., Hollenbach, D., & Bally, J. 1998, *ApJ*, 499, 758
Jurić, M., & Tremaine, S. 2008, *ApJ*, 686, 603
Kasting, J. F., Whitmire, D. P., & Reynolds, R. T. 1993, *Icarus*, 101, 108
Kenyon, S. C. 2002, *PASP*, 114, 265
Kenyon, S. J., & Hartmann, L. 1987, *ApJ*, 323, 714
Kenyon, S. J., & Luu, J. X. 1998, *AJ*, 115, 2136
Kirsh, D. R., Duncan, M., Brasser, R., & Levison, H. F. 2009, *Icarus*, 199, 197
Kokubo, E., & Ida, S. 1998, *Icarus*, 131, 171
Kokubo, E., Kominami, J., & Ida, S. 2006, *ApJ*, 642, 1131
Konacki, M., Torres, G., Jha, S., & Sasselov, D. D. 2003, *Nature*, 421, 507
Königl, A. 1991, *ApJ*, 370, L39
Korycansky, D. G., & Asphaug, E. 2006, *Icarus*, 181, 605
Kuiper, G. P. 1951, *Proc. Natl. Acad. Sci.* 37, 1
Lada, C. J., & Lada, E. A. 2003, *ARA&A*, 41, 57
Latter, H. N., & Ogilvie, G. I. 2006, *MNRAS*, 372, 1829
Laughlin, G., & Bodenheimer, P. 1994, *ApJ*, 436, 335
Laughlin, G., Steinacker, A., Adams, F. C. 2004, *ApJ*, 608, 489
Lecar, M., Podolak, M., Sasselov, D., & Chiang, E. 2006, *ApJ*, 640, 1115
Lee, M. H. 2000, *Icarus*, 143, 74
Lee, M. H., & Peale, S. J. 2002, *ApJ*, 567, 596
Leinhardt, Z. M., & Richardson, D. C. 2002, *Icarus*, 159, 306
Leinhardt, Z. M., & Stewart, S. T. 2009, *I*carus, 199, 542
Lesur, G., & Papaloizou, J. C. B. 2009, *A&A*, 498, 1
Levison, H. F., & Agnor, C. 2003, *AJ*, 125, 2692
Levison, H. E., Morbidelli, A., Gomes, R., & Backman, D. 2007, *Protostars and Planets V*, eds B. Reipurth, D. Jewitt, and K. Keil, University of Arizona Press, Tucson
Lin, D. N. C., Bodenheimer, P., & Richardson, D. C. 1996, *Nature*, 380, 606
Lin, D. N. C., & Ida, S. 1997, *ApJ*, 477, 781
Lin, D. N. C., & Papaloizou, J. 1980, *MNRAS*, 191, 37
Lin, D. N. C., & Pringle, J. E. 1990, *ApJ*, 358, 515
Lissauer, J. J. 1993, *ARA&A*, 31, 129
Lithwick, Y. 2009, *ApJ*, 693, 85
Livio, M., & Pringle, J. E. 2003, *MNRAS*, 346, L42
Lodders, K. 2003, *ApJ*, 591, 1220
Lovett, E. O. 1895, *AJ*, 15, 113
Lubow, S. H., Seibert, M., & Artymowicz, P. 1999, *ApJ*, 526, 1001
Lynden-Bell, D. 1969, *Nature*, 223, 690
Lynden-Bell, D., & Pringle, J. E. 1974, *MNRAS*, 168, 603
Malhotra, R. 1993, *Nature*, 365, 819
Malhotra, R. 1995, *AJ*, 110, 420
Marois et al. 2008, *Science*, 322, 1348
Marchal, C., & Bozis, G. 1982, *Celestial Mechanics*, 26, 311
Marcy, G., Butler, R. P., Fischer, D., Vogt, S., Wright, J. T., Tinney, C. G., & Jones, H. R. A. 2005, *Progress of Theoretical Physics Supplement*, 158, 24
Marzari, F., & Weidenschilling, S. J. 2002, *Icarus*, 156, 570
Masset, F. S., D'Angelo, G., & Kley, W. 2006, *ApJ*, 652, 730
Masset, F. S., & Ogilvie, G. I. 2004, *ApJ*, 615, 1000
Mathis, J. S., Rumpl, W., & Nordsieck, K. H. 1977, *ApJ*, 217, 425
Matsumura, S., & Pudritz, R. E. 2006, *MNRAS*, 365, 572
Matt, S., & Pudritz, R. E. 2005, *ApJ*, 632, L135
Mayer, L., Lufkin, G., Quinn, T., & Wadsley, J. 2006, *ApJ*, 661, L77
Mayor, M., & Queloz, D. 1995, *Nature*, 378, 355
Mayor, M., & Udry, S. 2008, *Physica Scripta*, 130, 014010
Militzer, B., Hubbard, W. B., Vorberger, J., Tamblyn, I., & Bonev, S. A. 2008, *ApJ*, 688, L45
Miller, K. A., & Stone, J. M. 2000, *ApJ*, 534, 398
Mizuno, H. 1980, *Progress of Theoretical Physics*, 64, 544
Moorhead, A. V., & Adams, F. C. 2008, *ApJ*, 193, 475
Morbidelli, A., Bottke, W., Nesvorny, D., & Levison, H. F. 2009, *Icarus*, in press (http://arxiv.org/abs/0907.2512)
Morbidelli, A., Chambers, J., Lunine, J. I., Petit, J. M., Robert, F., Valsecchi, G. B., & Cyr, K. E. 2000, *Meteoritics & Planetary Science*, 35, 1309
Morbidelli, A., Levison, H. F., Tsiganis, K., & Gomes, R. 2005, *Nature*, 435, 462
Murray, C. D., & Dermott, S. F. 1999, *Solar System Dynamics*, Cambridge University Press
Murray-Clay, R. A., & Chiang, E. I. 2005, *ApJ*, 619, 623
Nelson, R. P. 2005, *A&A*, 443, 1067
Nelson, R. P., & Papaloizou, J. C. B. 2004, *MNRAS*, 350, 849

