Protostellar Disks and Planet Formation

P. Cassen

1 Introduction

The idea that the planets of the Solar System formed from a "protoplanetary disk" of material swirling about the primitive Sun follows naturally from the observation that the planetary orbital angular momentum vectors are nearly aligned with each other and that of the Sun itself. The existence of such a progenitor disk was implicit in the ideas of Descartes, and has been a common feature of scientific attempts to explain the systematic aspects of the Solar System since then. A corollary of these "nebular" theories is that planetary systems are an ordinary consequence of star formation. Modern astronomy has confirmed the essential aspects of the hypothesis by revealing the common existence of planets around other stars, and disks around young stars.

But the idea that planets form from circumstellar disks carries the further implication that the properties of planetary systems are somehow related to those of their parent disks; that is, an understanding of disk evolution leads naturally to an understanding of the nature of the resultant planetary system. Certainly this premise has been adopted in much of the theoretical work on the formation of the Solar System; indeed, it is the basis upon which much of the content of these lectures is organized. It may be, however, that nature contrives to obscure the conditions of planetary formation to the extent that the disk-planetary system connection is no longer recognizable, at least in some cases. We will discuss some lines of theoretical argument which suggest that this is the case. Certainly the dynamical properties of the extrasolar planetary objects discovered so far are not obviously associated with specific disk properties, or even a disk origin, despite the overwhelming circumstantial evidence that they must have formed from disks. It remains to be seen to what degree these bodies, or the Solar System, are representative of planetary systems in general.

Whether or not the particular properties of a protostellar disk are eventually reflected in the properties of a planetary system, the disk origin of planets is on firm ground. Furthermore, disks are complex and interesting objects in

Cassen P (2006), Protostellar disks and planet formation. In: Mayor M, Queloz D, Udry S and Benz W (eds) Extrasolar planets. Saas-Fee Adv Courses vol 31, pp 369–448
DOI 10.1007/3-540-29216-0_3

their own right. Thus the first half of these lectures are devoted to methods of elucidating their properties. The second half deals with theories of planet formation. The literature on these subjects is enormous, so my choice of specific subjects to be treated, as well as the references supplied, are somewhat subjective. I have tried to tie arguments to first principles wherever possible, although references must be relied upon for many details. Inevitably, the derivations of some results are too cumbersome for inclusion here, in which cases the reader is directed to the appropriate references.

2 Observations of Protostellar Disks

2.1 T Tauri Stars

Most of what we know about protostellar disks is derived from observations of T Tauri stars. These are pre-main sequence stars of spectral class G, K and M, originally identified by their observational characteristics: prominent Balmer emission lines, excess ultraviolet (UV) and infrared (IR) emission, variability and evidence for outflows. It is now known that the defining emission characteristics of T Tauri stars are due to the presence of circumstellar *accretion* disks. Lynden-Bell and Pringle (1974) identified the source of IR radiation as dissipation within the disk, due to angular momentum transport and the release of gravitational energy, as material is fed through the disk to the star. The UV radiation was attributed to gas heated to high temperatures by dissipation in a narrow boundary layer between the (rapidly rotating) inner edge of the disk and the surface of the (slowly rotating) star. It turns out that starlight shining on the disk, which is absorbed and re-emitted at longer wavelengths, also contributes substantially to the IR radiation. Also, there is evidence that the UV radiation is due to disk material falling onto the star along stellar magnetic flux tubes, rather than through the viscous boundary layer imagined by Lynden-Bell and Pringle (1974). Nevertheless, the theory developed by these authors (and their predecessor, (Lüst 1952)) forms the basis of the current understanding of protostellar disks, and correctly describes the continuum spectrum a T Tauri star in terms of several components (see Fig. 1): the nearly blackbody radiation from the star itself; a broad IR component from the optically thick part of the disk; microwave (submillimeter-to-millimeter) emission from the more distant, optically thin, parts of the disk; and an optical and UV component from hot gas transferred from the disk to the star.

Figure 2 shows the Hertzprung–Russellure (H–R) diagram for classical T Tauri stars and "weak-line" T Tauri stars (WTTS; premain sequence stars without evidence for disks) in the Taurus–Auriga star-forming region, along with theoretically derived pre-main sequence evolutionary tracks and their isochrons. Note that although some disks apparently last as long as 10^7 years, they appear to be gone by the time a star reaches the main sequence. Also,

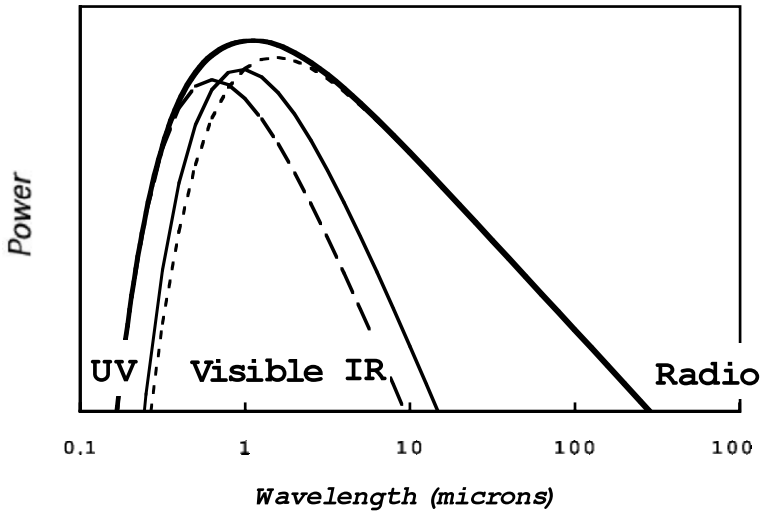

Fig. 1. Accreation disk theory predicts that the continuum spectrum of a pre-main sequence star and disk (*thick solid line*) is the sum of three components: the nearly blackbody radiation from the star (*thin solid line*), a broad component from the disk (*short dashed line*),and a blackbody-like component from hot gas being accreted from the disk to the star (*long dashed line*)

there are some stars that appear to be quite young and yet show no evidence for disks. These stars are of particular interest: either disks never formed; they did form, but were rapidly dissipated; or they rapidly became invisible because the dust in them coagulated to form larger objects (e.g., planets) which are not detected by standard means. These young "weak-line" T Tauri stars deserve more systematic study than they have so far received.

2.2 Interpretation of T Tauri Spectral Energy Distributions

An array of observational methodologies are now used to probe the properties of protostellar disks. High-resolution spectroscopy, optical and near-IR imaging, microwave interferometry (reviewed in chapters of *Protostars and Planets IV*) have all yielded fascinating, detailed information on individual objects. So far, however, the most general information has come from the determination of the spectral energy distributions (SEDs) of a large number star/disk systems, by multiwavelength photometry. The SED is defined as the quantity $\lambda F_\lambda(\lambda)$, where F_λ is the measured flux per unit wavelength λ [or equivalently, in terms of frequency, $\nu F_\nu(\nu)$]. The SED constrains the trends of several properties potentially important for planet formation: disk mass, temperature, accretion rate, lifetime and the degree to which solid particles have coagulated. Therefore, in the remainder of this section, we concentrate on the basics of interpreting SEDs, with an emphasis on the derivation of

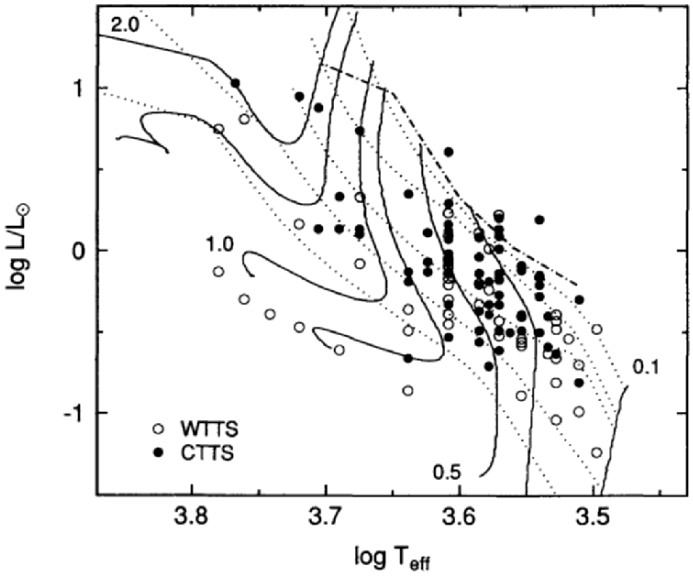

Fig. 2. Positions on the H–R diagram of classical T Tauri stars (*filled circles*) and "weak-line" T Tauri stars (*open circles*) for stars in the Taurus–Auriga star-forming region. The solid lines are theoretical evolutionary tracks, labeled by mass in units of $M_\odot$, from D'Antona and Mazzitelli (1994) The dashed lines are isochrons, corresponding to (*from the top down*): $10^5, 10^6, 3\times10^6, 10^7$ and 3×10^7 years, respectively. The dot-dash line is a theoretical upper limit pre-main sequence locations in the diagram (Stahler 1983). (Figure from Kenyon and Hartmann 1985)

fundamental relationships and the approximations commonly employed to obtain useful estimates of disk properties.

We start with some concepts from basic radiative transfer theory (Mihalas 1978). Define the *specific intensity* I_ν, to be the radiant energy $\mathrm{d}E$ flowing in direction $\mathbf{k}$ through an area $\mathrm{d}A$ (with normal in direction $\mathbf{n}$), in time $\mathrm{d}t$, frequency interval $\mathrm{d}\nu$ and solid angle $\mathrm{d}\Omega$:

$$\mathrm{d}E = I_\nu \mathbf{k}\cdot\mathbf{n}\,\mathrm{d}A\mathrm{d}\nu\mathrm{d}\Omega\ .$$

Then the *flux vector* is defined to be the moment

$$\mathbf{F}_\nu = \int I_\nu \mathbf{k}\,\mathrm{d}\Omega\ .$$

Thus the flux component F_ν, coming from a protostellar disk and passing into a telescope pointed at the disk is given by

$$F_\nu = \int (2\pi r\mathrm{d}r)\,\frac{\cos\theta}{d^2} I_\nu\ ,$$

where the integral is to be taken over all disk radii. Interpretation of the measurement of F_ν therefore requires that the quantity I_ν be understood in terms of disk properties. Now the significance of I_ν is that it obeys a conservation law of the form

$$\frac{\partial I_\nu}{\partial t} + c\mathbf{k}\cdot\nabla I_\nu = \text{sources} + \text{sinks} \ . \tag{1}$$

The first term on the left is negligible in cases where changes in the source and intervening medium are slow compared to the light travel time. In a vacuum, the right hand side is zero. In the presence of matter, the source and sink terms are expressed as

$$\begin{aligned} \text{sources} &= \frac{\varrho\varepsilon_\nu}{4\pi} + (\text{scattering terms}) \\ \text{sinks} &= \varrho\kappa_\nu I_\nu + (\text{scattering terms}) \ , \end{aligned}$$

where ϱ is the material mass density, ε_ν is the emissivity/mass and κ_ν is the absorption opacity (the latter two being wavelength-dependent). The scattering terms quantify the amount of radiation scattered into and out of the relevant direction and wavelength interval, and we assume them to be unimportant for the present purposes. In the case of *local thermodynamic equilibrium* (LTE), in which the radiative state of material is defined solely by its temperature, the relationship $\varepsilon_\nu = 4\pi\kappa_\nu B_\nu(T)$ holds, where $B_\nu(T)$ is the Planck function:

$$B_\nu(T) = \frac{2h\nu^3/c^3}{\mathrm{e}^{\,h\nu/kT} - 1} \ .$$

Thus, (1) can be written

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}\tau_\nu} + I_\nu = S_\nu, \tag{2}$$

where

$$\tau_\nu = \int_0^S \varrho\kappa_\nu \, \mathrm{d}s'$$

is the optical depth along the ray path length s, and the source function S_ν is given by

$$S_\nu = \frac{\varepsilon}{4\pi\kappa_\nu} = B_\nu(T) \ .$$

The last equality holds for LTE. Equation (1) has the formal solution

$$I_\nu(\tau_\nu) = I_\nu(0)\mathrm{e}^{\,-\tau_\nu} + \int_0^{\tau_\nu} B_\nu(T) \exp\left[-\left(\tau_\nu - \tau'_\nu\right)\right] \mathrm{d}\tau'_\nu \ .$$

For the radiating disk, we can measure the optical depth along the normal to the disk ($\tau_{\nu\perp}$) and from the midplane, in which case $I_\nu(0) = 0$ and $\tau_\nu = \tau_{\nu\perp}/\cos\theta$. If $B_\nu(T)$ was independent of τ_ν, the integral would yield

$$I_\nu = B_\nu(T)\left(1 - \exp\left(-\tau_{\nu\perp}/\cos\theta\right)\right)$$

and

$$F_\nu = \int (2\pi r\,\mathrm{d}r)\,\frac{\cos\theta}{d^2} B_\nu(T)\left(1 - \exp\left(-\tau_{\nu\perp/\cos\theta}\right)\right)\ . \tag{3}$$

Beckwith et al. (1990) used this expression to derive estimates of disk masses and temperature distributions from the SEDs of about 40 T Tauri stars, as described below. Of course the function $B_\nu(T)$ is generally *not* independent of τ_ν, so (3) represents an approximation. The approximation, however, does make sense in the limits of large and small disk optical depth, as long as care is taken in the interpretation of the value of the temperature T in $B_\nu(T)$.

The inner parts of protostellar disks are usually optically thick, even at far IR and millimeter wavelengths. For those regions, the exponential term in (3) can be neglected, and T is clearly to be regarded as a photospheric, or effective radiating temperature, T_e. If one represents its radial distribution as a power law, $T_e = T_{eo}\,(r/r_o)^{-q}$, a transformation of the variable of integration leads to

$$\nu F_\nu = A\,(T_{eo})\,\nu^{4-\frac{2}{q}}\left(\frac{cos\theta}{d^2}\right)\int \frac{x\,\mathrm{d}x}{\exp\left(x^q\right) - 1}\ .$$

Thus the slope and magnitude of the SED (the latter modulated by the inclination and distance to the disk) correspond to a particular power law effective temperature distribution.

Simple theoretical arguments led to the expectation that the value of q would be 3/4. First, the power radiated by an accretion disk at any distance was expected to scale as the radial mass flux through an annulus of the disk, times the gravitational energy per unit mass of the annulus. The expression of this power in terms of an effective temperature yields

$$2\,(2\pi r\mathrm{d}r)\,\sigma T_e^4 \propto \dot{M}_d d\left(-\frac{GM}{2r}\right) = \dot{M}_d \frac{GM}{2r^2}\,\mathrm{d}r$$

or $T_e^4 \propto r^{-3/4}$. (This expression turns out to be correct, except for the constant of proportionality; see Sect. 3.1)Second, stellar radiation impinging on a flat, optically thick disk would be absorbed and re-radiated at a temperature T_e' according to

$$T'^4_e = T_*^4\left(\frac{r_*}{r}\right)^2 \sin\alpha \approx T_*^4\left(\frac{r_*}{r}\right)^2\left(\frac{r_*}{r}\right)$$

(where α is the angle upon which starlight impinges the disk), which again implies $T'^4_e \propto r^{-3/4}$. In fact, Beckwith et al. (1990) found q to be greater than 3/4 (Figs.3 and 4), with the most frequent value being close to 1/2, which corresponds to a flat SED in the IR. This observation led to the development of more sophisticated models of disk thermal structure than provided by the simple arguments given above. It was recognized, for instance, that if the surface

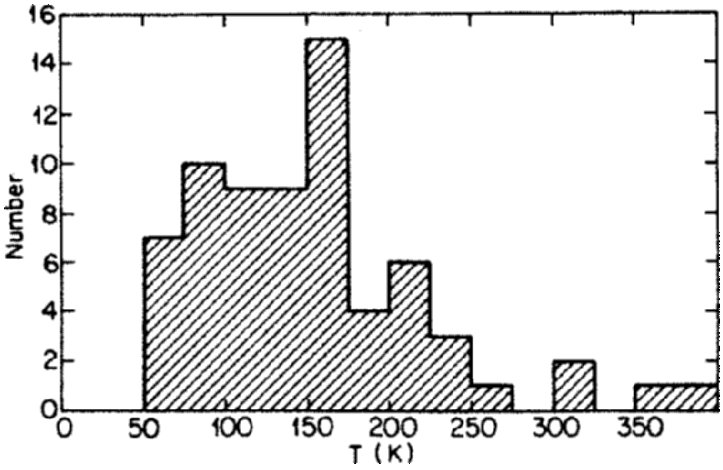

Fig. 3. Histogram of disk temperatures at 1 AU inferred by modeling SEDs, for a sample of T Tauri stars. (Figure from Beckwith et al. 1990)

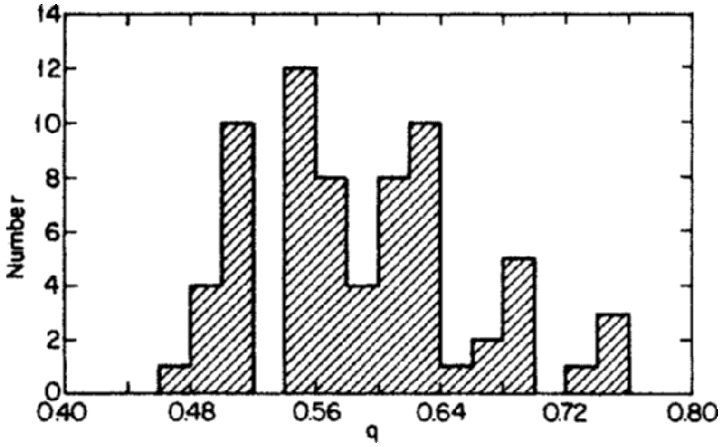

Fig. 4. Histogram of disk temperature power law indices, q, inferred from SEDs, for a sample of T Tauri stars. Most values of q are less than that predicted for accretion or reprocessing from a flat disk, which indicates that SEDs are generally flatter than predicted by the simplest models. (Figure from Beckwith et al. 1990)

of the disk flared upwards with distance from the star, proportionally more stellar radiation would be absorbed and re-emitted at larger distances, which would produce a flatter SED in the IR (Kenyon and Hartmann 1985). In particular, Chiang and Goldreich (1997) analyzed the two-layer model illustrated in Fig. 5.

Dust particles in the upper layer of the disk absorb stellar radiation and emit IR radiation, both outward to space and downward, toward the disk midplane. Because they are smaller than IR wavelengths ($\leq 1\,\mu$m), the particles are inefficient emitters and must become superheated (relative to the gas) in order to attain thermal equilibrium with the radiation they absorb. The lower layers of the disk are heated by the radiation from the dust. The vertical extent of the disk is determined by a balance between pressure and the vertical component of stellar gravity at every radius. In this model, the net effect of this balance is that the thickness of the disk increases faster than linearly with radius, i.e., it flares upward. Both the superheating of the dust and the flaring contribute to flattening the SED relative to that which would correspond to $T'_e \propto r^{-3/4}$, as shown in Fig. 6. The essential features of the Chiang and Goldreich (1997) model are evident in the results of D'Alessio et al. (2001), who analyzed disk structure in much greater detail. This work shows that heating by stellar radiation is the primary determinate of $T_e\,(r)$ beyond a few AU, and therefore of the slope of the SED at IR and submillimeter wavelengths. Within a few AU, internal

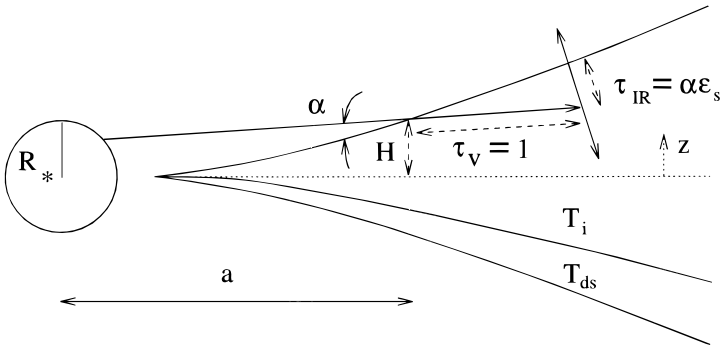

Fig. 5. The two-layer disk model of Chiang and Goldreich (1997), used to explain the flat continuum spectrum of many pre-main stars. Stellar radiation, incident on the disk at distance a, height H, and angle α, is absorbed by dust within one (visible wavelength) optical depth along its path into the upper layers of the disk. The radiation is re-emitted at infrared wavelengths, outward to space and inward toward the midplane. Because the dust particles are too small to be efficient radiators in the infrared, their temperature T_{ds} exceeds that of the surrounding gas, and the temperature T_i of the interior dust (which absorbs and emits the IR radiation from the upper layers with equal efficiency). (Figure from Chiang and Goldreich 1997)

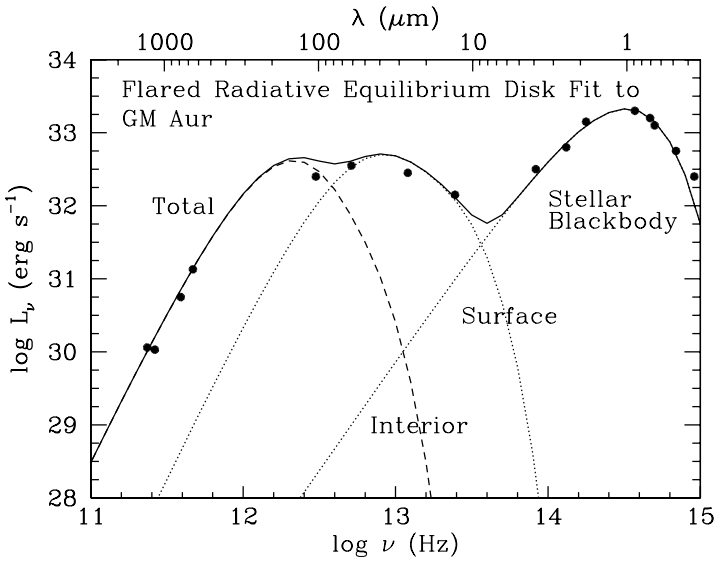

Fig. 6. A fit to the SED of the T Tauri star GM Aur, using the two-layer model. (The SED here is plotted as a function of frequency, with wavelength along the top. Astronomers differ in their preferences of independent variable.) The contribution of the hot surface dust layer dominates throughout the IR, falling below that from the interior only at wavelengths long enough for the disk to be optically thin. (Figure from Chiang and Goldreich 1997)

dissipation associated with disk accretion is important for determining the thermal structure.

At wavelengths where large parts of the disk are optically thin (typically, $\lambda > 1$ millimeter), (3) may be written

$$F_\nu = \frac{1}{d^2} \int (2\pi r \, \mathrm{d}r) \, \frac{2\nu^2 kT}{c^2} \Sigma(r) \kappa_\nu \ .$$

The quantity $\Sigma(r)$ is the surface density. To obtain this expression, the exponential was approximated in the limit of $\tau_{\nu\perp} << 1$, the Planck function was approximated by the Rayleigh–Jeans form $B_\nu \approx 2\nu^2 kT/c^2$, and the substitution $\tau_{\nu\perp} = \Sigma(r)\kappa_\nu$ was made. In protostellar disks, as in the cool interstellar medium, the opacity is determined essentially by the size and composition of dust. Except near resonant absorptions, it can be expressed as a power of frequency, $\kappa_\nu = \kappa_0 (\nu/\nu_0)^\beta$, in which case

$$\nu F_\nu = \frac{A}{d^2} \nu^{3+\beta} \kappa_0 \int (2\pi r \, \mathrm{d}r) \, T(r) \, \Sigma(r) \ .$$

The quantity A is a known constant. In principle, β can be determined by observations alone, through measurements of the same system at different frequencies. The integral then represents a "temperature-weighted" disk mass.

Beckwith et al. (1990) constructed composite models of the SEDs of stars observed at 1.3 and 2.7 mm, by combining their radio data with ground-based optical photometry and IR measurements obtained by the Infrared Astronomical Satellite, and so obtained the disk mass estimates shown in Fig. 7. The masses are mainly in the range 10^{-3}–$10^{-1}\,M_\odot$, and so are typically within an order of magnitude of the minimum mass inferred for the primitive solar nebula, $10^{-2}\,M_\odot$ (Weidenschilling 1977). The lack of correlation of disk mass with age is not understood, but may reflect the fact that it is really *dust* that is being observed, and the mass of dust present in a system could be the fluctuating result of a residual interstellar component, loss by coagulation, and production by collision and fragmentation of larger bodies.

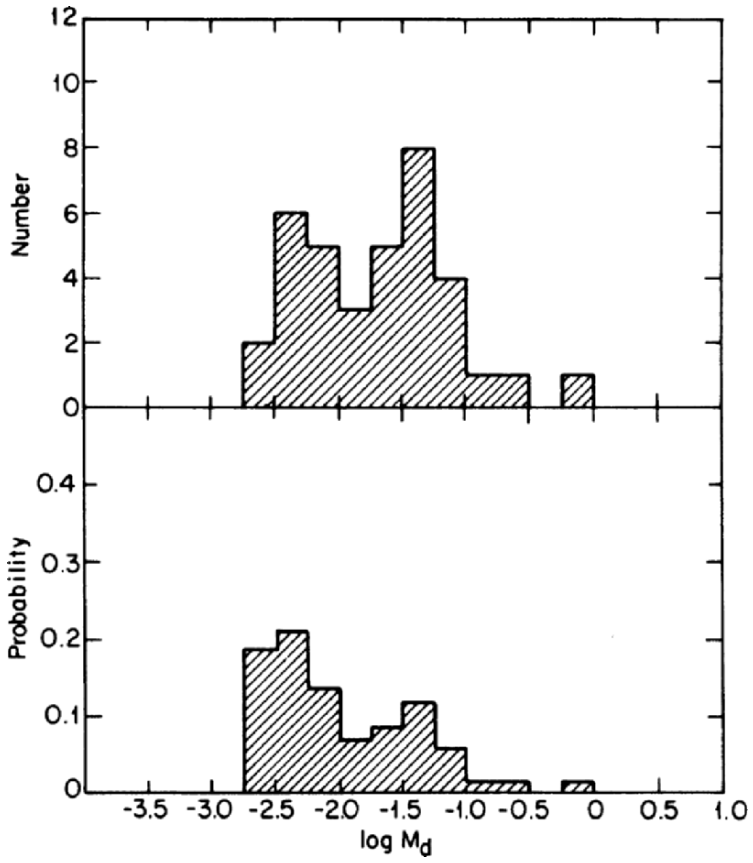

Fig. 7. Histograms of disk masses (measured in $M_\odot$) inferred by modeling SEDs, for a sample of T Tauri stars. The upper panel is that deduced directly from the data; the lower panel is an inferred distribution which accounts for sampling bias. Values of disk masses are subject to uncertainties associated with dust emissivity and abundance. (Figure from Beckwith et al. 1990)

The exponent β is expected to lie in the range $0 < \beta < 2$, with the smaller values favored as particles grow. Comparisons of observed values of β and κ_0 with laboratory measurements and theoretical calculations of absorptivity should then yield information on particle growth in disks (**?**). There is, in fact, evidence from multiwavelength imaging and SED fits (D'Alessio et al. 2001; Throop et al. 2001) that particle growth to mm-sized pebbles is observed, but compelling, quantitative results are difficult to obtain because of ambiguities in the modeling; see Beckwith et al. (2000) for a discussion.

2.3 Accretion Rates of Protostellar Disks

The accretion rate of a protostellar disk is an important quantity because it determines the amount of energy released *within* the disk, which is the primary determinant of midplane temperatures in optically thick parts of the disk. Optical and UV radiation provide better diagnostics of accretion rates than IR radiation, because the latter tends to be dominated by reprocessing of stellar emission. Figure 8 shows a comparison between the optical spectrum of the T Tauri star BP Tau and the WTTS LkCa7. The notable features of the BP Tau spectrum are the excess emission at shorter wavelengths, the prominent Balmer lines (at 4861, 4340, 4101 Å...; H_α at 6563 Å not shown),

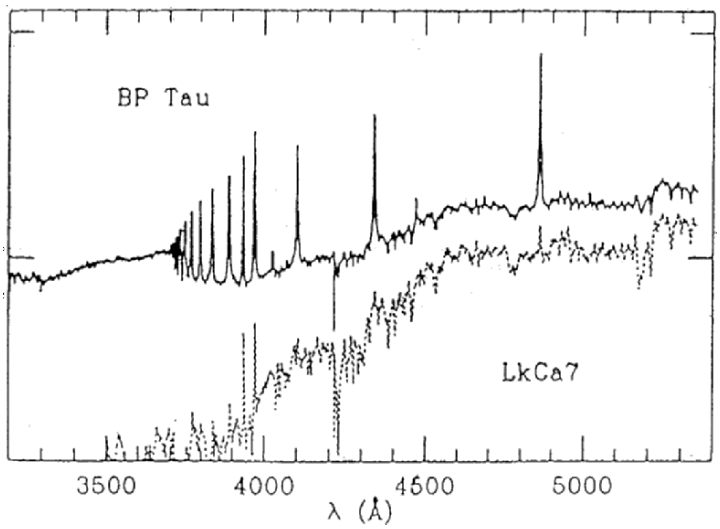

Fig. 8. Optical spectra of the T Tauri star BP Tau (*solid*) and the weak-line T Tauri star LkCa7 (*dotted*). The BP Tau spectrum exhibits excess emission at short wavelengths, prominent Balmer lines and the narrower absorption lines, features which are typical of stars with disks. They are explained by the accretion of hot, infalling gas onto the star. The normal, pre-main sequence photosphere of LkCa7 indicates that this star is not undergoing accretion; its lack of IR excess indicates that it has no disk. (Figure from Hartmann 1998)

and the narrower absorption lines. These features are typical of T Tauri stars, and are attributed to the presence of hot (several thousand K) gas, overlying a normal, pre-main sequence photosphere, such as that of LkCa7. There is evidence that the hot gas is produced by infalling material, channeled along stellar magnetic field lines, from the inner edge of the accretion disk to the surface of the star (Hartmann 1998).

Gullbring et al. (1998) and Calvet and Gullbring (1998) have derived accretion rates for T Tauri stars based on this concept. They assume that the two components (photosphere and hot, infalling gas) contribute to the optical spectrum, so the observed optical and UV luminosity $L_{\rm obs} = L_{\rm photosphere} + L_{\rm accretion}$. The excess continuum emission due to accretion is determined by the subtraction of a "template" spectrum provided by a WTTS of the same spectral type, by the method described by Hartigan et al. (1998). Special procedures must be used to account properly for extinction and UV emission which cannot be observed from the ground. Accretion is then modeled as a one-dimensional flow from a distance R_i (the inner edge of the disk, typically several stellar radii) down a magnetic flux tube which intersects some fraction f of the stellar surface (Fig. 9). The inflow becomes supersonic, and must therefore pass through a shock before hitting the star.

The accretion rate is given by

$$\dot{M}_d = \varrho \left(4\pi r_*^2 f\right) \nu_s \ ,$$

where ν_s and ϱ are the gas velocity and density at the shock. The former is taken to be the free-fall velocity from R_i:

$$\nu_s^2 = \frac{2GM_*}{r_*} \left(1 - \frac{r_*}{R_i}\right) \ .$$

The density is determined by the assumption that the shock location occurs at optical depth unity above the star, where pressure balance between the stellar atmosphere and the inflow requires:

$$p_s = \frac{1}{2}\varrho\nu_s^2 = g_* \Sigma_s = g_* \frac{\tau}{\kappa} \ .$$

Here g_* is the stellar gravity, Σ_s is the surface density of the shock layer, κ is the mean opacity and the optical depth $\tau = 1$. Thus, for known M_*, r_*, g_*, R_i and κ, the flux of energy from the accretion column can be calculated:

$$F_{\rm accretion} = \left(\frac{1}{2}\varrho\nu_s^2\right) \nu_s \ .$$

The total accretion luminosity is

$$L_{\rm accretion} = \left(4\pi r_*^2 f\right) F_{\rm accretion} \ ,$$

which then, in principle, determines f. Comparisons of calculated spectra with excess emission determined from observations, and corresponding values of $F_{\rm accretion}$ and f, are shown in Fig. 10. An important result is that accretion apparently occurs on only 1% or less of the stellar surface.

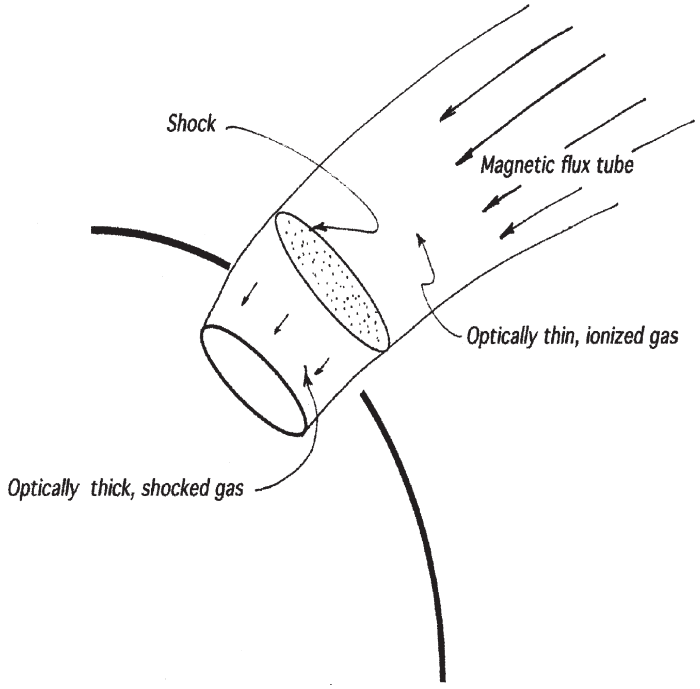

Fig. 9. Accretion rates can be obtained from models in which it is assumed that gas from the inner edge of the disk falls onto the star along stellar magnetic flux tubes. Optically thin gas, flowing supersonically toward the star, passes through a shock wave near the stellar surface. Radiation from the shocked gas heats both the stellar photosphere and the infalling gas above it, which produce excess UV and optical radiation and the characteristic emission features of accreting stars

Accretion rates determined by these methods are shown as a function of stellar age in Fig. 11. Note that a typical accretion rate of $10^{-8}\,M_{\odot}$/year is consistent with a disk mass of $10^{-2}\,M_{\odot}$ for a million year old disk. Note also that active (accreting) disks last for many thousands of dynamical (rotation) periods; thus disk evolution is gradual and the processes that cause it are subtle.

2.4 Internal Temperatures of Protostellar Disks

Temperatures within protostellar disks are not directly observable because most disks are optically thick within a few AU of the star, even at long ($\lambda \geq 1\,\mathrm{mm}$) wavelengths. Therefore, some model of the vertical structure, like those referred to above, must be constructed to relate the observed surface fluxes to the internal state. A key parameter in such a model is the rate of energy dissipation within the disk, which is related to the accretion rate, as described below in Sect. 1.3.1. Midplane temperatures for disks around T

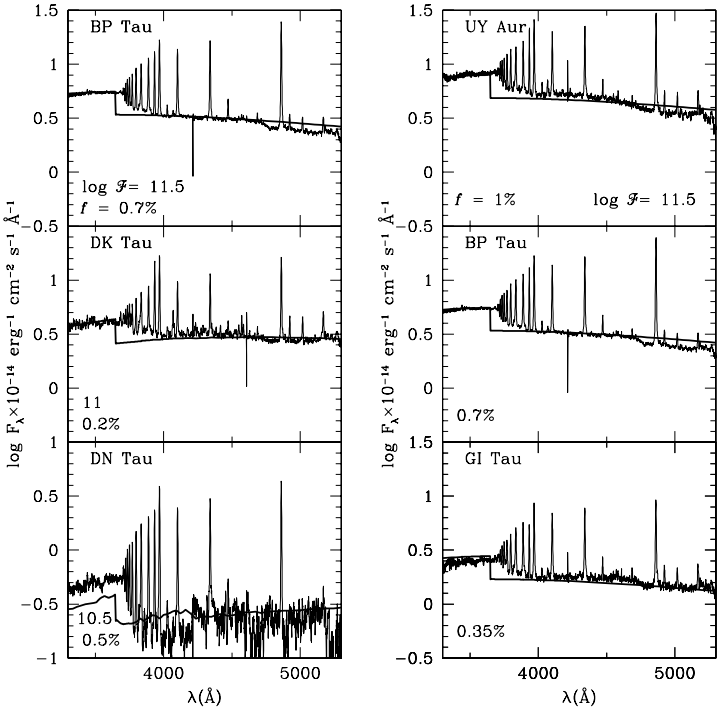

Fig. 10. Comparisons of calculated continuum spectra (*thick line*) with excess emission determined from observations, for a sample of T Tauri stars. Values of $F_{\rm accretion}$ and f determined from the model illustrated in Fig. 10 are given. Note that $f \leq 0.1$; that is, accretion apparently occurs on only 1% or less of the stellar surface. (Figure from Calvet and Gullbring 1998)

Tauri stars were estimated by Woolum and Cassen (1999), by combining the results described above for disk surface temperatures, masses and accretion rates, with a simple model based on radiative transport through a plane-parallel atmosphere. The observed IR flux was treated as the sum of that due to internally released accretion energy and a reprocessed stellar component. In optically thick parts of the disk, the former largely controls the midplane temperature, while the latter dominates the observed flux. They concluded that midplane temperatures at 1 AU are mainly in the range 200–800 K, for disks with ages of about 1 million years. At the low pressures in these disks ($< 10^{-3}$bars), H_2O would exist as vapor within a few AU of the star (wherever the midplane temperature exceeds 160 K), and so would not be readily incorporated into planetary objects. Icy objects could form beyond 2–3 AU. They also argued that if very young disks were characterized by accretion rates as high as 10^{-6}M$_\odot$/year (Fig. 11), midplane temperatures would then be high enough to vaporize even the rock-forming elements (primarily Fe, Mg, Si) at a few AU.

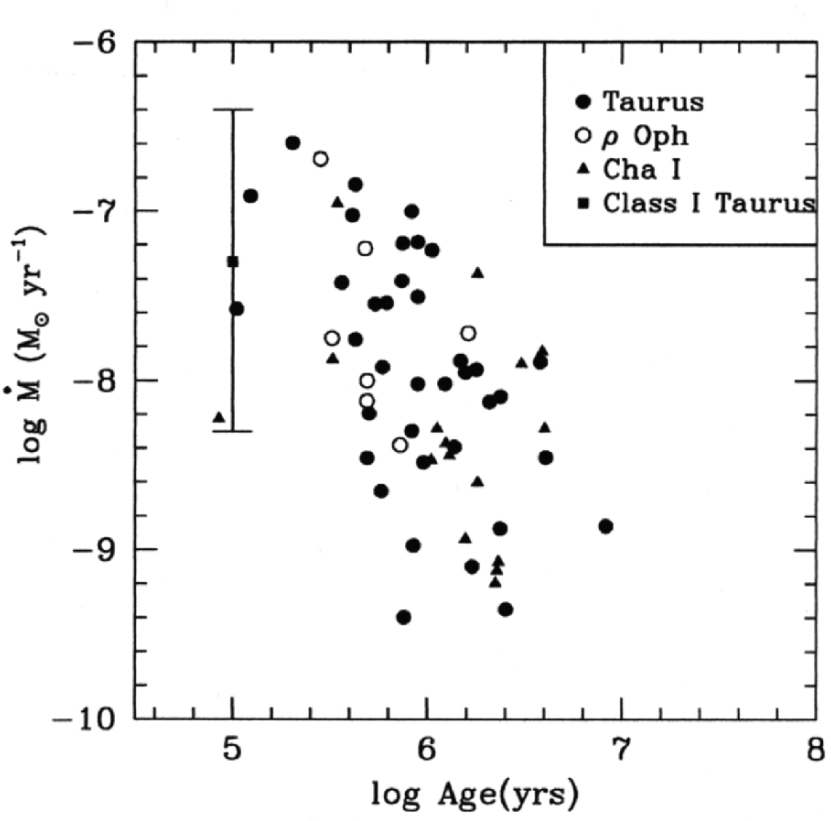

Fig. 11. Mass accretion rates and ages inferred for T Tauri stars in three star-forming regions. The vertical line indicates the mean and dispersion of accretion rates estimated for embedded stars (assumed to be 10^5 years old). (Figure from Calvet et al. 2000)

3 Theory of Disk Structure and Evolution

3.1 Conservation Equations

The equations that govern the structure and evolution astrophysical disks are derived from the equations of mass, momentum and energy conservation, which have the general form

$$\frac{\partial}{\partial t}\begin{pmatrix}\varrho\\ \mathbf{v}\\ e\end{pmatrix}+\nabla\cdot\left[\mathbf{v}\begin{pmatrix}\varrho\\ \mathbf{v}\\ e\end{pmatrix}\right]=\begin{pmatrix}0\\ \text{forces}\\ \text{sources}+\text{transport}\end{pmatrix}\ .$$

When applied without reference to a specific geometry, they lead to the equations of fluid motion, which are usually expressed as

$$\frac{\partial \varrho}{\partial t} + \nabla \cdot (\varrho \mathbf{v}) = 0 \quad \text{(mass)}$$

$$\varrho \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) = -\nabla p + \nabla \cdot \mathbf{w} - \varrho \nabla \Phi \quad \text{(momentum)}$$

$$\varrho \left(\frac{\partial e}{\partial t} + \mathbf{v} \cdot \nabla e \right) = -p \nabla \cdot \mathbf{v} - \nabla \cdot \mathbf{F} + D \quad \text{(energy)} \ .$$

In these equations, $\mathbf{w}$ is the non-diagonal part of the stress tensor, w_{ij} being the stress on a j-facing surface in the i-th direction; Φ is the gravitational potential; $\mathbf{F}$ is the energy flux vector due to radiation, conduction, or other means; and D is the rate of energy dissipation associated with stresses:

$$D = w_{ij} \frac{\partial v_i}{\partial w_j} \ .$$

Magnetic fields have been ignored in these equations, as they will be in the rest of this section; but they can be important and will be discussed later.

The equations of disk structure and evolution are most illuminatingly derived by application of the conservation equations in their general form to a control volume such as that shown in Fig. 12. Variations in ϕ are not considered. Because the disk is assumed to be thin, it is practical to consider the values of z-integrated quantities, such as the surface density:

$$\Sigma = \int_{-\infty}^{+\infty} \varrho \, \mathrm{d}z \ .$$

Thus, the conservation of mass equation , integrated over z, is:

$$\frac{\partial \Sigma}{\partial t} + \frac{1}{r} \frac{\partial \dot{m}}{\partial r} = 0, \tag{4}$$

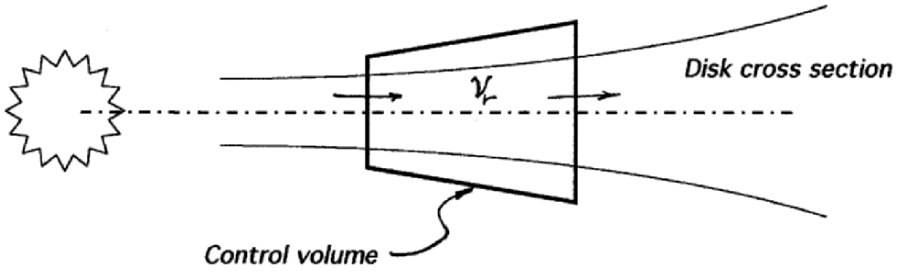

Fig. 12. Control volume for the derivation of conservation equations for a disk with radially flowing material and no azimuthal gradients. The radial velocity v_r is taken positive outward by convention, but will generally be inward in the inner parts of the disk and outward in the outer parts

where

$$\dot{m} = \int_{-\infty}^{+\infty} \varrho v_r r \, \mathrm{d}z$$

is $1/2\pi$ times the radial mass flux through the disk (the "reduced" mass flux). In what follows, it is assumed that the radial velocity v_r varies little over the disk thickness, so $\dot{m} = \varrho v_r r$.