Nesvorný, D., Alvarellos, J. L. A., Dones, L., & Levison, H. F. 2003, *AJ*, 126, 398

Nettelmann, N., Holst, B., Kietzmann, A., French, M., Redmer, R., & Blaschke, D. 2008, *ApJ*, 683, 1217

O'Dell, C. R., Wen, Z., & Hu, X. 1993, *ApJ*, 410, 696

Ogilvie, G. I., & Lubow, S. H. 2003, *ApJ*, 587, 398

Oishi, J. S., Mac Low, M.-M., & Menou, K. 2007, *ApJ*, 670, 805

Papaloizou, J. C. B., & Lin, D. N. C. 1984, *ApJ*, 285, 818

Papaloizou, J. C. B., Nelson, R. P., & Masset, F. 2001, *A&A*, 366, 263

Papaloizou, J. C. B., & Terquem, C. 1999, *ApJ*, 521, 823

Perri, F., & Cameron, A. G. W. 1973, *Icarus*, 22, 416

Pessah, M. E., Chan, C., & Psaltis, D. 2008, *MNRAS*, 383, 683

Podolak, M. 2003, *Icarus*, 165, 428

Pollack, J. B., Hubickyj, O., Bodenheimer, P., Lissauer, J. J., Podolak, M., & Greenzweig, Y. 1996, *Icarus*, 124, 62

Popham, R., Narayan, R., Hartmann, L., & Kenyon, S. 1993, *ApJ*, 415, L127

Pringle, J. E. 1977, *MNRAS*, 178, 195

Pringle, J. E. 1981, *ARA&A*, 19, 137

Pringle, J. E. 1989, *MNRAS*, 236, 107

Pringle, J. E. 1991, *MNRAS*, 248, 754

Pringle, J. E., Verbunt, F., & Wade, R. A. 1986, *MNRAS*, 221, 169

Quillen, A. C., Blackman, E. G., Frank, A., & Varniére, P. 2004, *ApJ*, 612, L137