To a high degree of accuracy, the vertical momentum balance is hydrostatic, just as in a planetary atmosphere. Pressure forces are balanced by gravity:

$$0 = -\frac{\partial p}{\partial z} + \varrho \frac{\partial \Phi}{\partial z} \ .$$

The gravitational potential consists of two terms, one due to the vertical component of stellar gravity and one due to the self-gravity of the disk. The latter is negligible for stable, unperturbed disks, so

$$\frac{\partial \Phi}{\partial z} = \frac{\partial \Phi_*}{\partial z} = -\frac{GM_*}{(r^2+z^2)} \frac{z}{r} \approx -\frac{GM_*}{r^2} \frac{z}{r} = -\Omega_K^2 z \ .$$

The quantity $\Omega_K^2 = GM_*/r^3$ is the square of the Keplerian frequency, i.e., the angular velocity of a freely orbiting object in a circular orbit at distance r from the star. Thus,

$$\frac{1}{\varrho} \frac{\partial p}{\partial z} = -\Omega_K^2 z \ . \tag{5}$$

If we characterize the distance z by the thickness of the disk measured, say, in terms of a scale height h, and note that the quantity p/ϱ is, to order unity, the sound speed c_s, we find that the relative thickness $h/r = c_s/r\Omega_K = c_s/V_K << 1$, for a thin disk. ($V_K$ is the Keplerian velocity.) Disk models typically yield values of h/r in the range 10^{-2}–10^{-1}.

The radial momentum equation is:

$$\frac{\partial \left(v_r \Sigma\right)}{\partial t} + \frac{1}{r} \frac{\partial \left(v_r \dot{m}\right)}{\partial r} = \Sigma \left(\frac{\partial \Phi}{\partial r} + r\Omega^2 \right) - \frac{\partial P}{\partial r},$$

where $P \equiv \int_{-\infty}^{+\infty} p \, \mathrm{d}z$ and $\partial\Phi/\partial r = -GM_*/r^2$. It is readily seen that, as long as the characteristic evolution time r/v_r is much longer than the orbital period $2\pi/\Omega$, the terms on the left are much smaller than the first term on the right-hand side. Also, the pressure gradient is a factor of $(h/r)^2$ smaller than the first term on the right. Thus, to this order, $\Omega = \Omega_k$; the orbital motion is Keplerian. (It turns out that the small radial pressure gradient is important when considering the fate of solid objects in the disk, as discussed in Sect. 1.4.1.).

Disk evolution is addressed directly by the angular momentum equation, which is:

$$\frac{\partial\left(\Sigma j\right)}{\partial t}+\frac{1}{r}\frac{\partial\left(\dot{m}j\right)}{\partial r}=\frac{1}{r}\frac{\partial\left(r^2W_{r\phi}\right)}{\partial r}+\frac{1}{2\pi}\frac{\partial T}{\partial r}\ .$$

Here, j is the angular momentum/mass $= r^2\Omega_K = rV_K = (GM_*r)^{1/2}$, T is any externally applied torque, and

$$W_{\phi r}=\int_{-\infty}^{+\infty}w_{\phi r}\,\mathrm{d}z\ .$$

The mass conservation equation (4) can be used to solve for the reduced mass flux:

$$\dot{m}=\frac{r}{j}\frac{\partial}{\partial r}\left(\frac{T}{\pi}+2r^2W_{\phi r}\right)\ . \tag{6}$$

This equation and (4) specify the rearrangement of surface density in terms of the forces that affect angular momentum (we omit the external torque term from now on, for brevity):

$$\frac{\partial\Sigma}{\partial t}=-\frac{2}{r\sqrt{GM_*}}\frac{\partial}{\partial r}\left[r^{1/2}\frac{\partial}{\partial r}\left(r^2W_{\phi r}\right)\right]\ .$$

Now it is usually supposed that $W_{\phi r}$ is a *Newtonian stress*; that is, it is linearly proportional to the rate of strain (as is true for many viscous fluids). If this is the case,

$$W_{\phi r}=\Sigma\nu r\frac{\partial\Omega}{\partial r},$$

where we follow standard notation and use ν for the kinematic viscosity (not to be confused with frequency, in Sect. 1.2). The disk evolution equation is then

$$\frac{\partial\Sigma}{\partial t}=\frac{3}{r}\frac{\partial}{\partial r}\left[r^{1/2}\frac{\partial}{\partial r}\left(r^{1/2}\Sigma\nu\right)\right]\ . \tag{7}$$

This equation has the form of a diffusion equation for Σ, and the problem of disk evolution has been essentially reduced to one of determining the proper expression for ν. For instance, if ν were given as any function of r, the equation would be linear, and could be solved by any of a number of standard techniques (Lynden-Bell and Pringle 1974). In general, solutions for disks which conserve overall angular momentum (but can lose energy) have the properties shown schematically in Fig. 13. Material in the inner parts of the disk lose angular momentum and spiral in to the central star, while material in the outer disk gains angular momentum and expands outward. This very

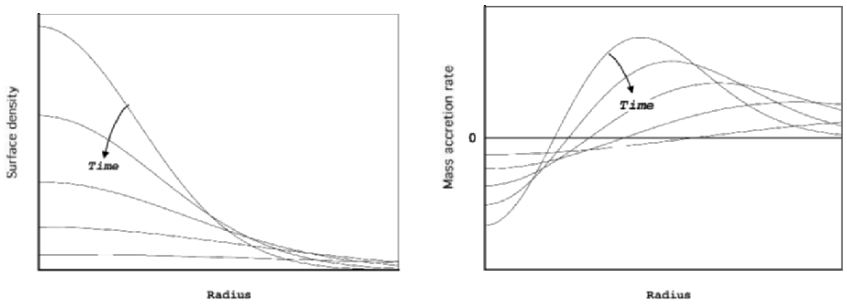

Fig. 13. Schematic representation of the variations of surface density (*left*) and accretion rate (*right*), as functions of time and radius. Surface density decreases in the inner parts of the disk as the disk spreads. The mass accretion rate is negative (inwards) in the inner parts, nearly independent of radius and diminishing in magnitude with time. There is a stagnation radius, at which the radial motion changes direction; the location of the stagnation radius moves outward with time

fundamental behavior of accretion disks can be demonstrated by the application of simple conservation arguments, and is independent of the particular mechanism(s) of angular momentum transport (see Lynden-Bell and Pringle 1974).

What is the value of ν indicated by observations? The viscous evolution time is given by r_d^2/ν, where r_d is a characteristic disk size, say, 100 AU. A disk age of 1 million years then yields a kinematic viscosity of about $10^{15}\,\mathrm{cm}^2/\mathrm{sec}$, far greater than the ordinary molecular viscosity of hydrogen gas, which is about $10^6\,\mathrm{cm}^2/\mathrm{sec}$ for conditions appropriate for these disks. It is therefore commonly supposed that turbulence in the disks enhances the viscosity far above the molecular value. The turbulent kinematic viscosity (or eddy diffusivity, as atmospheric scientists call it) can be expressed as the product of a turbulent velocity and a mixing length. If the latter is as large as the scale height of the disk, relatively modest turbulent velocities, about $10^3\,\mathrm{cm/sec}$ would suffice to provide the inferred value of the viscosity.

The concept of turbulent viscosity gives rise to a prescription originally postulated by Shakura and Sunyaev (1973), in another context. They assumed

$$\nu = \alpha c_s h$$

where the parameter $0 < \alpha < 1$, as expected if the scale height and sound speed are upper limits to the mixing length and turbulent velocity, respectively. This widely adopted formulation shifts the burden of quantifying angular momentum transport from ν to α. If one adopts expected values of c_s and h (say 1 km/sec and 0.1 AU, respectively), a value of α between 10^{-3} and 10^{-2} would provide the necessary viscosity. Useful as such prescriptions may be for some purposes, one should bear in mind that the clear inadequacy of molecular viscosity renders even the assumption of Newtonian viscosity

suspect. For this reason, it is prudent to recognize the more basic forms of the evolution (4) and (6).

The energy lost from an evolving disk contributes to the observed radiation and provides diagnostic information, as discussed in Sect. 1.2. If the dissipated energy is expressed in terms of an effective radiating temperature T_e, the energy conservation equation is

$$\frac{\partial\left(\Sigma e\right)}{\partial t}+\frac{1}{r}\frac{\partial\left(\dot{m}e\right)}{\partial r}=\frac{1}{r}\frac{\partial\left(rW_{\phi r}v_{\phi}\right)}{\partial r}-2\sigma T_e^4+\left(\text{external irradiation}\right)\ . \quad (8)$$

The thermal energy is $(h/r)^2$ times smaller than the kinetic and gravitational energies, so the pressure work term has been ignored, and a good approximation for the energy/mass is

$$e=\frac{v_{\phi}^2}{2}+\Phi=-\frac{GM_*}{2r}\ .$$

The last term in (8) may include energy deposition from the star, the background radiation field, and radiation from other parts of the disk itself. Ignoring it for the moment, and using the mass conservation equation to eliminate terms in (8), one finds

$$2\sigma T_e^4=rW_{\phi r}\frac{\partial\Omega_k}{\partial r}=-\frac{3}{2}W_{\phi r}\Omega_K\ .$$

This formula can be expressed in terms of the disk mass flux by noting that solutions of the disk evolution equation generally produce an r-independent $\dot{m}$ over some portion of the inner disk. In that region, (5) can be integrated to obtain

$$W_{\phi r}=\frac{\dot{m}j}{r^2}\ .$$

Therefore

$$\sigma T_e^4=\frac{3\dot{M}_dGM}{8\pi r^3}\ ,$$

where $\dot{M}_d=-2\pi\dot{m}$. Note that this frequently encountered formula for the distribution of internally generated, radiated energy does not depend on the nature of the stresses causing angular momentum transport. Its validity does require that energy be dissipated "locally"; that is, that radial transport of energy is negligible compared to vertical transport, and that the disk material behave as a Newtonian fluid. And, as discussed above, it is not the sole source of the observed radiated energy, which can be dominated by external irradiation which is absorbed and re-emitted by the disk.

Finally, it is important to have a description of the vertical structure of the disk, especially when considering the accumulation of solid material into planets. Equation (5) gives the mechanical balance. The thermal state must be ascertained by an energy equation of the form

$$\frac{\partial F_z}{\partial z} = -\left(\text{dissipation}\right) + \left(\text{external irradiation}\right),$$

where F_z is the flux of energy in the z direction and the terms on the right must now be specified as functions of z. Although F_z may have radiative and convective components, the latter is usually unimportant for heat transport in disks (Cassen 1993; D'Alessio et al. 1999). Several other factors, however, do complicate vertical structure models. First, without a detailed model for the stresses responsible for angular momentum transport, some assumption must be made regarding the vertical distribution of dissipation. Usually it is assumed that the dissipation rate is proportional to the local mass density, although this need not be the case in a real disk. Second, the evaporation, condensation and coagulation of dust and ice, which are the major contributors to the opacity (Pollack et al. 1994; Henning and Stognienko 1996), must be solved for self-consistently. Third, the effect of the external radiation can be a complicated function disk radial structure (Bell 1999).

The nature of the most useful approximations and assumptions employed to determine vertical structure may depend on the particular objective. Studies directed toward the astronomical appearance of disks must properly account for the external illumination, but may simplify the issues of opacity structure and internal dissipation, as in, for instance, Chiang and Goldreich (1997). Studies directed toward understanding the thermal conditions of planet-building may simplify the treatment of external irradiation, which has a minor effect on midplane temperatures where disks are optically thick, but must account for the vertical distributions of dissipation and opacity (e.g., Cassen 2001).

A simplification that is usually valid is to treat the vertical structure as "quasi-steady". That is, the time for even an optically thick disk to adjust to new heating and cooling conditions,

$$t_{\text{thermal}} = \frac{\Sigma c_s^2}{2\sigma T_e^2}$$

is usually much shorter than other evolutionary timescales (e.g., the characteristic dust coagulation time or the time over which the local accretion rates changes). There are conditions, however, when even this simplification is invalid (e.g., during outbursts). See the papers by D'Alessio et al. (1998, 1999, 2001), Bell and Lin (1994) and Bell et al. (1997) and Cassen (2001) for further applications of vertical structure modeling.

3.2 Turbulence in Disks

The problem of disk evolution is that of determining the precise mechanisms of angular momentum transport. What are the processes that produce the torques and stresses that appear in (6) and how are they quantified? Instability leading to turbulence is the most frequently invoked phenomenon for providing the required viscosity, but the nature of the instability and the consequences of the turbulence remain controversial, despite assertions that the problem has been solved (e.g., Balbus and Hawley (1991, 2000). The problem begins with the realization that a Keplerian disk, although possessing a strong radial gradient of azimuthal velocity (shear), does not suffer the kind of shear-induced instability known to produce turbulence in other situations. An annulus of disk material displaced outward from radius r_1 to r_2, while conserving its angular momentum, experiences a centrifugal force j_1^2/r_2^3 at its new location. The centrifugal force required to maintain it in equilibrium is j_2^2/r_2^3. So, as long as $j_1 < j_2$, as it is in a Keplerian disk, the material experiences a net force which restores it toward its original location. The disk satisfies Rayleigh's criterion for stability,

$$\frac{\mathrm{d}j^2}{\mathrm{d}r} > 0$$

and some other source of turbulence must be sought.

Other instabilities exist, of course. For instance, it was proposed by Lin and Papaloizou (1980) that convective instability, which stirs a fluid when radiation alone would require a superadiabatic temperature gradient, could provide the required radial mixing of angular momentum. They envisioned a *feedback loop*, in which internally dissipated energy in an optically thick disk would drive convection; the turbulence so produced would give rise to $W_{\phi r}$ stresses; the stresses would result in the net outward transfer of angular momentum from the (more rapidly rotating) inner annuli to the (less rapidly) rotating outer annuli; the loss of angular momentum by the inner annuli would release gravitational energy, which would be the ultimate source of the dissipated energy driving convection. It is just such a feedback process that is required to overcome the inherent stability conferred by rotation.

But strong arguments have been presented against the existence of *any* such feedback process, in the absence of magnetic forces Balbus et al. (1996). The issue can be addressed by examining the equations of turbulent motion, derived from the mass and momentum equations by separating variables into an *average* part and a *fluctuating* part. For instance, for the velocity component v_i,

$$v_i = \langle v_i \rangle + u_i$$
$$\langle \nu_i \rangle = \frac{1}{2\pi \Delta r} \int v_i \, \mathrm{d}\phi \, \mathrm{d}r \, \mathrm{d}z,$$

where the r-integration is taken over some suitably small interval. In particular, one can derive a set of energy equations by multiplying each component of the momentum equation by its respective velocity component, and dropping terms smaller than second order in the fluctuating quantities (and other terms assumed to be small). The results for the r and ϕ directions, derived by Balbus et al. (1996), are:

$$\frac{\partial}{\partial t}\left\langle \frac{\varrho u_r^2}{2} \right\rangle = 2\Omega \left\langle \varrho u_r u_\phi \right\rangle - \left\langle u_r \frac{\partial p}{\partial r} \right\rangle - \left\langle \varrho \nu |\nabla u_r|^2 \right\rangle$$

$$\frac{\partial}{\partial t}\left\langle \frac{\varrho u_\phi^2}{2} \right\rangle = -\frac{\langle \varrho u_r u_\phi \rangle}{r} \frac{\mathrm{d}\left(r^2\Omega\right)}{\mathrm{d}r} - \left\langle \frac{u_\phi}{r} \frac{\partial p}{\partial \phi} \right\rangle - \left\langle \varrho \nu |\nabla u_\phi|^2 \right\rangle \ . \qquad (9)$$

These equations do not represent the energy conservation law, being derived from the momentum equations alone, but are relations that must hold between the mechanical energies associated with turbulent fluctuations and the mean flow, the latter being represented by Ω and its derivative. The last term on the right is the energy dissipated by viscosity; it always reduces the energy of the fluctuating part of the flow (the terms whose derivatives appear on the left). The first term on right prescribes the interaction of the turbulent stress, $\langle \varrho u_r u_\phi \rangle$, with the mean flow. The point stressed by Balbus et al. (1996) is that, in Keplerian disks (or any disk that is stable by Rayleigh's criterion) this interaction provides a *negative* feedback for the energy of ϕ-fluctuations if the stress is such as to transport angular momentum outward, i.e., $\langle \varrho u_r u_\phi \rangle > 0$. But it is necessary for turbulence to transport angular momentum outward in an accretion disk, because it is the loss of angular momentum that allows material to flow inward. Thus, turbulence that allows accretion appears to be self-defeating, no matter what the source of the turbulence. (True, the radial fluctuations are not damped, but *correlated* azimuthal fluctuations are necessary to produce any turbulent stress.) What about the ϕ-pressure gradient term? Balbus et al. (1996) argue that it cannot isotropize the turbulence enough to overcome the damping effect of the positive angular momentum gradient (while pointing out that long-range correlations in pressure fluctuations associated with *organized waves* can provide the desired effect, but would not be considered to be turbulence).

The conclusion that purely hydrodynamic turbulence cannot be self- sustaining in Keplerian disks is supported by numerical calculations of the evolution of a turbulent field in a "local" patch of a disk. In these calculations (e.g., Balbus et al. (1996) and Hawley et al. (1999), in which an appropriate form of periodic boundary conditions are imposed on a sheared, rotating fluid, instabilities are not manifested and an initially turbulent field decays. Furthermore, calculations in which convective turbulence is forced by imposition of an *ad hoc* heat source produce stresses which induce *inward* angular momentum transport, consistent with the arguments given above.

These results indicate that protostellar disk evolution must be driven by either magnetic instability or the action of waves, topics discussed below. And yet the issue should not be considered settled; not all researchers are ready to rule out the possibility of hydrodynamic turbulence in the unequivocal manner that its detractors have. For instance, calculations by Klahr and Bodenheimer (2000) indicate that *baroclinic* instabilities lead to sustained positive turbulent stresses and consequent outward angular momentum transport (see also Sheehan et al. (1999) and Li et al. (2001)). Baroclinic instability, well-known in planetary atmospheres, can occur when surfaces of constant pressure do not coincide with surfaces of constant density, i.e., when

$$\nabla p \times \nabla \varrho \neq 0 \ .$$

This condition is precluded in any calculation for which a barotropic relation between pressure and density, $p \sim \varrho^{\gamma}$, is assumed, for then

$$\nabla p \times \nabla \varrho \sim \nabla \varrho^{\gamma} \times \nabla \varrho = 0$$

and surfaces of constant p and ϱ do coincide. But it is virtually inevitable in an optically thick protostellar disk, where radial entropy gradients prevent the simple proportionality represented by the constant in the above equation. Local calculations, which assume a barotropic relation or do not account for a radial entropy gradient, and which exhibit decaying turbulence or negative turbulent stresses, do not allow the possibility of baroclinic instability. Now a common feature of the nonlinear development of baroclinic instabilities is the generation of relatively long-lived, organized structures, such as vortices, jet streams and spiral shock waves. Associated with these structures are nonlocal, correlated fluctuations which contribute to the transport of angular momentum. Related instabilities, which depend on the existence of a locally steep pressure gradient have similar properties (e.g., Lovelace et al. (1999) and Li et al. (2001). The full implications of these sources of turbulence (and possibly others; real disks are complicated structures) remain to be worked out (see the comments on Rossby waves, below).

In whatever way the issue of non-magnetic turbulence is resolved, Balbus and Hawley (1991, 1998) have shown that magnetically coupled protostellar disks are inevitably turbulent in a manner that produces outward angular momentum transport at the level required to drive the observed activity. Additional stress terms appear in (9) which can act as sources of azimuthal fluctuations. The instability that produces the turbulence relies on the fact that adjacent annuli are magnetically tethered, so they act like spring-coupled, rotating masses. If the spring constant is very strong, the instability is suppressed, but for a weak spring (magnetic field), perturbations in displacement grow. The reader is referred to the above references for a formal stability analysis. Here, we develop a useful analogy provided by Balbus and Hawley (1998) that gives some insight into the nature of the instability, and can be used to describe its essential properties.

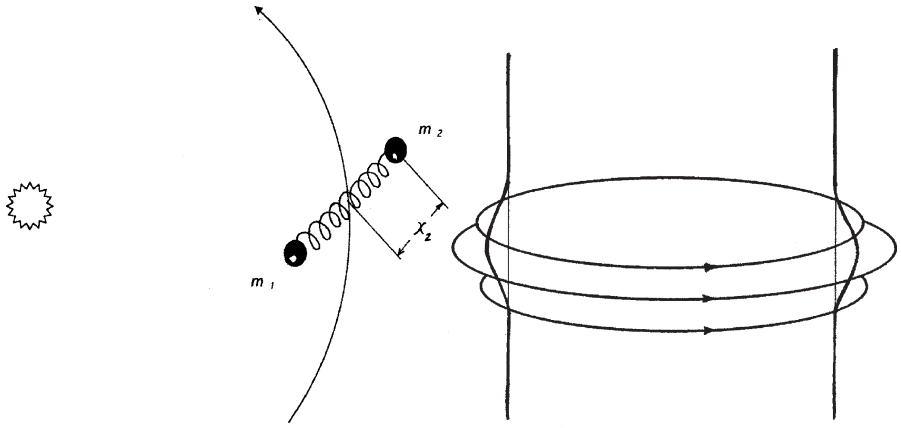

Fig. 14. Two masses, attached by a spring and displaced from a common orbit about a star (*left*), obey the same equations as orbiting, perfectly conducting fluid annuli threaded by a vertical magnetic fluid (*right*). Both systems are unstable for a range of values of the effective spring constant. This unstable range encompasses the only physically realistic values for the magnetic fluid

Consider two masses (identical for convenience) in orbit about a star, but attached by a spring with spring constant f_s (Fig. 14a) The equation of motion of either mass, in the rotating frame, is

$$\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}^2t} = \Omega \times (\Omega \times \mathbf{r}) + 2\Omega \times \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = -2f_s\mathbf{x} + \mathbf{g},$$

where $\mathbf{x}$ is the displacement from the equilibrium orbit. Let $\mathbf{r} = \mathbf{r}_0 + \mathbf{x}$ and expand $\mathbf{g}$ about its value at $\mathbf{r}_0$ to obtain the set

$$\begin{aligned} \ddot{x}_r - 2\Omega\dot{x}_\phi &= x_r\left(-2f_s + 3\Omega^2\right) \\ \ddot{x}_\phi + 2\Omega\dot{x}_r &= -2f_s x_\phi \ . \end{aligned}$$

The effective spring constant for the r motion is $2f_s - 3\Omega^2$, which must be positive for a restoring force. Now it turns out that these equations are exactly equivalent to those for the small displacement of annuli in a disk of perfectly conducting gas, threaded by a vertical magnetic field $\mathbf{B}$(Fig. 14b), with the correspondence

$$2f_s - 3\Omega^2 \rightarrow r\frac{\mathrm{d}\Omega^2}{\mathrm{d}r} + (\mathbf{k}\cdot\mathbf{u_a})^2 \ .$$

The last term is the scalar product of the wave vector of the disturbance and the Alfvén velocity, the latter defined by

$$\mathbf{u_a} \equiv \frac{\mathbf{B}}{\sqrt{4\pi\varrho}} \ .$$

Note that since $\mathrm{d}\Omega^2/\mathrm{d}r < 0$ in a Keplerian disk, stability requires

$$(\mathbf{k}\cdot\mathbf{u_a})^2 > -r\frac{\mathrm{d}\Omega^2}{\mathrm{d}r} \Rightarrow u_a^2 > \left|\frac{r}{k^2}\frac{\mathrm{d}\Omega^2}{\mathrm{d}r}\right| \ .$$

But k must be large enough to allow a wavelength to fit within the thickness of the disk: $k > \pi/h$. So the right hand inequality can be extended to

$$u_a^2 > \left|\frac{r}{k^2}\frac{\mathrm{d}\Omega^2}{\mathrm{d}r}\right| > \left|\frac{rh^2}{\pi^2}\frac{\mathrm{d}\Omega^2}{\mathrm{d}r}\right| = \frac{3\Omega^2h^2}{\pi^2} \approx \frac{3c_s^2}{\pi^2} \ .$$

But this condition for stability requires that the magnetic field pressure dominate the thermal pressure, a condition difficult to achieve in disks because magnetic buoyancy effects tend to expel such strong fields. Therefore, magnetic disks are unstable; in fact, they are unstable even in the limit of vanishingly weak field. Numerical simulations confirm the instability and indicate that the resulting turbulence would be sufficient to provide the inferred accretion rates of protostellar disks Hawley et al. (1995).