Rafikov, R. R. 2005, *ApJ*, 621, L69

Rasio, F. A., & Ford, E. B. 1996, *Science*, 274, 954

Raymond, S. N., Armitage, P. J., & Gorelick, N. 2009, *ApJ*, 699, L88

Raymond, S. N., Barnes, R., Armitage, P. J., & Gorelick, N. 2008, *ApJ*, 687, L107

Raymond, S. N., O'Brien, D. P., Morbidelli, A., & Kaib, N. A. 2009, *Icarus*, in press (http://arxiv.org/abs/0905.3750)

Raymond, S. N., Quinn, T., & Lunine, J. I. 2005, *ApJ*, 632, 670

Rebull, L. M., Stauffer, J. R., Megeath, S. T., Hora, J. L., & Hartmann, L. 2006, *ApJ*, 646, 297

Ribas, I., & Miralda-Escude, J. 2007, *A&A*, 464, 779

Rice, W. K. M., & Armitage, P. J. 2003, *ApJ*, 598, L55

Rice, W. K. M., Armitage, P. J., Bate, M. R., & Bonnell, I. A. 2003, *MNRAS*, 339, 1025

Rice, W. K. M., Armitage, P. J., Bonnell, I. A., Bate, M. R., Jeffers, S. V., & Vine, S. G. 2003, *MNRAS*, 346, L36

Rice, W. K. M., Lodato, G., & Armitage, P. J. 2005, *MNRAS*, 364, L56

Rice, W. K. M., Lodato, G., Pringle, J. E., Armitage, P. J., & Bonnell, I. A. 2004, *MNRAS*, 355, 543

Rice, W. K. M., Wood, K., Armitage, P. J., Whitney, B. A., & Bjorkman, J. E. 2003, *MNRAS*, 342, 79

Rossiter, R. A. 1924, *ApJ*, 60, 15

Ryan, E. V., & Melosh, H. J. 1998, *Icarus*, 133, 1

Rybicki, G. B., & Lightman, A. P. 1979, *Radiative Processes in Astrophysics*, (Wiley)

Ryu, D., & Goodman, J. 1992, *ApJ*, 388, 438

Safronov, V. S. 1969, *Evolution of the Protoplanetary Cloud and Formation of the Earth and the Planets*, English translation NASA TT F-677 (1972)

Salmeron, R., & Wardle, M. 2005, *MNRAS*, 361, 45

Sano, T., & Stone, J. M. 2002, *ApJ*, 577, 534

Sato, B., et al. 2005, *ApJ*, 633, 465

Sekiya, M. 1998, *Icarus*, 133, 298

Shakura, N. I., & Sunyaev, R. A. 1973, *A&A*, 24, 337

Shen, Y., Stone, J. M., & Gardiner, T. A. 2006, *ApJ*, 653, 513

Shlosman, I., & Begelman, M. C. 1989, *ApJ*, 341, 685

Shu, F. H., Johnstone, D., & Hollenbach, D. 1993, *Icarus*, 106, 92

Sicilia-Aguilar, A., Hartmann, L. W., Fürész, G., Henning, T., Dullemond, C., & Brandner, W. 2006, *AJ*, 132, 2135

Simon, M., & Prato, L. 1995, *ApJ*, 450, 824

Smoluchowski, M. V. 1916, *Physik. Zeit.*, 17, 557

Springel, V., et al. 2005, *Nature*, 435, 629

Stevenson, D. J. 1982, *Planetary and Space Science*, 30, 755

Stone, J. M., & Balbus, S. A. 1996, *ApJ*, 464, 364

Stone, J. M., Hawley, J. F., Gammie, C. F., & Balbus, S. A. 1996, *ApJ*, 463, 656

Strom, R. G., Malhotra, R., Ito, T., Yoshida, F., & Kring, D. A. 2005, *Science*, 309, 1847

Supulver, K. D., Bridges, F. G., Tiscareno, S., Lievore, J., & Lin, D. N. C. *Icarus*, 129, 539