Can we conclude that magnetic turbulence is the main process responsible for protostellar disk evolution? For this to be the case, the disk gas must be well-coupled to the magnetic field. This condition is quantified by the requirements that the magnetic Reynolds number be larger than unity and that collisional ion-neutral momentum exchange occur rapidly compared to, say, the orbital period. Disks are dense enough in most places to insure that the latter condition is fulfilled, but the former condition requires a level of ionization which, although not very high (ionization fraction $\approx 10^{-13}$), is still difficult to attain in their cold, dusty interiors. One can identify four sources of ionization: (1) galactic cosmic rays, (2) stellar energetic particles and x-rays, (3) radioactive nuclides and (4) thermal (collisional) excitation. Galactic cosmic rays penetrate no more than about $10^2\,\mathrm{gm/cm^2}$ of material, and so would be largely excluded from the inner disks ($r \leq 1\,\mathrm{AU}$) where the surface density is estimated to typically exceed $10^3\,\mathrm{gm/cm^2}$. (The current galactic cosmic ray flux is obviously incapable of significant ionization of most of the approximately $10^3\,\mathrm{gm/cm^2}$ of terrestrial atmosphere.) Furthermore, one expects that the intense stellar wind associated with young stars would attenuate the flux of such particles to levels well below that currently experienced by the solar system. Stellar particles and x-rays, although abundant, penetrate only about $1\,\mathrm{gm/cm^2}$,and so are expected to ionize only a very thin "ionodisk" at high altitudes. Similarly, the most abundant energetic particles from radioactive nuclides have very limited ranges. Only close to the star, perhaps within a few tenths of an AU, is the disk temperature expected to be high enough to evaporate most of the dust, a condition that is probably required to maintain the ionization level necessary for magnetically induced turbulence. Thus the issue of turbulent angular momentum transport throughout the disk remains open. For a recent discussion of relevant issues by the advocates of *exclusively* magnetic turbulence, see Balbus and Hawley (2000).

3.3 Waves in Disks

It was mentioned above that organized, non-axisymmetric structures transport angular momentum. These structures frequently have the form of waves, which may be thought of as coherent perturbations of the flow through which the fluid medium flows. An astounding variety of fluid waves have been identified, many of them familiar in our everyday experience; one should expect to find many of them in disks. Waves usually result from the interaction of some disturbing force (or instability) and the natural restoring forces present in the system, necessary for the existence of an equilibrium. The restoring forces can be associated with the natural frequencies of a system. Some of the important natural frequencies of a disk are its orbital frequency Ω, the epicyclic frequency (associated with the Coriolis force and characteristic of radial oscillations), and the Brünt-Väisälä frequency (associated with pressure forces and characteristic of vertical oscillations). In a thin disk, these frequencies are all of comparable magnitude, so one might expect that there are complicated interactions among various wave types.

The nature of a wave is quantitatively described by a dispersion relation, which is a relation between the frequency of oscillation of the wave, ω, and the wave number, $k = 2\pi/\lambda$ (λ is the wavelength): $\omega = \omega\,(k; \text{parameters})$. From the dispersion relation, one can determine the rate at which energy is carried by the wave ($\partial\omega/\partial k$), the speed at which the wave pattern moves (ω/κ), places where the waves can and cannot propagate, and so forth. The usual method of deriving a dispersion relation involves three steps: (1) all variables are expressed as the sum of their equilibrium values and small perturbations which oscillate in time and space; (2) the fluid equations are linearized by retaining only terms first-order in the perturbations and using the fact that the equilibrium values satisfy the steady-state equations exactly; and (3) spatial derivatives in the equilibrium values are considered negligible compared to the spatial oscillations of the perturbations (WKB approximation). These steps result in a set of linear, *algebraic* equations (algebraic because the spatial and temporal derivatives of oscillating quantities are proportional to the quantities themselves). If there are no explicit forcing functions, the equations are homogeneous and have solutions (the "free wave" solutions) only if there is a specific relation between ω (specifying the temporal oscillation) and k (specifying the spatial oscillation); this is the dispersion relation.

In Keplerian disks, the strong variation in rotation rate with radius tends to shear disturbances, so that waves commonly have a spiral pattern. The equation of a spiral is $\phi = \psi\,(r)$; the equation of a spiral rotating at frequency ω is $\phi = \psi\,(r) + \omega t$; and many such spirals, with the same shape and frequency, are described by $m\phi = \psi\,(r) + \omega t$, where m is an integer. Therefore, in analyzing disk waves, it is expeditious to express the oscillating perturbations in the form:

$$x = X\mathrm{e}^{\,\mathrm{i}[\omega t - m\phi + \psi(r)]}\ .$$

where X is the amplitude of the perturbation. The radial wavenumber is given by $k = \mathrm{d}\psi/\mathrm{d}r$ and the pattern speed is $\Omega_p = \partial\phi/\partial t = \omega/m$. A positive wavenumber corresponds to a *leading* wave (phase increasing with r) and a negative wave number corresponds to a *trailing* wave. It turns out that only trailing waves carry angular momentum in ways that are physically sustainable.

Waves in disks have been studied using many different approximations, assumptions and techniques, with the consequence that it is not always easy to relate the results of one study to those of another. There are approximations based on disk geometry: 1-D (axisymmetric), 2-D $(r, \phi\ )$, and "shearing sheet" (or local Cartesian). In the last approximation, the $(r, \phi\ )$ coordinate system is replaced by a local, rotating (x, y) system, and terms associated with the curvature of streamlines are neglected. The "tightly wound" (short radial wavelength) assumption exploits the spiral geometry of the wave and is usually equivalent to the WKBJ approximation. Physical forces that are frequently (but not always) neglected are viscous, magnetic and disk self-gravitational. The following approximations of the equation of state are often encountered: isothermal, polytropic and Boussinesq (in which density variations due to dynamic pressure are neglected, but not those in the base state). In particular, it is often assumed for mathematical convenience that the surface density (rather than the volume density) is a power of the vertically integrated pressure, a kind of polytropic approximation unjustifiable by any simple physical assumption. Finally, analyses can be Lagrangian, in which the fluid displacement is used as a dependent variable, or Eulerian, in which fluid velocities are the primary dependent variables. I will identify the specific assumptions used in the analyses described below.

Spiral density waves can be considered to be the primary wave-form in astrophysical disks. They were originally studied in connection with galactic structure (Lin and Shu 1964), but observational confirmation of the theory was obtained in the structure of Saturn's rings. A standard analysis (Shu 1992) considers the disk to be infinitely thin, so that perturbations are restricted to the (r, ϕ) plane and the volume density $\varrho = \Sigma\delta\,(z)$, where δ is the Dirac function. It is assumed that the vertically integrated pressure is a function of only the surface density: $P = P\,(\Sigma)$. Viscous stresses and dissipation are ignored, but the self-gravity of the disk is included. For this purpose, the standard conservation equations of Sect. 3.1 are supplemented by Poisson's equation for the gravitational potential:

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \Phi}{\partial z}\right) + \frac{1}{r^2}\frac{\partial^2 \Phi}{\partial r^2} + \frac{\partial^2 \Phi}{\partial z^2} = 4\pi G\delta\,(z) + \varrho_* \ .$$

Here, ϱ_* represents the contribution to the stellar potential. Shu (1992) shows how this equation can be simplified and solved in the WKB approximation to yield a simple relation between the amplitudes of the surface density and

gravitational potential perturbations. The following dispersion relation can then be derived from the linearized conservation equations:

$$(\omega - m\Omega)^2 = \Omega^2 + k^2 c_s^2 - 2\pi G\,|k|\,\Sigma_0 \ . \tag{10}$$

Subscript 0 refers to the unperturbed state. (Here and in the following it is assumed that the orbital and epicyclic frequencies are the same, as they nearly are whenever the disk mass is small compared to the stellar mass. More general relations can be derived, in which κ appears instead of Ω.) The derivation of (10) involves some subtle issues of ordering. It must be recognized that the perturbation to the surface density is intrinsically larger than the velocity, displacement and potential perturbations. This effect is illustrated geometrically in Fig. 15, where a density wave has been constructed by representing the perturbed streamlines of the disk by a set of nested ellipses, each ellipse being rotated slightly with respect to its neighbor. Note how small displacement perturbations (eccentricities) produce large surface density perturbations (tightly bunched streamlines).

The dispersion relation yields an axisymmetric ($m = 0$) stability criterion. Recognize that $\omega^2 > 0$ (i.e., ω real) is required for oscillating (not

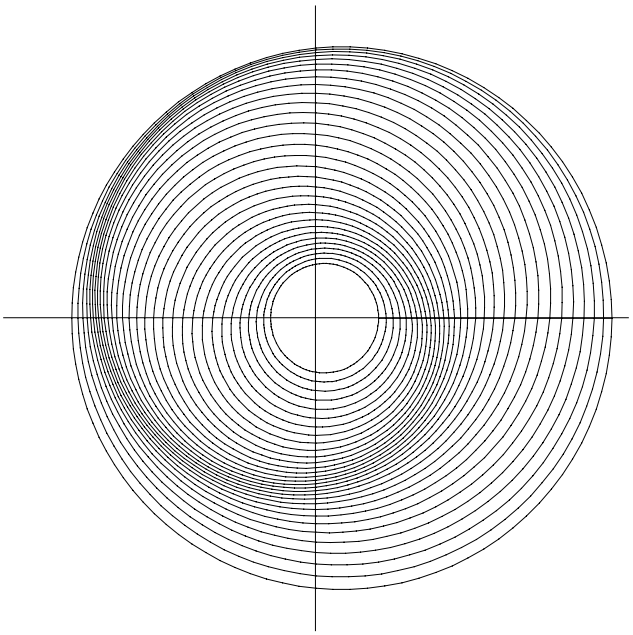

Fig. 15. Motions in a spiral density wave can be represented by elliptical streamlines with radially varying phases. Even a small displacement of the streamlines from circular can produce a large surface density perturbation, as seen in the tightly bunched streamlines

exponentially growing) perturbations, and find from (10) (with $m = 0$) that the condition

$$Q \equiv \frac{c_s \Omega}{\pi G \Sigma} > 1 \tag{11}$$

is required for stability; i.e., $\omega^2 > 0$ positive for all values of k. This important inequality is commonly known as Toomre's stability criterion (Toomre 1964), although Toomre derived a slightly different relation for a galactic disk of stars, and it has been derived by in other forms by a number of people (perhaps first by Safronov (1960)). (The analysis of non-axisymmetric perturbations is more complicated, but it is found that values of Q only somewhat greater than unity are required for stability.) It is seen that the effects of pressure and rotation are stabilizing, while the effect of gravity is destabilizing. Estimates of the temperatures and surface densities obtained by the methods described in Sect. 1.2 usually indicate that protostellar disks are stable by this criterion. It is quite possible, however, that a disk could become unstable at some time during its evolution. This might happen during the formation of the disk, if mass builds up faster than it is accreted by the star, or at later times if the outer part of the disk (for instance) becomes sufficiently cool. The question of what happens to an unstable disk is of great importance to the issue of giant planet formation, and will discussed in Sect. 1.6.2.

The dispersion relation also reveals other important properties of the waves. These are conveniently described in terms of the nondimensional wavenumber and frequency defined by

$$k' = \frac{k 2\pi G \Sigma_0}{\Omega^2}$$
$$\omega' = \frac{\omega - m\Omega}{\Omega} \ .$$

The radius at which $\omega' = 0$ is the *corotation* radius , where the pattern speed $\Omega_p = \omega/m$ matches the orbital frequency (which, for our Keplerian protostellar disks is the natural frequency for radial oscillations). The quantity ω' thus measures the "distance" from corotation in frequency space. The locations at which $\omega' = \pm 1$ also have special significance. This condition defines the *Lindblad resonances*, where the orbital frequency is exactly an integer times the difference in the orbital and wave pattern frequencies:

$$\Omega = \pm m \left(\Omega_p - \Omega \right) \ .$$

The resonant radii are found from the Keplerian rotation law:

$$r_L = \left(\frac{m \pm 1}{m} \right)^{1/3} r_{\text{corotation}} \ \ (m = 1, 2, 3, ...) \ . \tag{12}$$

The plus and minus signs correspond to *outer* and *inner* Lindblad resonances, respectively. At these locations, the disturbance caused by the wave is in

phase with the local natural frequency of oscillation, and one expects that disturbances would interact strongly with the disk material there.

The dispersion relation, solved for k', then becomes

$$|k'| = \frac{2}{Q^2}\left[1 \pm \sqrt{1 - Q^2\left(1 - \omega'^2\right)}\right] \ . \tag{13}$$

The requirement that the radical be real excludes wave propagation from within a certain distance of corotation:

$$\omega'^2 > 1 - \frac{1}{Q^2} \ .$$

Furthermore, the fact that the right-hand side of (13) must be positive indicates that waves associated with the minus sign ("long" waves) cannot propagate where $\omega'^2 > 1$. For very stable disks ($Q >> 1$), the range of long wave propagation is very limited. These waves are governed primarily by rotational and gravitational forces. The propagation of "short" waves (for the plus sign), which are like acoustic waves, is not so restricted. The situation is illustrated in Fig. 16. Spiral density waves can be excited by instability if $Q < 1$ somewhere in the disk, or by an object embedded in, or external to, the disk, such as a planet or stellar companion. In fact, they can be excited by any number of disturbances. Their importance for disk evolution lies in their ability to transport angular momentum. One can get some idea about how this works by examining the fate of energy, E_w, and angular momentum, J_w, carried by

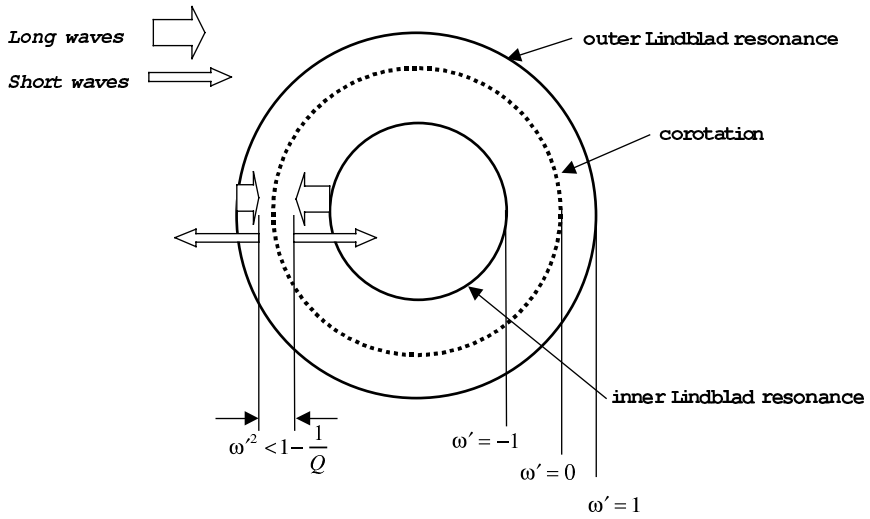

Fig. 16. Long (gravity-like) waves of a given spiral mode (m) are restricted to propagate in a limited range between the inner and outer Lindblad resonances, the boundaries of which exclude the corotation radius. Short (acoustic-like) waves can propagate within the inner Lindblad resonance and beyond the outer Lindblad resonance, but are also excluded by the same boundaries near corotation

a wave. In the absence of dissipation, wave energy and angular momentum are conserved, in the sense that they obey conservation equations of the form:

$$\frac{\partial\left(E_w, J_w\right)}{\partial t}+\frac{1}{r}\frac{\partial\left[rc_{gr}\left(E_w, J_w\right)\right]}{\partial r}=0,$$

where c_{gr} is the group velocity, and the quantities (E_w, J_w) represent the azimuthally averaged energy and angular momentum, per unit surface area, associated with the wave. They are, of course, functions of the wave parameters ω, k, m and the radius r. Their specific functional dependencies need not concern us now (see Shu (1992) for formulae), except to note that quite generally, $J_w = E_w/\Omega_p$. In the presence of dissipation, waves exchange energy with the disk material, and the wave conservation equations take the form

$$\frac{\partial E_w}{\partial t}+\frac{1}{r}\frac{\partial\left[rc_{gr}E_w\right]}{\partial r}=-D-\frac{\partial E_d}{\partial t}+\frac{1}{r}\frac{\partial\left[rv_rE_d\right]}{\partial r}$$
$$\frac{\partial J_w}{\partial t}+\frac{1}{r}\frac{\partial\left[rc_{gr}J_w\right]}{\partial r}=-\frac{\partial J_d}{\partial t}+\frac{1}{r}\frac{\partial\left[rv_rJ_d\right]}{\partial r},$$

where D is the appropriately averaged dissipation rate and

$$E_d=-\frac{GM_*\Sigma}{2r}$$
$$J_d=\Sigma\sqrt{GM_*r}\ .$$

From these equations, with the aid of the azimuthally averaged mass conservation equation, and assuming that the disk quantities do not change in the time of wave propagation, one can derive an explicit formula for the radial mass flux induced by angular momentum deposition from the attenuated waves:

$$2\pi\Sigma rv_r=\frac{4\pi D}{\Omega\left(\Omega_p-\Omega\right)}\ .$$

This relation demonstrates that the mass flux is proportional to the dissipation, and that it is inward inside corotation and outward outside of corotation. The relation does not hold at corotation because it was derived from the free wave dispersion relation, which must be violated at the corotation resonance. We return to the connection between waves and resonances in the discussion of planet-disk interactions (Sect. 1.7.1).

What about the *vertical* structure of these waves? The 2-D analysis just described cannot address this question, of course. In fact, a complete 3-D analysis has yet to be accomplished. Interesting results have been obtained, however, by means judicious approximations (Lubow and Pringle (1993); Korycansky and Pringle (1995); Lubow and Ogilvie (1998); Ogilvie (1998)). The strategy employed is the following. The conservation equations, in the shearing sheet approximation, including z and r dependencies, are linearized in the usual way, but only axisymmetric waves are considered. The WKB approximation is used for r, but not z dependencies, so r takes the role of a parameter.

This procedure reduces the problem to a set of 1-D (in z) differential equations (eigenvalue problems), with coefficients dependent on r. These equations are solved for the z structure of the waves. The radial propagation properties are diagnosed by solving the equations at different r and constraining the solutions to obey conservation of wave energy. It is then argued that the results are applicable to non-axisymmetric waves, with the replacement of ω by $\omega - m\Omega$, as long as the waves are tightly wound in the sense that $m << kr$. Self-gravity is not included, so only the counterparts of the short (acoustic) modes described above are present. Finally, it is assumed that the fluid obeys a polytropic equation of state: $p \propto \varrho^{\gamma}$, as might be the case for an optically thick disk. This point is important for the following reason: the vertical extent of a locally polytropic disk is finite. That is, there is an altitude at which temperature, density and pressure fall to zero, which defines the thickness of the disk. To see this, solve the vertical hydrostatic equation (5) to find that the state variables are proportional to a power of the quantity $\left(1 - z^2/H^2\right)$, where

$$H^2 = 2\gamma p_0 / \left(\gamma - 1\right) \Omega^2 \varrho_0$$

and subscript 0 refers to values at $z = 0$. This kind of vertical structure has the effect of a waveguide, and strongly affects the nature of waves and their propagation properties, as described below.

Resolving the vertical disk structure reveals a multiplicity of modes, each with its own characteristic vertical oscillations, some of which are associated with buoyancy restoring forces. Some modes have counterparts in stellar oscillations, and are therefore designated accordingly. The p-waves are compressible, with the main restoring force provided by pressure. The g-waves are essentially gravity waves, and are incompressible, with buoyancy providing the restoring force. Waves of the third mode, designated r-waves, have no counterpart in (non-rotating) stars; they are inertial waves, with angular momentum and buoyancy providing the restoring forces. Each of these modes possess z-symmetric and antisymmetric components, and each is represented by a series of waves with n nodes in the z direction. In addition, there are two *fundamental* (designated f) modes (z- symmetric and antisymmetric), which turn out to be particularly important because they carry almost all of the angular momentum contained in waves stimulated at resonances. They are the dominant 3-D counterparts of the short, 2-D spiral density waves.

Like their 2-D counterparts, the radial propagation of these waves is constrained, as shown in Fig. 17. Note that vertical resonances (where radial oscillations are in phase with vertical oscillations) now play a role in restricting the radial propagation of the p-waves. The waves also become vertically constrained, in some cases severely so, as they propagate away from resonances. The f, p and g modes become confined to within a wavelength of the surfaces of the disk; the r modes become restricted to a distance of about $(\lambda H)^{1/2}$ from the midplane. An example of the confinement of the f mode is shown in Fig. 18. This effect is due to the waveguide character of disks with finite

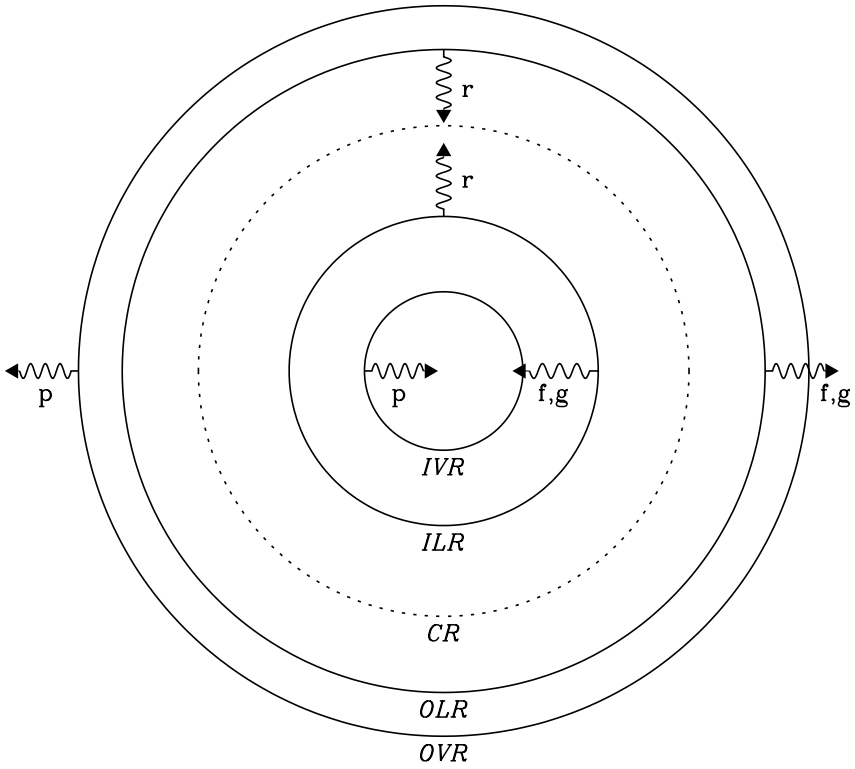

Fig. 17. When the vertical structures of spiral density waves are resolved, variety of additional modes are discovered. These also have restricted ranges of propagation, some of which are bounded by the locations of vertical resonances, as well as the Lindblad resonances. Shown are the boundaries of allowed propagation for the $r-$, $p-$, $g-$ and $f-$ modes discussed by Lubow and Ogilvie (1998). I and O refer to inner and outer, LR and VR are Lindblad and vertical resonances, respectively, and CR is corotation. (Figure from Lubow and Ogilvie 1998)

vertical extent. In fact, it disappears in disks that are vertically isothermal, for such disks formally extend to infinite z, as demonstrated by the appropriate solution of (5). So what about a more realistic disk, which might resemble the polytropic model except near the surfaces, where an optically thin, isothermal atmosphere should exist? This situation was analyzed by Ogilvie and Lubow (1999), who found that wave confinement still occurred, but was less severe.

Now recall that the angular momentum carried by a wave is deposited in the disk according to how the wave energy is dissipated. The fact that the energy of spiral density waves tends to be concentrated toward the disk surfaces could have a profound effect on their non-linear behavior and the ultimate manner in which they are dissipated. They might form shocks or "break" in other ways. The problem deserves further attention, probably most fruitfully by the application of high resolution numerical simulations.

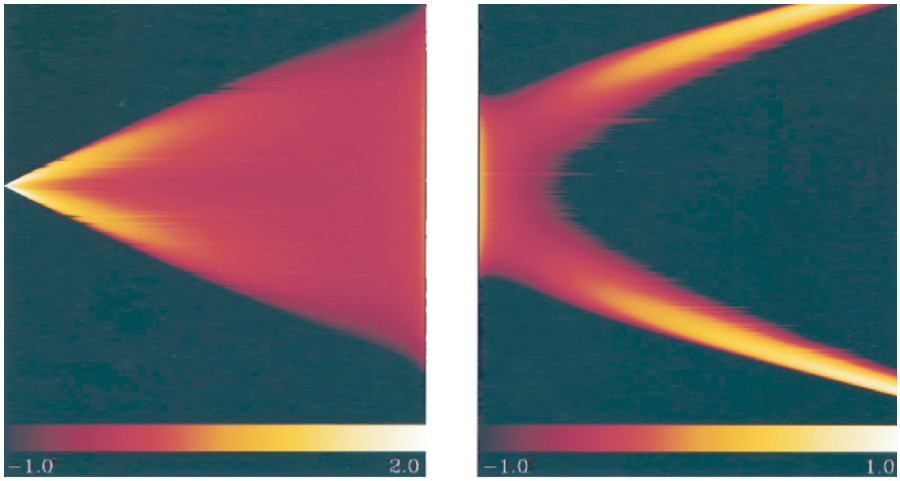

Fig. 18. A representation of the distribution in disk cross-section of the wave energy associated with the fundamental, $m = 2$ mode, launched at an inner (**a**) and outer (**b**) Lindblad resonance. The scale refers the logarithm to the base 10 of a dimensionless energy. The plot illustrates the channeling effect imposed by vertical structure. (Figure from Lubow and Ogilvie 1998)

Before leaving the subject, something should be said about another kind of wave, the Rossby wave. Rossby waves have the property of propagating *vorticity*, which is a measure of the local shear in a flow, defined as the curl of the velocity. These waves are well-studied in the context of planetary atmospheres and have been shown to be potentially important in protostellar disks. Indeed, the baroclinic instability (and related instabilities) referred to in Sect. 3.2 is known to stimulate such waves, which may then couple with other modes in complicated ways. The Rossby wave dispersion relation, as derived by Sheehan et al. (1999), is

$$(\omega - m\Omega)^2 = \Omega^2 + k^2 c_s^2 + m^2 c_s^2 / r^2 \ . \tag{14}$$

This looks very much like (10), except that the gravitational term has disappeared (because self-gravity was not included in the analysis), replaced by the last term on the right. This term was neglected in the spiral density dispersion relation because it was smaller than other terms, for short wavelengths ($kr >> 1$) in thin disks. But Rossby waves have long wavelengths and are revealed only when this term is retained. For this reason, radial gradients of the base state should not be suppressed, i.e., the WKBJ approximation is not valid. [In other respects, the assumptions leading to (14) are the same as those for (10).]

In planetary atmospheres, Rossby waves are known to be responsible for the extraction of energy from external sources (such as solar radiation or internally generated heat) and its transfer to organized motions. For instance,

large storms and jet streams in the terrestrial atmosphere are manifestations of Rossby waves, as are the vortices and high speed belts on Jupiter. The role of these waves in disk angular momentum transport has yet to be determined. Their effectiveness will depend on the existence of appropriate instabilities and the establishment of a feedback loop that permits energy extraction from the disk itself, as discussed above.

This ends our discussion of disks as astrophysical objects independent of their potential for forming planets. The remaining sections focus on the theory of how they produce planetary systems.

4 Dust-Gas Dynamics

4.1 Drift and Settling Velocities in the Absence of Turbulence

In a cool protostellar disk of the same composition as our Sun, about 0.4 % of the mass is in the form of rock-forming solids (mainly iron and magnesium silicates). If it is cold enough (less than about 160 K), another 1.5 % exists as H_2O ice. Initially, these solids are in the form of sub-micron particles, or "dust", inherited from the interstellar medium, or newly condensed, if and where the disk was once hot. Small dust particles are well-coupled to the gas, and so follow the overall gas motions closely. While dispersed, their main effect is as a source of opacity; they have no direct effect on the gas dynamics, because they comprise such a small mass fraction. Occasionally they collide with each other, stick together (see **?**, and references therein), and gradually accumulate, reducing the opacity and altering the nature of their dynamical interaction with the gas. So begins, it is believed, the process of rocky planet formation.

The dynamical interaction of gas and solids is described by the equations of motion for the solid particles, which may be written (e.g. Dubrulle et al. 1995)

$$\begin{aligned} \frac{\mathrm{d}v_{rp}}{\mathrm{d}t} &= g_r + \frac{j_p^2}{r^3} - \frac{v_{rp} - v_r}{t_r} \\ \frac{\mathrm{d}j_p}{\mathrm{d}t} &= -\frac{j_p - j}{t_f} \\ \frac{\mathrm{d}v_{zp}}{\mathrm{d}t} &= g_z - \frac{v_{zp} - v_z}{t_f} \ . \end{aligned} \tag{15}$$

Subscript p refers to the particles. The last term on the right hand side of each equation is the frictional force between the gas and the particles. The "friction time" t_f is the characteristic time in which a particle of given size and mass exchanges momentum with the gas (sometimes referred to as the "stopping time"). The way it is calculated depends on whether the mean free path in the gas is larger or smaller than the particle and whether the initial

relative velocity is super- or sub-sonic. The mean free path for gas at 1 AU in a typical protostellar disk is a few tens of centimeters. For smaller particles moving subsonically through such a gas,

$$t_f = \frac{\varrho_p r_p}{\varrho c_s},$$

where ϱ and r_p are the particle's density and radius, respectively. This formula can be derived by equating the frictional force to the rate at which thermal gas molecules transfer momentum to a solid, spherical particle and solving the resulting momentum equation. (Of course, the particles are likely to be irregularly shaped rather than round, in which case r_p must be interpreted to be an appropriate dimension. For instance, it might be the characteristic size of the individual particles making up a fractal structure.) For particles larger than the mean free path, the friction time depends on the relative velocity itself:

$$t_f = \frac{8\varrho_p}{3\varrho C_d \left(v_p - v\right)} \ .$$

Here, C_d is the drag coefficient, which also depends on the relative velocity as well as the particle size and shape, and gas viscosity; however, its value is generally of order unity. In most of the following, we will be concerned with particles smaller than the mean free path.

The first important point to be learned from these equations is the following: solid particles experience an azimuthal drag due to the fact that the gas motion is slightly sub-Keplerian, and this drag can cause a substantial radial drift as particles lose angular momentum. In Sect. 3.1, we used the gas radial momentum equation to derive the fact that, to order, $(h/r)^2$, $\Omega = \Omega_K$. But if the pressure gradient term is retained in that equation, one finds for the gas

$$(r\Omega)^2 = (r\Omega_K)^2 + \frac{r}{\varrho}\frac{\partial p}{\partial r}$$

or

$$j^2 = j_K^2 + \frac{r^3}{\varrho}\frac{\partial p}{\partial r} \ .$$

With this expression, the solid particle radial momentum equation (the first of 15) yields

$$v_{rp} - v_r = \frac{t_f}{\varrho}\frac{\partial p}{\partial r} \ .$$

Since the pressure gradient is usually negative, the effect produces an inward drift. Weidenschilling (1977) calculated the drift rate in a model disk, for a variety particle sizes and gas densities, for a disk in which $v_r = 0$. His results are shown in Fig. 19, in which it is seen that very small particles are well-coupled to the gas and have small drift rates, and large solid objects

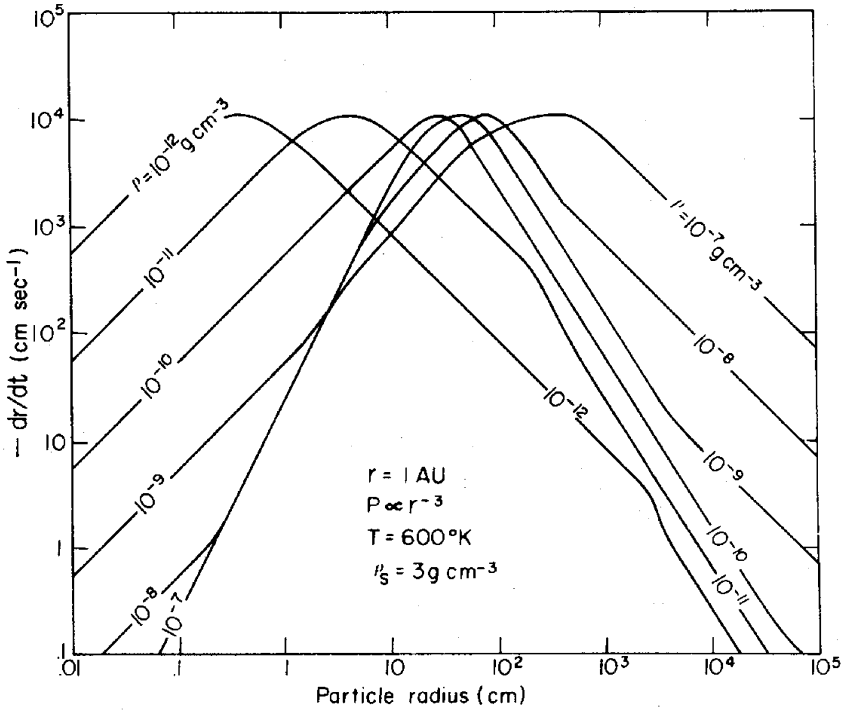

Fig. 19. Particle drift rates due to loss of angular momentum by gas drag, as a function of particle size and gas density, for typical disk conditions. Small particles have small drift rates because they are easily carried along by the gas; large objects have small drift rates because drag forces are small compared to inertial forces. Intermediate-sized particles, in this case about a meter in size, drift rapidly. (Figure from Weidenschilling 1977)

are decoupled from the gas and also have small drift rates. Intermediate-sized particles, however, can drift rapidly. The peak drift rates in Fig. 19, 10^4 cm/sec, would deliver a particle from 1 AU to the Sun in only 100 years. The survival of planetary-sized objects therefore implies that growth through the critical size range (about 1 meter, for a typical gas density) was rapid, or that there were sustained, systematic outward gas velocities which were larger than the drift rate, or that collective effects (for which formula derived for individual, isolated particles do not apply) protected the particles from drift loss.

The solid particle equations also quantify the rate at which vertical settling and concentration at the midplane occurs. For the gas velocity $v_z = 0$, and writing $v_{zp} = \mathrm{d}z/\mathrm{d}t$ and $\mathrm{d}v_{zp}/\mathrm{d}t = \mathrm{d}^2z/\mathrm{d}t^2$, one finds

$$\frac{\mathrm{d}^2 z}{\mathrm{d}t^2} + \frac{1}{t_f}\frac{\mathrm{d}z}{\mathrm{d}t} + z\Omega^2 = 0 \ .$$

This is the equation of a damped harmonic oscillator, which is critically damped for $t_f = 1/2\Omega$. Particles larger than $r_p = \Sigma/4\varrho_p$ follow orbits which oscillate through the midplane; smaller particles gradually spiral toward the midplane. If $t_f << 1/\Omega$, the terminal velocity $v_{zp} = -z\Omega^2 t_f$ is quickly attained. Suppose $\Omega = 2 \times 10^{-7}\,\mathrm{sec}^{-1}$, the orbital frequency of the Earth, and $\varrho_0 = 10^{-9}\,\mathrm{gm/cm}^3$, a typical value at 1 AU. Then, if we measure the quantity $\varrho_p r_p$ in cgs units, we find $t_f \approx 5 \times 10^3\,(\varrho_p r_p)$ sec, $v_{zp} \approx 200\,(\varrho_p r_p)$ cm/sec, and the settling time $= h/v_{zp} \approx c_s/v_{zp}\Omega \approx 10^3/\Omega\varrho_p r_p\,\mathrm{sec}$. The density ϱ_p is of order unity, so one sees immediately that micron-sized particles would take millions of years to settle, and the least breeze in the vertical direction would inhibit any concentration by settling. They must grow first. On the other hand, a golf-ball sized rock would settle, in the absence of vertical gas velocities, in 10^3–10^4 years.

4.2 Particle Growth and Trajectories in the Absence of Turbulence

Differential velocities among particles of different sizes cause them to collide, stick and grow. The rate of growth of a particle and the path that it follows as it grows can be determined by integration of the set (15), augmented by an equation for the growth rate . The latter is given by

$$\frac{\mathrm{d}m_p}{\mathrm{d}s} = \pi r_p^2 \varrho_s,$$

where m_p is the particle mass (assumed here to be $4\pi r_p^3 \varrho_p/3$), s is the path length along the trajectory, and ϱ_s is the volume density of solids encountered along the trajectory. This expression is valid if all solids encountered by the growing particle stick to it, as is expected when relative velocities between particles are less than about 1 m/sec (Dominik and Tielens 1997; **?**). If this is not the case, the right-hand side must be appropriately reduced by some efficiency factor. If relative velocities are too high, fragmentation occurs, rather than growth; see the discussion in Wurm et al. (2001). Note also that the density of solids, ϱ_s, is an evolving quantity whose value depends on how all of the particles are settling. Nakagawa et al. (1986) calculated growth and trajectories by assuming that at early times ϱ_s was well represented by its initial value, a constant fraction of the gas density. At later times, as settling proceeded, it was approximated by Σ_s/z, where Σ_s is the initial column density of solids. They also adopted typical forms for the distribution of gas density, and assumed that there were no gas velocities. Their results are shown schematically in Fig. 20, for a particle starting at altitude Z_0. There is little radial drift until the particle has settled and grown. At Z_1, it is large enough to drift radially, but rapidly accumulates smaller (background) particles, and begins to spiral vertically again at Z_3. But by this time, it has entered a region (at Z_2) in which settling has caused the dust density to be comparable to the gas density. Further details are given in Fig. 21. Most of the settling time is spent at high altitudes, while the particle grows. The total

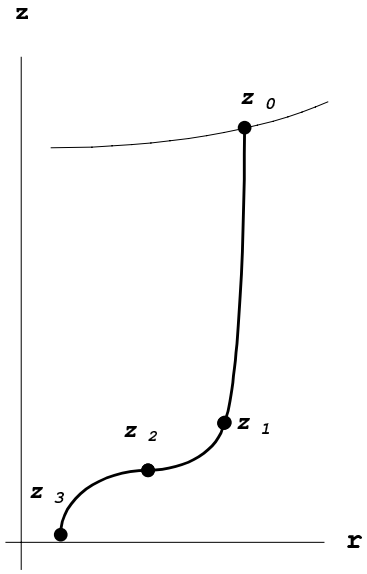

Fig. 20. Schematic representation of the trajectory of a particle settling and growing in a non-turbulent disk. The initial position of the particle is at altitude Z_0. It settles mainly vertically until Z_1, where it has become large enough to drift radially. It then rapidly accumulates smaller particles, whereupon, at Z_2, it enters a region where the dust density is comparable to the gas density. After Z_3 it spirals vertically toward the midplane. Small amounts of turbulence, however, can prevent the formation of the dense dust layer; see Fig. 23 and the discussion in Sect. 4.3. (Figure adapted from Nakagawa et al. 1986)

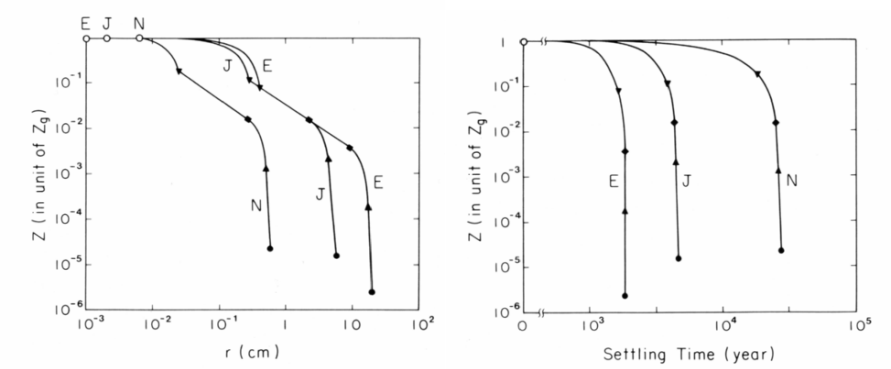

Fig. 21. Quantitative results for settling and growth in a non-turbulent disk: altitude vs. particle size (a), and altitude vs. time. The quantity z_g is the disk scale height. The curves labeled E, J and N refer to locations in a disk at the orbits of Earth, Jupiter and Neptune, respectively. The solid symbols correspond to the points z_1, z_2 and z_3 of Fig. 21, and the endpoint of the calculation. (Figure from Nakagawa et al. 1986)

radial drift experienced by these particles is limited to a small fraction of an AU, because rapid radial drift is associated with rapid growth through the size of maximum drift.

4.3 The Effect of Turbulence on Particle Settling

Although the calculation just described is useful for understanding the fundamental processes and timescales, even a small degree of turbulence could have a substantial impact on some of the quantitative details. The rather small values of the settling velocity estimated above motivate the following question: What values of the turbulence parameter α would prevent (or inhibit) settling? Suppose we calculate

$$\left|\frac{v_{zp}}{v_{turb}}\right| = \frac{z\Omega^2 t_f}{\alpha c_s h/l_{turb}} = \frac{2}{\alpha\Sigma}\frac{zl_{turb}}{h^2}\varrho_p r_p \ .$$

Here, l_{turb} is the turbulent mixing length. Realizing that $zl_{turb}/h^2 < 1$, one concludes that

$$\left|\frac{v_{zp}}{v_{turb}}\right| < \frac{2}{\alpha\Sigma}\varrho_p r_p = r_p\frac{2\times 10^{-3}}{\alpha}$$

for $\varrho_p = 1\,\mathrm{g/cm^3}$ and a typical value (at 1 AU) of $\Sigma = 10^3\,\mathrm{gm/cm^3}$. Turbulent velocities would therefore exceed the settling velocities of small grains ($r_p << 1\,\mathrm{cm}$) even for values of α well below that required to evolve a disk in a million years (10^{-3}–10^{-2}). Therefore, even if turbulence is not the primary agent for transporting angular momentum in disks, it could still be important for controlling the initial distribution and evolution of the solid component.