Syer, D., & Clarke, C. J. 1995, *MNRAS*, 277, 758

Tabachnik, S., & Tremaine, S. 2002, *MNRAS*, 335, 151

Takeuchi, T., Clarke, C. J., & Lin, D. N. C. 2005, *ApJ*, 627, 286

Takeuchi, T., & Lin, D. N. C. 2002, *ApJ*, 581, 1344

Takeuchi, T., Miyama, S. M., & Lin, D. N. C. 1996, *ApJ*, 460, 832

Terquem, C., & Papaloizou, J. C. B. 2002, *MNRAS*, 332, L39

Tanaka, H.,. Takeuchi, T., & Ward, W. R. 2002, *ApJ*, 565, 1257

Thommes, E. W., Duncan, M. J., & Levison, H. F. 1999, *Nature*, 402, 635

Thommes, E. W., Duncan, M. J., & Levison, H. F. 2003, *Icarus*, 161, 431

Thommes, E. W., & Murray, N. 2006, *ApJ*, 644, 1214

Throop, H. B., & Bally, J. 2005, *ApJ*, 623, L149

Toomre, A. 1964, *ApJ*, 139, 1217

Torres, G., Konacki, M., Sasselov, D. D., & Jha, S. 2004, *ApJ*, 609, 1071

Torres, G., Winn, J. N., & Holman, M. J. 2008, *ApJ*, 677, 1324

Trujillo, C. A., Jewitt, D. C., & Luu, J. X. 2001, *AJ*, 122, 457

Tsiganis, K., Gomes, R., Morbidelli, A., & Levison, H. F. 2005, *Nature*, 435, 459

Turner, N. J., Sano, T., & Dziourkevitch, N. 2007, *ApJ*, 659, 729

Turner, N. J., Willacy, K., Bryden, G., & Yorke, H. W. 2006, *ApJ*, 639, 1218

Umebayashi, T. 1983, *Progress of Theoretical Physics*, 69, 480

Umebayashi, T., & Nakano, T. 1981, *PASJ*, 33, 617

Valtonen, M., & Karttunen, H. 2006, *The Three-Body Problem*, Cambridge University Press

Velikhov, E. T. 1959, *Sov. Phys. JETP*, 36, 995

Veras, D., & Armitage, P. J. 2004, *MNRAS*, 347, 613

Wadhwa, M., Amelin, Y., Davis, A. M., Lugmair, G. W., Meyer, B., Gounelle, M., & Desch, S. 2007, *Protostars and Planets V*, eds B. Reipurth, D. Jewitt, and K. Keil, University of Arizona Press, Tucson

Ward, W. R. 1997 *Icarus*, 126, 261
Weidenschilling, S. J. 1977, *Astrophysics and Space Science*, 51, 153
Weidenschilling, S. J. 1977, *MNRAS*, 180, 57
Weidenschilling, S. J., & Marzari, F. 1996, *Nature*, 384, 619
Wetherill, G. W., & Stewart, G. R. 1993, *Icarus*, 106, 190
Whipple, F. L. 1972, in *From Plasma to Planet, Proceedings of the Twenty-First Nobel Symposium*, editor Aina Evlius. Wiley Interscience Division (New York), p. 211
Winn, J. N. 2009, *ApJ*, 700, 302
Winters, W. F., Balbus, S. A., & Hawley, J. F. 2003, *MNRAS*, 340, 519
Wolf, S., & D'Angelo, G. 2005, *ApJ*, 619, 114
Wolk, S. J., & Walter, F. M. 1996, *AJ*, 111, 2066
Wolszczan, A., & Frail, D. A. 1992, *Nature*, 355, 145
Yang, C.-C., Mac Low, M.-M., & Menou, K. 2009, *ApJ*, submitted (http://arxiv.org/abs/0907.1897)
Youdin, A. N., & Chiang, E. I. 2004, *ApJ*, 601, 1109
Youdin, A. N., & Goodman, J. 2005, *ApJ*, 620, 459
Youdin, A. N., & Shu, F. H. 2002, *ApJ*, 580, 494
Zhu, Z., Hartmann, L., & Gammie, C. 2009, *ApJ*, 694, 1045