It is often useful to characterize the turbulent field by an ensemble of transient *eddies* (or vortices), with some wavelength distribution of energy, usually expressed as a power law in the wavenumber k:

$$E_k = \frac{v_k^2}{k_0}\left(\frac{k}{k_0}\right)^{-a} \ .$$

The characteristic eddy velocity and size are v_k and $2\pi/k$, respectively. In random turbulence, eddies typically last only a single turnover time, so their lifetime t_k is given by $1/kv_k$. The motion of a solid particle in a turbulent field is then a diffusive process, as its velocity is altered by random encounters with turbulent eddies. The velocity perturbations depend on the particle size and the gas properties, as characterized by t_f. For instance, if $t_f > t_k$, the eddy disappears before the particle is entrained in it. If the particle has a systematic velocity v_p, and if the "eddy crossing time" $t_{cross} \equiv 1/kv_p$ is greater than t_k, the particle can be entrained, but will also encounter smaller eddies within the eddy of size $2\pi/k$. If $t_{cross} < t_f$, the particle passes through the eddy without being entrained. Such considerations must be accounted for when deriving

a diffusion coefficient κ_p for the particles (Volk et al. 1980). The diffusion coefficient can then be represented as an integral of the effects of each eddy size. For particles settling in a turbulent disk, Dubrulle et al. (1995) derive

$$\kappa_p = \kappa_0 \left[\int_{k_0}^{\infty} \frac{E_k\,(k)\;\mathrm{d}k}{k^2} \right]^{1/2} \quad ,$$

where the coefficient κ_0 represents the effects of the largest eddies, which dominate the transport, and is a function of t_f, v_{zp} and k_0 This diffusion coefficient κ_p is to be used in the equation governing the diffusive settling of particles:

$$\frac{\partial \varrho_s}{\partial t} + \frac{\partial\,(\varrho_s v_{zp})}{\partial z} = \frac{\partial}{\partial z} \left[\varrho \kappa_p \frac{\partial\,(\varrho_s/\varrho)}{\partial z} \right] \quad .$$

Dubrulle et al. (1995) consider a disk with an isothermal vertical structure and calculate the evolutions of initially uniform dust distributions, of a single particle size, for various values of α. They find that stationary solutions, shown in Fig. 22 (for $\alpha = 2 \times 10^{-3}$), are attained in a few times the timescale

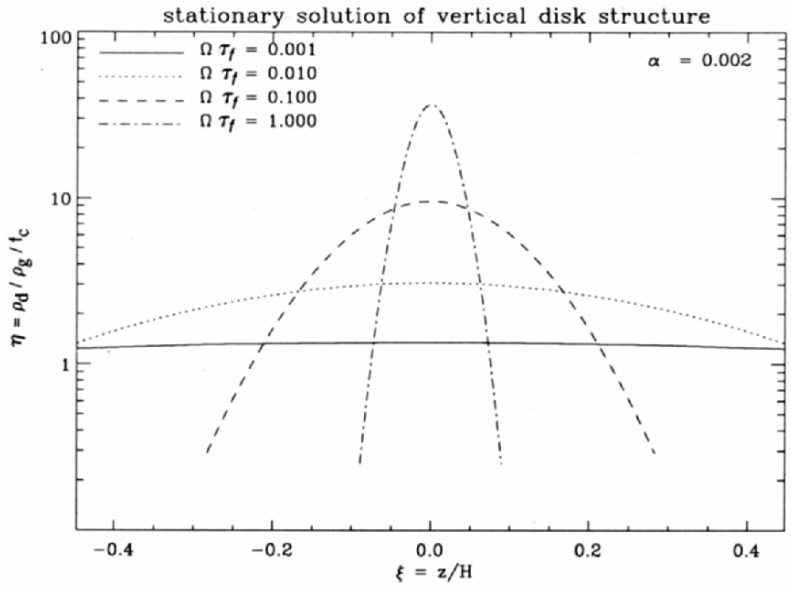

Fig. 22. Steady state, vertical concentrations of solid particles in a turbulent disk with turbulence parameter $\alpha = 0.002$. The quantity plotted on the vertical axis is the enhancement in the concentration of solids above the initial, uniform value. Particles are assumed to be of uniform size. The curves correspond to particle sizes of 250 cm (*dot-dashed*), 25 cm (*dashed*), 2.5 cm (*dotted*) and 0.25 cm (*solid*), for values of the disk surface density of 10^3 gm/cm^2 and particle density 2 gm/cm^3. (Figure from Dubrulle et al. 1995)

$t_f/\left(t_f\Omega\right)^2$. (These solutions can be scaled to other values of the parameters by preserving $\alpha/t_f\Omega$.) Strong concentrations about the midplane are only attained by particles considerably larger than 1 cm. The theory predicts that only large rocks ($r_p > 2.5$ m) could settle to a layer in which the density of solids approaches that of the gas (but the analysis breaks down for such large objects, which could oscillate about the midplane). Even if turbulence was vanishingly small in the ambient disk, a dense layer of solids would produce local turbulence by virtue of the velocity difference between it and the gas (Weidenschilling and Cuzzi 1993), so the formation of such a layer appears to be self-inhibiting. The conclusion is that substantial particle growth must occur before a dense, dusty (or rocky) layer can form at the midplane.

4.4 The Initial Stages of Accumulation

The conclusion of the last section has important consequences for the initiation of planet-building. If dense particle layers could form by settling, as in the calculation by Nakagawa et al. (1986), they might form clumps due to a gravitational instability in the layer (Safronov 1969; Goldreich and Ward 1973). The instability would be like that which occurs in a gas disk when the criterion (11) is violated. This is an appealing prospect, because the most primitive meteorites, which are objects representative of the earliest stage of accumulation, appear to be made by the indiscriminant collection of nebular solids, as might occur in such gravitational clumping. But the results described above indicate that gravitationally unstable layers are not readily formed; accumulation must proceed snowball fashion, as individual particles and collections of particles collide. Calculations of this kind of accumulation in the presence of turbulence have been performed (see Wasson 1985) and indicate that it could produce rocky objects on a timescale of 10^4 years. But the effects of fragmentation are either ignored or represented by untested hypotheses; furthermore, the detailed interactions of particles with the turbulent field are approximated by averages which may obscure important physics.

An attempt to address these issues more rigorously was undertaken by Cuzzi et al. (1996, 2001), with interesting results. They noted that the instantaneous distribution of solid particles entrained in a turbulent flow tends to be inhomogeneous; particles are transiently concentrated at stagnant regions of the flow, between eddies. Moreover, there is a preferred particle size for concentration, selected by the condition $t_f = t_\eta$, where $\eta = k_{max}$ and k_{max} is the wavenumber of the smallest eddies. (This kind of concentration is not the same as that which is sometimes seen at the center of vortices, which only occurs when long-lived, large scale eddies are present.) The expected concentrations and preferred sizes depend on the turbulent characteristics.

Now the properties of the smallest eddies can be found by using the well-known Kolomogorov rules of steady turbulence (Tennekes and Lumley 1972). The starting point is the observation that, in steady, 3-D turbulence, energy does not accumulate at any wavenumber despite the fact that it is being

continuously transferred from larger to smaller eddies. Thus, $\dot{E}_k$, the rate at which energy is transferred at wavenumber k, must be independent of k:

$$\dot{E}_k = \frac{v_k^2}{t_k} = \frac{v_k^2}{1/kv_k} = v_k^3 k = v_0^3 k_0 = v_\eta^3 k_\eta \ .$$

Subscript 0 refers to the largest wavelength. Introduce the *Reynolds number*, a dimensionless combination that measures the ratio of inertial forces to viscous forces:

$$R_e = \frac{\text{velocity} \times \text{length}}{\text{kinematic viscosity}} \ .$$

The large scale turbulence is characterized by

$$R_e = v_0/k_0\nu_{mol},$$

where ν_{mol} is the molecular viscosity. The smallest scale is that where energy is dissipated by molecular viscosity, so

$$R_{e\eta} = v_\eta/k_\eta\nu_{mol} \ .$$

From these relations, and the invariance of $\nu_k^3 k$, one finds

$$\eta/k_0 = R_e^{3/4}$$
$$v_\eta/v_0 = R_e^{-1/4}$$
$$t_\eta/t_0 = R_e^{-1/2} \ .$$

These expressions give the turbulent characteristics at the dissipation scale in terms of the large scale turbulent characteristics. So what is the value of R_e in a protostellar disk? It can be expressed in terms of the turbulence parameter α as follows:

$$R_e = \frac{v_0}{k_0\nu_{mol}} = \frac{\alpha c_s h}{\nu_{mol}} \approx \frac{\alpha h}{l_{mfp}} \approx \alpha \times 10^{11} \ .$$

Here we have used the fact that $\nu_{mol} \approx c_s l_{mpf}$, and chosen representative values of $l_{mpf} = 10\,\text{cm}$ and $h = 10^{12}\,\text{cm}$ to obtain the numerical factor. The preferred size for concentration is given by the condition $t_f = t_\eta = R_e^{-1/2} t_0$, so for this we need to know the lifetime of the largest eddies. It seems safe to assume that, in random turbulence, this lifetime would be comparable to an orbital period. With this assumption, the expression for preferred particle size becomes

$$t_f = \frac{\varrho_p r_p}{\varrho c_s} = \frac{1}{R_e^{1/2}\Omega}$$

or

$$r_p = \frac{\Sigma}{\varrho_p R_e^{1/2}} = \alpha^{-1/2} 10^{-3} \ .$$

The numerical factor in the last equality corresponds to $\Sigma = 10^3\,\mathrm{gm/cm^2}$ and $\varrho_p = 3\mathrm{gm/cm^3}$. Values of α commonly expected for disk evolution, or derived from numerical simulations of angular momentum transport, (say, 10^{-4}–10^{-2}) produce concentrations of particles in the size range 10^{-2}–10^{-1}. Cuzzi et al. (2001) have emphasized that this is the range of sizes of *chondrules*, the abundant and ubiquitous igneous pebbles found in primitive meteorites. In fact, the *distribution* of sizes they find in concentrations from numerical simulations matches that of chondrules in meteorites remarkably well (Fig. 23). Also, from such simulations, and a fractal description of the turbulent properties as a function of R_e, they predict large concentration factors for the high

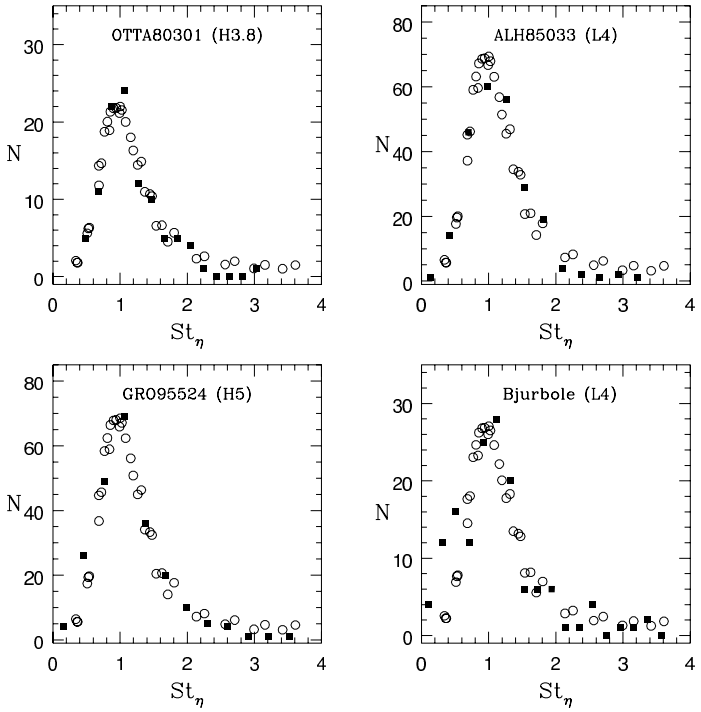

Fig. 23. Size distributions of chondrules from four meteorites (*solid squares*), compared with predicted size distributions from numerical simulations of turbulent concentration (*open circles*). The quantity plotted on the horizontal axis is the Stokes number, a non-dimensional parameter that measures (in this case) the ratio of the friction time, t_f, to the frequency of the smallest turbulent eddies; it is proportional to the product of particle size and density. Thus the plots show, in effect, particle size vs. density. (Figure from Cuzzi et al. 2001)

Reynolds numbers of protostellar disks (see their Fig. 4, where concentrations are expressed as the fraction of total particles expected to exceed a given enhancement in spatial density, as a function of R_e).

From this analysis, Cuzzi et al. (2001), noting that settling is inhibited by turbulence, propose that the initial stages of rocky body accumulation occurred by the formation of chondrule precursors by sticking collisions, their subsequent transformation into chondrules (an enigmatic process, not understood despite much attention and applied creativity), and concentration in the manner described above. However, it must be emphasized that turbulent concentrations are transient; particles actually flow through the regions of concentration, so some further development must be postulated to attain larger solid objects. Perhaps concentrations are high enough to promote gravitational instability or some other positive feedback due to heavy mass loading on the gas by solids. Incidentally, I note that an equivalent analysis of systematic inhomogeneities produced by waves has not been performed.

Other mechanisms for the concentration of solid material at the earliest stage of planet formation have also been suggested. An example is the two-phase (gas-dust) fluid instability proposed by Goodman and Pindor (2000), which produces radial fluctuations in a dust layer at the midplane. Again, the perturbations must grow to the point where gravitational instability sets in, if the process is to lead to the formation of long-lived solid objects.

The fact is, we do not yet know exactly how the first rocks form from the dusty component of protostellar disks. This is the case despite the fact that we actually have samples of these rocks, in the form of meteorites, for our own Solar System. The most primitive of the meteorites look like sediments, the result of a gentle accumulation of disparate components, the ensemble of which escaped familiar planetary equilibrating processes (heat and pressure). Our lack of understanding of how this happened does not prevent us from analyzing the later stages of planet formation, which, in some ways, is a less complicated problem.

5 Growth of Planetesimals to Planets

5.1 Basics of Collisional Planet-building; Runaway Growth

For the present purposes, we define a planetesimal to be a body large enough to gravitationally perturb its neighbors, but smaller than a full-grown planet, say Moon-sized. The first part of this definition can be quantified by saying that a planetesimal produces gravitationally induced velocity perturbations larger than typical drift or settling velocities, perhaps 10^3 cm/sec. The velocity perturbations are comparable to the escape velocity, so we are talking about rocky objects of roughly a kilometer in radius or greater. Somehow these objects formed, but, as indicated in the preceding discussion, gasdynamic effects are sufficiently complex that the mechanisms of growth to this stage

remain poorly described. Further growth is dominated by the gravitational interactions among bodies, so the nature of the problem changes (although gasdynamic effects are not necessarily negligible).

The number of kilometer-sized objects required to build the terrestrial planets is about 10^{11}. Even with modern computers and sophisticated algorithms for integrating the equations of motion, this is far too many bodies to treat by any direct numerical means; statistical methods are required. Despite the difficulty of the problem, a great deal of progress has been made in understanding planetesimal growth in the past two decades. The reasons, I believe, are to be found not only in the advances in computational capability and the technical creativity of researchers, but in a cooperative effort in which a combination of numerical and statistical treatments have been continuously tested against each other to resolve discrepancies.

The statistical approach to understanding the coalescence of many small objects into a few large ones starts with the coagulation equation:

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

Here, n_k denotes the number of objects with mass m_k, and A_{ij} is the probability of collision (and merger) of bodies with masses m_i and m_j. The first term on the right is the gain of bodies with mass m_k and the second term is the loss of bodies with m_k that merge to make bigger bodies. All of the physics is in the collision term A_{ij}, which usually depends in complicated ways on the relative velocities of colliding bodies, their masses, number densities and so forth. Thus one must understand both the detailed physics of collisions (including the effective cross-section for collision) and the dynamical evolution of the ensemble, in order to calculate relative velocities and mass distributions.

Let us start with a description of collisional growth in the simplest situation. If all collisions result in merger, one can say

$$\frac{\mathrm{d}m_p}{\mathrm{d}t} = \left(\pi r_p^2\right) v_{rel} \varrho_s F_g \ .$$

The first factor on the right is the geometric cross-section of the particle, v_{rel} is the mean relative velocity of the ensemble of particles, ϱ_s is their spatial density and F_g is a gravitational enhancement factor, which represents an increase in effective cross-section due to the gravitational bending of particle paths toward the growing object. In the simplest case, the growing body sweeps up smaller particles (with much smaller masses) by two-body interactions. It is readily shown by solution of this "scattering" problem that

$$F_g = 1 + \left(v_{esc}^2 / v_{rel}^2\right) \ ,$$

where the escape velocity is given by

$$v_{esc}^2 = \frac{2Gm_p}{r_p} = \frac{8}{3}\pi r_p^2 \varrho_p \ .$$

Thus

$$\frac{\mathrm{d}m_p}{\mathrm{d}t} = \left(\pi r_p^2\right) v_{rel} \varrho_s \left(1 + \frac{8\pi r_p^2 \varrho_p}{3V_{rel}^2}\right) .$$

Note that the dependence of growth rate on the radius of the growing particle depends strongly on the relative magnitudes of v_{esc} and v_{rel}. If $v_{esc} << v_{rel}$, $\mathrm{d}m_p/\mathrm{d}t \propto r_p^2$; in the opposite limit, $\mathrm{d}m_p/\mathrm{d}t \propto r_p^4$. In general, $\mathrm{d}m_p/\mathrm{d}t \propto r_p^y$, where $2 \leq y \leq 4$. This dependence turns out to be extremely important in determining the nature of solutions to the coagulation equation and the physics of planetary growth.

To see this, consider the growth of two large objects sweeping up smaller objects by two-body interactions. Calculate the time dependence of their mass ratio:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{m_1}{m_2}\right) = \left[\left(\frac{r_1}{r_2}\right)^{y-3} - 1\right] .$$

The "+" term on the right indicates a positive definite factor, the details of which are unimportant for the present point. What is important, is that the ratio m_1/m_2 continuously grows if $y > 3$ (as for $v_{esc} >> v_{rel}$). The possibility of this "runaway growth" of one (or a few) body(ies) with respect to others has a substantial effect on how the planetesimal growth problem must be solved. For instance, because the coagulation equation involves probabilities and deals with mean quantities (as do all statistical treatments), particular care must be used in evaluating such quantities which can be skewed in unphysical ways if the distribution of masses (or other properties) becomes discontinuous. (In fact, there are other aspects of the coagulation equation that are potentially unphysical; see Wetherill (1990), for a discussion.) Because runaway growth occurs when relative velocities are smaller than escape velocities, it is important to have an accurate representation of the dynamical state of the entire ensemble. The central problems of planetesimal growth are therefore those of obtaining real collisional cross-sections and outcomes (including the effects of fragmentation), and describing the real velocity evolution of the system.

5.2 Three-Body Effects on Collision Cross-Section

Greenzweig and Lissauer (1990) used numerical integrations of the restricted three-body problem to determine the gravitational enhancement factor, F_g, for an object sweeping up massless "test" particles in orbit about a star. Averages of the outcomes of many integrations, starting from different initial conditions, were used to evaluate

$$F_g = F_g\left[i,\, e, (r_p/r_{hill})\right] ,$$

where i and e are inclination and eccentricity of a test particle, respectively, and r_{hill} is the radius of the "Hill sphere":

$$r_{hill} = \left(\frac{m_p}{3M_*}\right)^{1/3} a_p \ .$$

The quantity a_p is the semi-major axis of the growing planetesimal. (The Hill sphere is a measure of the extent of the gravitational influence of a secondary compared to the gravitational influence of the primary. Its radius is defined to be the distance from the secondary, in the direction of the primary, at which the potential is a minimum, in a frame rotating with the secondary; that is, the distance to the inner Lagrange point. It arises as a natural parameter in the "shearing sheet" coordinate system.) By performing many integrations and relating the eccentricities and inclinations to relative velocities, the results shown in Fig. 24 were obtained. Scaling relations allow the results to be applied to general values of m_p and a_p. Although the two-body formula for F_g indicates divergence as relative velocities vanish, three-body effects limit its value. It is noteworthy that the orbits of test particles near the growing planetesimal can be extremely complicated (see Fig. 1 of Greenzweig and Lissauer 1990). These orbits should give pause to anyone contemplating a brute force, numerical attack on the entire planetesimal growth problem.

Armed with the details of these encounters, one can say something about the *limits* to runaway growth. Runaway slows when a planetesimal has

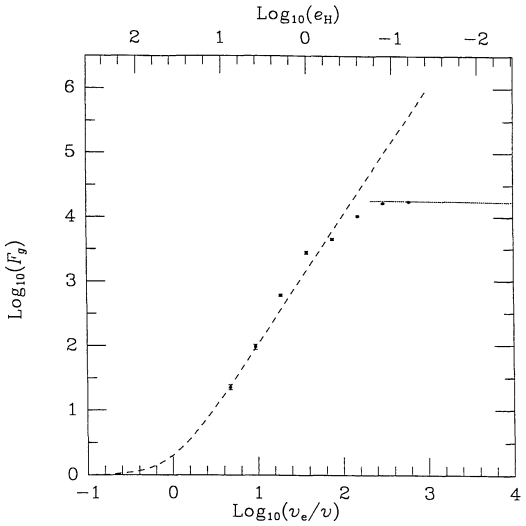

Fig. 24. The gravitational enhancement in collision cross-section, including three-body effects, as a function of the ratio of escape velocity to planetesimal velocity dispersion (or planetesimal eccentricity). The dashed line indicates the two-body approximation. (Figure from Lissauer 1993)

substantially cleared its neighborhood, after its own random velocity has diminished so that the boundaries of its neighborhood are reasonably stable. From numerical integrations like those referred to above, this occurs after test particles are depleted from a zone Δr of a few r_{hill} in width about the planetesimal. The mass contained in such an annulus therefore represents the mass attained after runaway:

$$m_{\text{runaway}} \approx (\text{few times}) 2\pi a_p \Delta r \Sigma_s \approx \frac{\left(10\pi a_p^2 \Sigma_s\right)^{3/2}}{\left(3M_*\right)^{1/2}}, \tag{16}$$

where Σ_s is the surface density of solids. Note that this is not the upper limit to the mass that may ultimately be attained; it is merely the mass that corresponds to the end of the runaway phase, after which the growth rate diminishes substantially. We note here that, for typical parameters, m_{runaway} in the terrestrial planet region is about $0.04 M_\oplus$, or 3 lunar masses.

5.3 Evolution of the Velocity Distribution

One of the most successful techniques for describing the evolution of the planetesimal velocity distribution is based on classical methods of statistical mechanics (Stewart and Kaula 1980; Lissauer and Stewart 1993). It employs the collisional Boltzmann equation, or more specifically a Fokker-Planck equation, in which encounters are regarded as instantaneous and local, and motions between encounters are effectively freely orbiting. (These techniques have also been used in stellar dynamics.) As will be seen, distant, non-local encounters are also important, so modifications must be made to incorporate their effects. The basic premise is that the planetesimal ensemble can be described by as a perturbation to a Boltzmann-like distribution function, $f(\mathbf{r}, \mathbf{v})$, expressed in terms of inclination and eccentricity as $f(i, e)$, where the number density of planetesimals is

$$n = \int f \, \mathrm{d}^3\mathbf{v} \ .$$

Evolution obeys the collisional Boltzmann equation

$$\frac{\partial f}{\partial t} + \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\frac{\partial f}{\partial \mathbf{r}} + \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}\frac{\partial f}{\partial \mathbf{v}} = \left.\frac{\delta f}{\delta t}\right|_{encounter} \ . \tag{17}$$

It is assumed that encounters cause perturbations to an equilibrium distribution, $f = f_0 + f_1$, where $f_1 << f_0$ and

$$f_0(i, e) = \frac{4\Sigma_s}{m_p} \frac{ei}{\langle e^2 \rangle \langle i^2 \rangle} \exp\left(-\frac{e^2}{\langle e^2 \rangle} - \frac{i^2}{\langle i^2 \rangle}\right) \ . \tag{18}$$

Equation (17) is linearized, using the fact that, because f_0 is an integral of free orbital motion, it satisfies the homogeneous equation. The result is a linear, inhomogeneous equation for f_1.

It remains to specify the encounter term on the right hand side of (17). The reader is referred to Wetherill and Stewart (1989) for mathematical details; here I only describe the physical mechanisms that are included. The term is usually divided into five components:

1. Gravitational "viscous stirring", due to close encounters between orbiting bodies converts ordered, orbital energy to random kinetic energy. This component is always positive, in the sense that it increases relative velocities.
2. Inelastic collisions also produce "viscous stirring", which increases random energy at the expense of ordered, orbital motion. Cross-sections for collision are provided by data such as that shown in Fig. 24.
3. Inelastic collisions can, of course, also extract energy from the random component, which reduces relative velocities.
4. The effect of "dynamical friction" is also included in the encounter term, although this effect really depends on the long-range, collective interaction of large numbers of small objects with their larger counterparts. It acts to drive the system toward a state of energy equipartition, in which different masses have the same energy. Thus, it reduces the relative velocities of the more massive bodies, and is a thereby a critical driver of runaway accretion.
5. Gas drag extracts energy from both the ordered orbital motion and the random velocities, but it can actually increase the relative velocities of disparate masses by differential drag.

It is useful to examine steady state velocity distributions (the mean relative velocity as a function of mass) derived by setting the sum of encounter terms equal to zero. An example is shown in Fig. 25. The index q defines the power law mass distribution; the value $q = 2$ gives an equal mass in each logarithmic mass interval. When mass is distributed rather smoothly among the bodies ($q < 2$), relative velocities are relatively uniform. For steeper power laws, and particularly for $q > 2$, all relative velocities are less, but dynamical friction caused by the relatively large number of smaller bodies dramatically decreases those of the largest bodies. This, of course, promotes runaway accretion. (See the discussion in Lissauer and Stewart 1993).

5.4 Calculating Planetesimal Growth

Modern computer codes that calculate the evolutionary development of an ensemble planetesimals employ complicated algorithms that have been tuned extensively by comparison with direct numerical simulations of more restricted (and therefore more manageable) problems, careful comparison with exact solutions of the coagulation equation (Wetherill 1990), and comparisons among different methods (e.g., Stewart and Ida (2000) and Inaba et al. (2001)). Many issues, not only of physics, but of numerical stability and accuracy must be

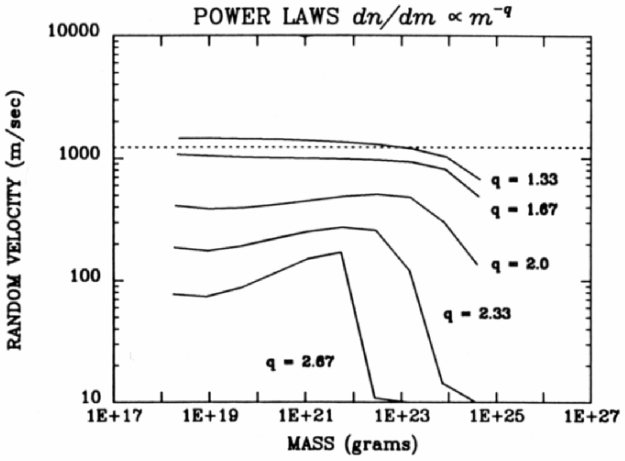

Fig. 25. Equilibrium velocity dispersions as a function of planetesimal mass, for different mass power law distributions. Larger objects tend to have lower random velocities. As the mass power law steepens, the random velocities of the largest objects are significantly depressed, a result of the equipartitioning effect of dynamical friction. (Figure from Lissauer and Stewart 1993)

confronted. To give some flavor of what is involved in a comprehensive calculation, we describe, with unfortunate but necessary brevity, the calculation of Wetherill and Stewart (1993).

As in all such statistical treatments, a "box" of planetesimals is considered to be representative of the ensemble evolution at a given distance from the star. Wetherill and Stewart (1993) start with a box at 1 AU, of radial width $\Delta a = 0.17$ AU, containing approximately 10^9 equally massive objects (i.e., all objects reside in a single mass bin, $m_1 = 4.8 \times 10^{18}$ gm). At $t = 0$, the distribution of velocities is given by (18), from which mean horizontal and vertical relative velocities, and collision probabilities can be found. The number of mergers, and the change in relative velocities in timestep Δt due to encounters, is then calculated from $\delta f/\delta t|_{encounter}$. The number of mass bins in increased to two; one containing objects with initial mass m_1, and one containing objects with mass m_2 ($= 2m_1$, if all collisions resulted in perfect mergers). Fragmentation is accounted for by creating a multi-mass bin, into which fragments are distributed according to a fragmentation size distribution law. The smallest of these fragments are particularly vulnerable to the gas drag term in $\delta f/\delta t|_{encounter}$, and may be lost from the system (dragged into the Sun). The populations and mean relative velocities are calculated for each mass bin, and each is advanced through Δt, the number of mass bins being increased as bodies merge. But soon the number of objects in the largest mass bins becomes rather small, challenging the validity of mean quantities calculated for those bins. So an algorithm for identifying and treating such runaways must

be implemented. Wetherill and Stewart (1993) find that these runaways tend to be on circular, coplanar orbits, and so become isolated from each other (but not from smaller objects). They define a "gravitational interaction radius" $R_g = R_g\,(r_{Hill}, a, e)$ and regard as *isolated* the largest N bodies, where N is defined by the condition that

$$\sum_N R_g \geq \Delta a \ .$$

These objects are not permitted to collide with each other, but they do interact with objects in bins of smaller mass. For more details the reader should consult Wetherill and Stewart (1993) and subsequent literature.

A representative of the (Wetherill and Stewart 1993) results are shown as successive cumulative mass distributions in Fig. 26. After 10^3 years, 52 bodies

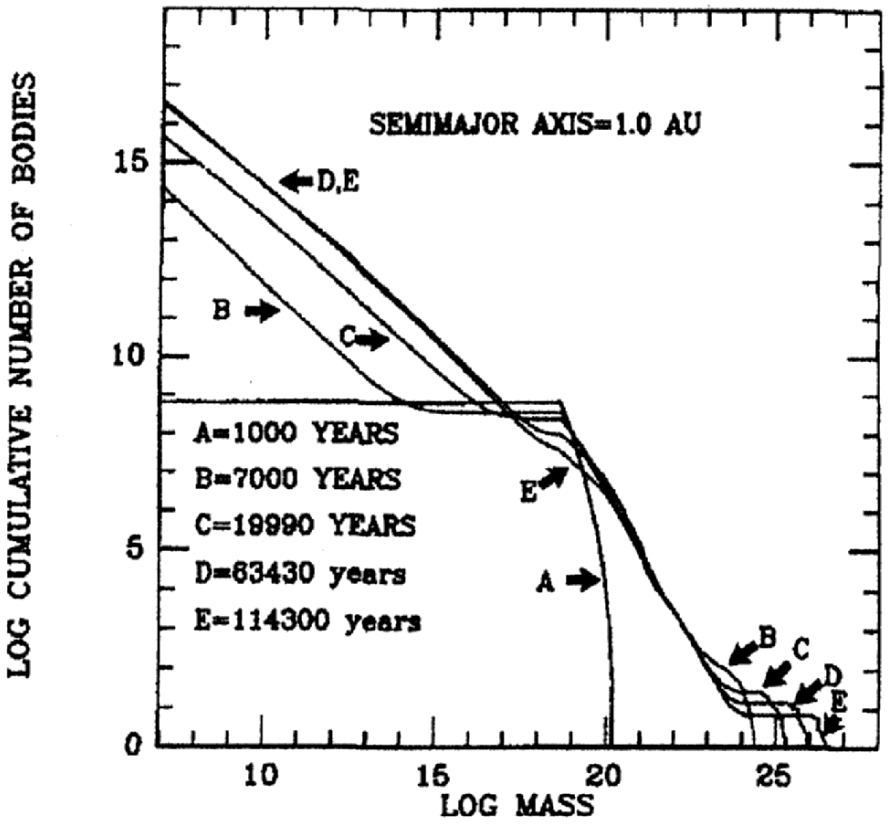

Fig. 26. The results of a calculation of the evolution of cumulative mass distribution in the terrestrial planet. These calculations considered planetesimals in a box, 0.17 AU on a side, located at 1 AU, using special methods to treat the runaway bodies. (Figure from Wetherill and Stewart 1993)

with $m > 30m_1$ have formed; after 7×10^3 years years, 50 bodies have become isolated (the largest being bigger than 10^3 km in radius); by about 20×10^3 years, fragmentation has begun to increase the small body population noticeably; and by 1.2×10^5 years there are 7 runaway bodies containing roughly half the mass of the system, the largest body being the size of Mercury. The results of another calculation is shown in Fig. 27 (Weidenschilling et al. 1997). This simulation is based on similar techniques, but treats the isolation of the largest objects directly by explicitly calculating the properties of individual runaway bodies, and including 100 radial zones, encompassing $a = 0.5$–1 AU. Their results are similar in essential ways to those of Wetherill and Stewart (1993).

5.5 The Final Stage of Accumulation; Rocky Planets in the Terrestrial Planet Region

Planetesimal growth simulations indicate that a final stage of rocky planet accumulation begins when several (or perhaps a few tens) of lunar or Mars-sized protoplanets, which have grown by runaway accretion, become relatively dynamically isolated. Secular resonances among these objects and/or externally produced perturbations (i.e., those produced by giant planets) then must increase their eccentricities and inclinations to the point where their orbits cross and further collisional growth ensues. Modern computeralgorithms (e.g., Sugimoto et al. (1990), Wisdom and Holman (1991), Saha and Tremaine (1992), Saha and Tremaine (1994), Makino et al. (1997), Duncan et al. (1998) and Chambers (1999)) can integrate the orbits of several tens to hundreds of individual gravitating objects for hundreds of millions of years, and so can be applied to the problem of this final stage. Monte Carlo techniques (e.g., Wetherill (1992, 1994, 1996)) have also been applied with considerable success, although they neglect secular effects. Inevitably, such calculations are performed specifically for the terrestrial planets and asteroids, as they provide the only currently available quantitative test. Here, I only summarize the main features of current results.

A typical calculation which begins with several tens of "protoplanets" in the terrestrial planet region usually evolves to a situation in which only a few planets remain (Fig. 28). The entire process may take a few times 10^8 years. A few conclusions about the process are robust: (1) If there are no external perturbers, mass and angular momentum are conserved to a high degree (only a few percent lost in ejected bodies or collisions with the star). In the presence of Jupiter or Saturn, more objects are lost, the amount depending on how much material originally resided close to the regions most affected by the giant planet resonances (i.e., beyond about 2 AU). (2) After a few hundred million years, a few terrestrial-mass planets remain (if the appropriate mass was present to begin with), although the final orbital configurations and individual masses are stochastically determined. Thus, specific quantitative predictions about the precise nature of the final planets cannot be made.

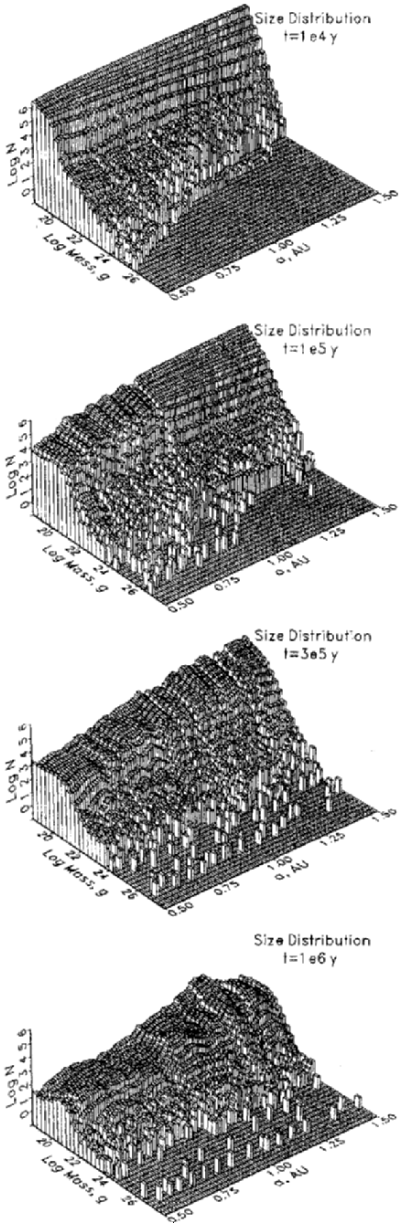

Fig. 27. The results of another calculation of the evolution of cumulative mass distribution in the terrestrial planet region. These calculations considered planetesimals between 0.5 and 1.5 AU, and allowed for the interactions between a continuum distribution of small bodies and a population of discrete runaway objects in individual orbits. There is qualitative agreement with the results shown in Fig. 27. (Figure from Weidenschilling et al. 1997)

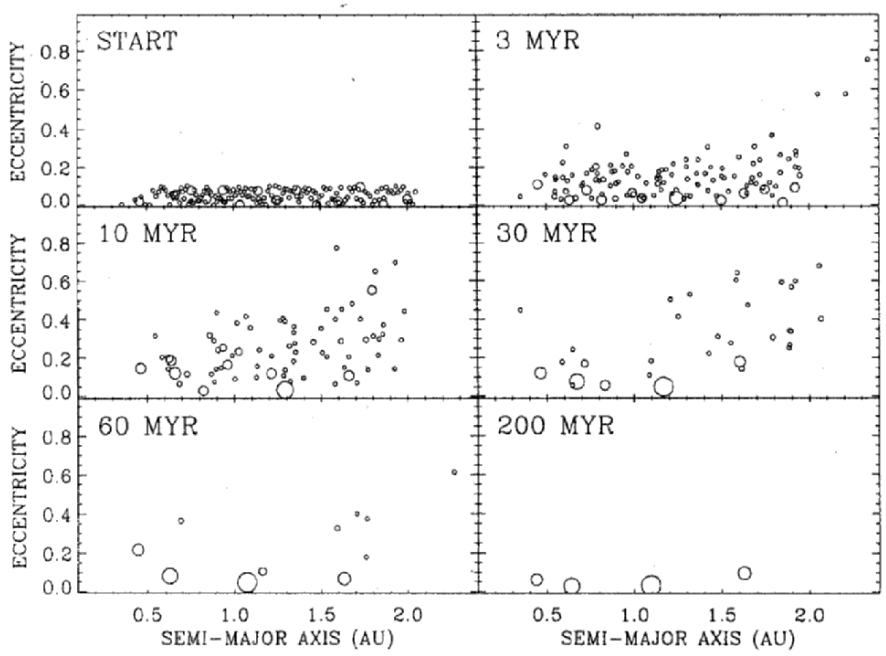

Fig. 28. Results of an N-body simulation (starting with about 150 bodies) of the final stage of planetary formation in the inner Solar System. The size of the symbols are proportional to the planetesimal radii. This particular simulation produces a reasonable facsimile of the terrestrial planets. (Figure from Chambers (2001))

(3) Giant collisions were inevitable during the last stages of rocky planet formation. (4) There is some radial mixing of material, but not complete homogenization, during the period of final assembly.

Some problems remain. Because it is believed that the giant planets formed in less than a hundred million years, their perturbations on terrestrial planet formation are included in the most recent calculations. Yet, when this is done, there is a tendency for N-body simulations to produce eccentricities and inclinations higher than those of the present terrestrial planets. Sometimes this results in only one or two massive planets being formed in the simulations, because more eccentric orbits permit more mergers. The results appear to be somewhat sensitive to the initial mass distribution, and it might be that the configuration of the terrestrial planet system reflects initial conditions not well-represented by the calculations so far. But it is also clear that the locations of the giant planets, and the timing of their growths can have important consequences. See Chambers and Wetherill (1998) and Chambers (2001) for recent discussions of these issues.

6 The Formation of Gas Giant Planets

6.1 Atmospheric Capture or Gravitational Collapse?

Although Jupiter and Saturn contain a greater mass of hydrogen and helium than rock-forming and icy material, they are nevertheless enhanced in the heavy elements relative to solar abundances. This fact excludes gravitational collapse from the protoplanetary disk, a compositionally indiscriminant process, as the sole mechanism by which these planets formed. Moreover, Uranus and Neptune, which have managed to retain large amounts of hydrogen and helium, are still made mostly of the heavier elements. Thus a great deal of attention has been directed toward understanding the growth of planets, by the means described in Sect. 5.5, to the point where they can capture a massive atmosphere. To be sure, one can imagine that these planets formed by gravitational collapse followed by some other process (e.g., preferential loss of gases, or gain of rocky planetesimals) which resulted in their present compositions. Furthermore, what appear to be gas giant planets have been detected around other stars, and we do not know whether or not their compositions are the same as that of their parent star; these might have formed by gravitational collapse. So two distinct modes of giant planet formation may occur; both are interesting.

6.2 Giant Planet Formation by Atmospheric Capture

Let us begin by asking: What do calculations predict for planetesimal growth at 5 AU and beyond? Growth rate is proportional to collision frequency, which is less at 5 AU than at 1 AU, because both orbital frequencies and (presumably) the surface density of solids decrease with distance from the star. Note, however, that settling in a non-turbulent nebula would still be expected to proceed on a 10^3–10^4 year timescale (see Fig. 21). As a planetesimal grows, it is capable of retaining an increasing mass of gas in its atmosphere. In fact, upon reaching a critical mass, it can induce the collapse of all the gas available in its neighborhood, limited only by a finite reservoir, or by dynamical effects associated with rotation (see Sect. 7.1). The planetesimal becomes the core of a gas giant planet.

This mechanism of gas giant planet formation is frequently called the "core instability" model, although it can be caused by evolution to a state in which there is no static equilibrium. The phenomenon is a consequence of the simultaneous constraints imposed by hydrostatic equilibrium and radiative transfer. A simple illustration is given by Stevenson (1982), although his argument involves some sleight-of-hand. Hydrostatic equilibrium requires that the atmosphere above the core satisfy

$$\frac{1}{\varrho}\frac{\mathrm{d}p}{\mathrm{d}r} = -\frac{GM_r}{r^2},$$

where M_r is the mass contained within radius r. Radiative transport in the optically thick atmosphere is governed by

$$\frac{16\sigma T^3}{3\kappa\varrho}\frac{\mathrm{d}T}{\mathrm{d}r} = -\frac{L}{4\pi r^2},$$

where the luminosity L (supplied by core accretion) and opacity κ are assumed constant, for the sake of example. As long as M_r is dominated by the core mass (and is therefore nearly constant), these equations are satisfied by state variables with the dependencies

$$\varrho \propto 1/r^3, \quad p \propto 1/r^4, \quad T \propto 1/r\ .$$

But when the core becomes large enough to attract an envelope which itself begins to contribute substantially to the total mass, no physically realistic solutions exist; the gravitational force within the atmosphere cannot be balanced by the pressure gradient and the atmosphere collapses. Figure 29 shows the critical core masses calculated by Mizuno (1980), along with an those derived from an analytic expression derived by Stevenson (1982). Note that the critical core mass has a strong dependency on opacity; it also depends on the core mass accretion rate.

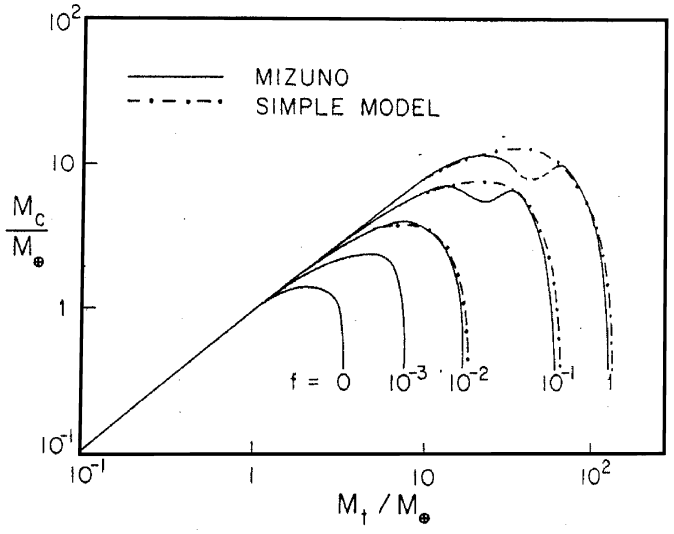

Fig. 29. Core mass as a function of total (core plus atmosphere) mass, for different (normalized) concentrations of dust in the atmosphere (or, equivalently, opacity normalized to a standard, solar composition). The *solid lines* are from calculations of Mizuno (1980); the *dot-dashed line* is derived by a simple model of Stevenson (1982). Portions of the curves with negative slope are unstable; no static equilibrium solutions exist to the right of the vertical portions of the curves. (Figure from Stevenson 1982)

What limits the mass once the nebular gas begins to collapse around the core? Undoubtedly, rotational effects become important at some point. Gravitational interactions with the disk, i.e., gap clearing Lin et al. (1996), may act to diminish Lubow and Artymowicz (2000) or terminate atmospheric accumulation. Or, it may be that the mass of available gas diminishes with time, so later-forming planets accrete less gas Shu et al. (1993). It has also been proposed that gas accumulation can be terminated by a hydrodynamic instability associated with accretion, if nebular densities are sufficiently low (Wuchterl et al. 2000), and references therein). There is currently no consensus on this important question.

The mass of Jupiter's core is uncertain. It could be as large as $15\,M_\oplus$, but data constraining its size also permit models with *no* core (Guillot et al. 1997). (If the latter were the case, the atmospheric capture model would be invalid, of course.) Can a core of $15\,M_\oplus$ grow before the gas of the protoplanetary disk has disappeared? Recall that the lifetimes of protostellar disks do not appear to exceed 10^7 years. Recall also that planetesimal growth slows considerably after the runaway phase, and that even in the terrestrial planet region (where planetesimal growth is faster) full-sized planets are attained only after 10^7–10^8 years. Thus it seems necessary that the critical core mass for Jupiter be attained before runaway ceases. The runaway mass is given by (16), which demands that $\Sigma_s \geq 10\,\mathrm{gm/cm^2}$ for $m_c = 15\,M_\oplus$. This surface density, if spread over the Jovian zone, would yield a mass of solids of about $50\,M_\oplus$, substantially greater than that which now resides in Jupiter. We shall return to the question of what happened to this "extra mass" in a moment, but supposing it to have existed, a critical mass core, consistent with the limits imposed by models of Jupiter's composition and structure, could have accumulated within about 10^6 years.

Pollack et al. (1996) have constructed quantitative models of the formation of Jupiter and Saturn, based on the atmospheric capture idea. The study incorporates a model for the growth of planetesimals and a calculation of the hydrostatic structure of the gaseous envelope, including its interaction with accreting planetesimals. They consider a zone in the disk centered on Jupiter's orbit (assumed not to change), with a width $\Delta a = \Delta a\left(r_{Hill}, \left\langle e^2 \right\rangle\right)$. That is, Δa grows as Jupiter's Hill sphere grows, and as the eccentricity dispersion of the accreting planetesimals increases. The planetesimals are assumed to be distributed instantaneously uniform throughout Δa, but their number diminishes as they are consumed by Jupiter (or increases as Δa expands). The fates of the planetesimals as they encounter Jupiter's atmosphere are calculated. They may have a grazing encounter, entering and leaving the atmosphere; they may be trapped by gas drag and disintegrate in the atmosphere; or they may plunge all the way to the core. Their kinetic energy is deposited accordingly. Calculations are performed for an ensemble of impact parameters. Statistical averages are then used as inputs to the atmospheric envelope growth calculation. The envelope is supported by the luminosity of accretion, and gas is supplied to it from the disk in accordance with the growth of the Hill sphere.

As might be suspected from the discussion above, opacity in the envelope is a key variable. Pollack et al. Pollack et al. (1996) assume that it corresponds to cosmic abundances.

Results of this study for the growth of Jupiter are shown in Fig. 30. (Calculations for Saturn were also performed.) Runaway growth of the core to about $12\,M_\oplus$ occurs in less than one million years, after which it becomes dynamically isolated and core growth slows. Atmospheric mass grows faster until, at about 8 million years, it exceeds the core mass and induces dynamical collapse. The authors point out that the quantitative aspects of this solution depend rather sensitively on the assumed surface density of solids. If Σ_s is much greater than $10\,\mathrm{gm/cm^2}$, the fraction of heavy elements (i.e., the core mass) is too large to be consistent with models of Jupiter. If Σ_s is much less than $10\,\mathrm{gm/cm^2}$, the core grows too slowly. Thus the model is essentially "tuned" to provide the maximum core mass allowed by the data, so that it can form within the inferred lifetime of the nebula. Variations on model parameters are discussed in Pollack et al. (1996).

What of the "extra mass" in the Jupiter region implied by this value of the surface density? Theory predicts that a few other potential giant planet cores would form from this material (e.g., Kokubo and Ida 1998). Thommes et al. (1999), noting the difficulty with which planetesimal accumulation theories have in explaining the formation of Uranus and Neptune (bodies become too isolated; growth is too slow) suggested that these planets formed between Jupiter and Saturn and were gravitationally scattered to their present orbital locations during the growth of Jupiter and Saturn. Their orbits were circularized, according to Thommes et al. (1999), by interaction with the residual

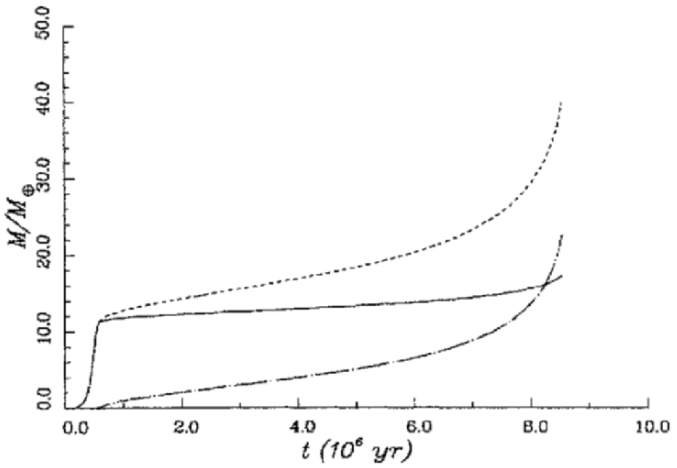

Fig. 30. Growth of Jupiter as a function of time, according to calculations of Pollack et al. (1996). Shown are the core mass (*solid line*), atmosphere mass (*dot-dashed line*) and total mass (*dotted line*).(Figure from Lissauer 1993)

planetesimal disk. This is an appealing hypothesis, potentially solving a few puzzles at once.

At this point, one should realize that the planetesimal accumulation theory of planet formation has attained considerable success in explaining the overall dynamical configuration of the Solar System, the only planetary system available for meaningful tests. But there remain difficulties and uncertainties concerning both the terrestrial and gas giant planets (and the asteroids, which I have not discussed). Moreover, the existence of extrasolar giant planets close to their parent stars is not readily explained by gradual planetesimal accumulation alone. We now turn to the possibility of planetary formation by gravitational instability.

6.3 Giant Planet Formation by Gravitational Collapse

The idea that planets could form, like stars, by direct gravitational collapse of gas and dust together, is an appealing one, particularly in the light of the discoveries of extrasolar planets. Gravitational collapse would likely be fast and efficient, taking only several orbital periods to isolate the planetary mass. This is an obvious advantage, as the atmospheric capture process, as applied to Jupiter, seems barely able to satisfy the requirement that the process be complete before the disk gas disappears. Most of the extrasolar planets exceed the mass of Jupiter and have non-circular orbits. It is not known how effective atmospheric capture might be in producing such objects, but if they formed rapidly by gravitational instability, one can readily imagine situations in which multiple protoplanets scatter each other into irregular orbits, with occasional mergers increasing the largest masses. Indeed, it has been argued (without notable acceptance, I should add) that the masses and distribution of eccentricities of extrasolar companions indicate that they represent a low mass extension of the stellar population, and that they therefore probably formed as stars do, by gravitational collapse, and should not even be called planets (Stepinski and Black 2001). It should be understood, however, that here we are talking about gravitational collapse *within a protostellar disk*, which represents quite a different situation than the collapse of a molecular cloud core.

Gravitational instability occurs if the criterion of (11) is violated. It is possible to estimate Q from the diagnostics outlined in Sect. 1.2.2 and theoretical models. Figure 31 shows the radial distribution of Q calculated from the models of D'Alessio et al. (1998) for typical T Tauri stars. It is seen that the disks are quite stable over all radii less than about 10 AU, but that it is only the effect of stellar radiation that stabilizes them beyond that distance. Thus, if some heating event in the inner disk caused it to swell vertically and shadow the outer disk (for instance), an instability might develop beyond 10 AU. Or it might be that disks were substantially more massive during their early histories, thereby lowering Q, as in the models of Boss (1997, 1998). Given the model results shown in Fig. 31, it is reasonable to suspect that instances of gravitational instability do occur.

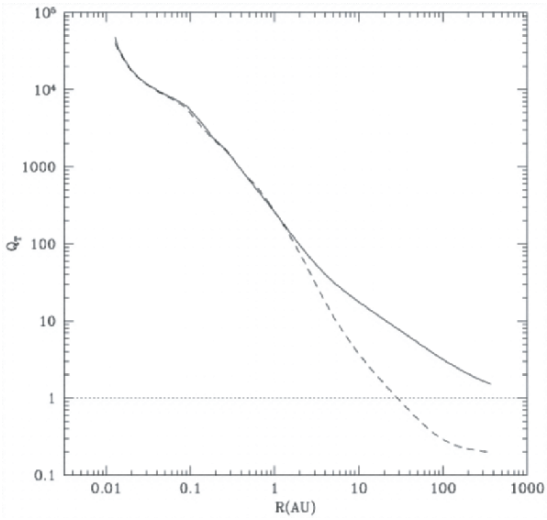

Fig. 31. The radial distribution of the gravitational stability parameter Q (*solid line*) for a typical T Tauri star, calculated from the models of D'Alessio et al. (1998). The dashed line indicates the value that would result if the disk was not illuminated by stellar radiation. Disks appear to be stabilized by stellar radiation over radii less than about 10 AU. They might be unstable at an earlier, more massive, stage, or if shadowed by a hot inner region.(Figure from D'Alessio et al. 1998)

In Sect. 3, I stated that gravitational instability produces spiral density waves. This fact has been confirmed in many numerical simulations, and is not surprising because almost *all* disturbances in highly sheared disks produce spiral structure. But the manifestations of instability are difficult to assess by direct observation. The waves seen vividly in Saturn's rings can usually be attributed to the disturbance of a satellite rather than instability, and the precise origin of galactic spiral structure is complicated by overlying processes such as star formation and molecular cloud evolution. Thus, so far, we have had to rely on the numerical simulations to determine the ultimate fate of unstable disks.

There are two fundamental problems that plague gas disk simulations, unique among numerical astrophysical problems. One source of difficulty is the intrinsic dynamic range of global phenomena in Keplerian disks, as defined by the variation of orbital period over the radial extent of the disk. Gravitational disturbances close to the star develop rapidly and can affect things far away, where the much slower orbital period can retard response. Thus a useful simulation must be sustained, and resolve a large range of orbital periods. This problem has been addressed in galactic simulations by the development of variable timestep techniques and tree codes, in which integrations are performed non-uniformly over the disk; equivalent techniques have not been developed for gas disks. The second problem stems from the fact that

small changes in kinetic and gravitational energy can cause large changes in internal energy, because the latter quantity is a factor of $(h/r)^2$ smaller than the former. And, as we shall see, the response of a disk's internal energy to instability plays a large role in determining the nonlinear development of the instability.

The first problem is somewhat mitigated in simulations of gas giant planet formation by gravitational instability, because the instability and interesting effects are presumed to occur in the outer disk, with minimal influence of the (highly stable) inner disk. Thus the inner disk has been effectively removed from some calculations (e.g., Boss (2000) and Pickett et al. (2000b)), although it remains to be demonstrated rigorously that this is an acceptable procedure. Up until recently, the calculations have also been performed under simplifying assumptions regarding the disposition of internal energy. It has been established that if an unstable disk remains everywhere at its initial temperature (as might be possible under optically thin conditions), instabilities proceed violently to a state dominated by strong density waves and high density contrasts (Boss 1997, 2000; Nelson et al. 1998; Pickett et al. 1998), as in Fig. 32. Boss (1997) argued that some of the density perturbations were, in fact, gravitationally bound and therefore could be regarded as potential protoplanets. But it has also been shown that disks which respond *adiabatically* to instability-generated disturbances sustain much milder density waves (Pickett et al. 2000a; Boss 2000; Nelson et al. 2000). That is, the energy loss rate determines the effect of the instability: isothermal disks, in which energy is easily lost (and gained, in gas undergoing expansion) remain unstable and evolve violently; adiabatic disks tend to heat up and become more stable.

It is clear that processes which transport and dissipate energy must be accurately represented if we are to determine, from numerical simulations, whether or not planets can form by gravitational instability and, if so, what their characteristics would be. I can't help mentioning that these calculations, as difficult (and therefore subject to error of assumption or algorithm) as the calculations of planetesimal growth, have been undertaken by relatively few researchers. A satisfactory theoretical understanding would be hastened, no doubt, by the kind of cooperative attention that has so advanced the planetesimal growth problem. Of course, determination of the heavy element abundances of extrasolar giant planets could provide a definitive test: compositional identity of planet and parent star would heavily favor gravitational collapse.

7 Planet Migration

7.1 Tidal Interaction and Angular Momentum Exchange Between Planet and Disk

The discovery of large planets in orbits very close to their parent stars has focussed attention on a theoretical result that had been previously disregarded

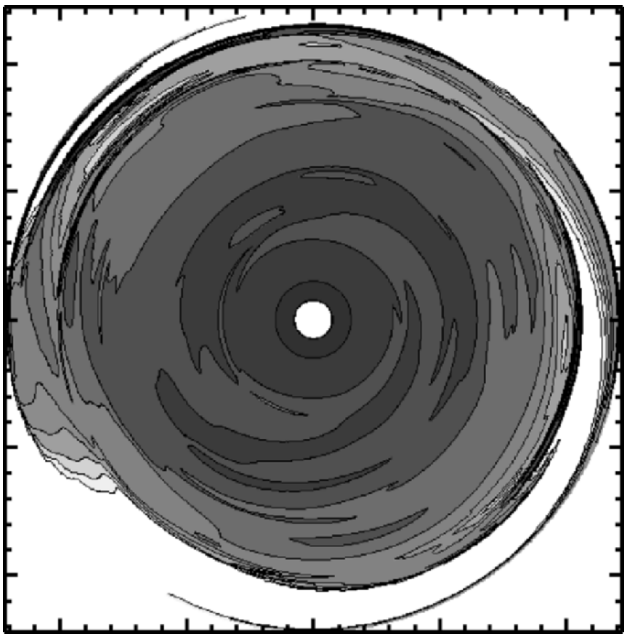

Fig. 32. A representation of equatorial density in a gravitationally unstable protostellar disk. In this numerical simulation, the evolution was assumed to be isothermal; that is, its temperature was fixed everywhere to its initial value. The contour/gray-scale interval is 0.5 in the log of the fractional density perturbation; it spans 4.5 orders of magnitude. An important unresolved question for giant planet formation is whether the high density (*dark*) arcs seen in this simulation eventually evolve into discrete, gravitationally coherent objects (protoplanets), or whether they are transient entities, like the many waveforms that dissolve into the background. (Figure from Pickett et al. 2000b)

in most theories of planet formation. The result states that angular momentum can be exchanged between a planetary body (or other stellar companion) and a circumstellar disk, due to tidal interaction, such that the net torque on the planetary body causes rapid orbital evolution (Goldreich and Tremaine 1979). Thus, even when a solid body has grown too large to be coupled to the gas by viscous drag, it may still be coupled by gravitational forces strong enough to affect its orbital motion.

An intuitive derivation of the tidal torque was given by Lin and Papaloizou (1979), by means of an "impulse approximation", in which the interaction is considered to arise from the local gravitational deflection of fluid parcels in

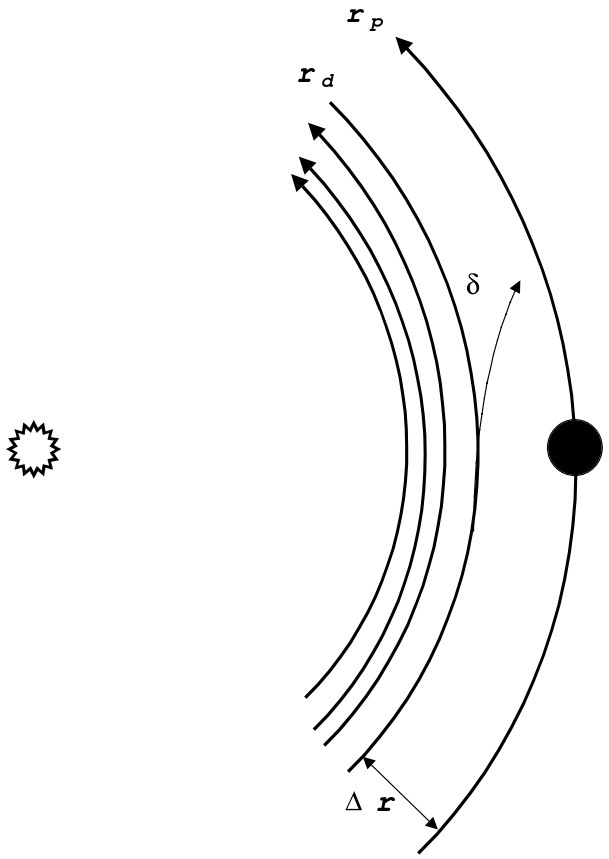

Fig. 33. A planet gravitationally deflects gas near the edge of a disk, changing its angular momentum. If close encounters between the planet and a given fluid element occur in a phase-coherent way, angular momentum is systematically exchanged between planet and disk. This tidal interaction can cause radial migration of the planet

the disk by the planet (Fig. 33). The fluid is deflected through an angle δ as it passes the planet, located Δr away:

$$\delta = \frac{\Delta v_r}{v_{rel}} = \frac{2Gm_p}{\Delta r v_{rel}^2} \ .$$

Here v_{rel} is the relative velocity between disk gas and planet and m_p is the mass of the planet. The change in the gas angular momentum (per unit mass) is

$$\Delta j = -v_{rel} r_d \left(1 - \cos\delta\right) \approx -v_{rel} r_d \frac{\delta^2}{2} = -\frac{2G^2 m_p^2 r_d}{\left(\Delta r\right)^2 v_{rel}^3} \ .$$

This interaction occurs every time the planet comes near the particular fluid under consideration, or every $\Delta t = 2\pi/|\Omega - \Omega_p|$, so the rate of angular momentum exchange is given by

$$\frac{\Delta j}{\Delta t} = \frac{\Delta j}{2\pi/|\Omega - \Omega_p|} = -\frac{G^2 m_p^2 |\Omega - \Omega_p|}{\pi r_p^2 (\Delta r)^2 (\Omega - \Omega_p)^3} .$$

We have used the fact that $v_{rel} = r_d\Omega - r_p\Omega_p$. To find the *total* angular momentum exchanged between planet and disk, this expression should be integrated over the whole disk:

$$\frac{\mathrm{d}J}{\mathrm{d}t} = \int_{\Delta r_d}^{\infty} \Sigma \frac{\Delta j}{\Delta t} 2\pi r \,\mathrm{d}(\Delta r) .$$

But it is assumed that the torques nearest the planet dominate, so Ω may be expanded

$$\Omega^2 = \left(\Omega_p + \left.\frac{\mathrm{d}\Omega}{\mathrm{d}r}\right|_p \Delta r + ...\right)^2 .$$

The result is

$$\frac{\mathrm{d}J}{\mathrm{d}t} = T = -\frac{8}{27}\left(\frac{r_p}{\Delta r}\right)^3 \left(\frac{m_p}{M_*}\right)^2 \Omega_p^2 \Sigma r_p^4 . \tag{19}$$

Remarkably, this formula for the torque on the disk close to the planet is almost correct, requiring only a modest adjustment of the constant to match a more rigorous derivation. Note that, according to this formula, the torque diverges as $\Delta r \to 0$, a fact which leads to the notion that the planet actually clears a gap around itself, by extracting angular momentum from the gas within its orbit and giving angular momentum to the gas outside of its orbit. The size of the gap is presumably set by the Δr at which the tidal torques are balanced by some other (e.g., viscous) torque. It is then the difference between the inner and outer torques that determines the net torque on the planet, i.e., how fast and in what direction the planet's orbit will change.

The problem with this derivation is that it obscures the real physics of the interaction, which is rich in hydrodynamical phenomena, and is really one of resonant (or near-resonant) interaction and wave excitation (Goldreich and Tremaine 1979, 1980). These aspects are revealed only by solving the full fluid equations, as in Sect. 3 for waves, but now including the forcing by the gravitational potential of the planet. The linearized equations that lead to the dispersion relation (9) then have inhomogeneous "driving" terms which represent the Fourier decomposition of this potential. Solutions of the inhomogeneous equations can be matched to the homogeneous, "wave" solutions to obtain a full description of the excitation of the disk caused by the planet (Goldreich and Tremaine 1979; Yuan and Cheng 1989). The torque is indeed concentrated near the planet, but at the special locations identified

previously as the Lindblad resonances (Sect. 3). These local disturbances excite spiral density waves which, behaving as previously described, propagate away from the resonances carrying angular momentum to be deposited wherever the waves damp. The effects associated with each resonance must then be summed to obtain the total angular momentum exchange rate. The torque exerted on the disk *near a single resonance* (designated as the mth Lindblad resonance) is given by Ward (1997a); cf. Artymowicz (1993) as:

$$T_m = -\frac{\pi^2 m \Sigma \psi_m^2}{r \dfrac{\mathrm{d}D_*}{\mathrm{d}r}} \ , \tag{20}$$

where the terms have the following definitions:

$$\psi_m = \frac{r\dfrac{\mathrm{d}\phi_m}{\mathrm{d}r} + 2m\phi_m f}{\sqrt{1 + 4\left(mc_s/\Omega r\right)^2}}$$

$$\phi_m = -\frac{Gm_p}{r_p} b_{1/2}^m \left(r/r_p\right)$$

$$b_{1/2}^m = \frac{2}{\pi}\int_0^\pi \frac{\cos m\theta \, \mathrm{d}\theta}{\sqrt{1 - 2\left(r/r_p\right)\cos\theta + \left(r/r_p\right)^2}}$$

$$f = m\left(\Omega - \Omega_p\right)/\Omega$$

$$D_* = \Omega^2 - m^2\left(\Omega - \Omega_p\right)^2 + \left(mc_s/r\right)^2 \ .$$

The quantity $b_{1/2}^m$ is the coefficient of the mth term in the Fourier expansion of the planet's gravitational potential (in the frame of the center of the disk). Examination of these terms shows that the numerator of (20) contains the intrinsic strength of the forcing term and the denominator contains the distance (measured in frequency space) from the resonance. Artymowicz (1993) showed that the terms proportional to $mc_s/r\Omega$, which had been neglected in previous analyses because this factor was considered to be small, of order the disk thickness ratio, are important in limiting the maximum value of torque. Recall that the impulse approximation indicates that the torque diverges as the distance to the planet Δr approaches zero. But note that the highest order resonances also occur closest to the planet, so the relevant m becomes large as Δr becomes small; thus terms proportional to $mc_s/r\Omega$ are not necessarily negligible. These terms reflect the displacement of the resonant locations due to the non-zero pressure gradient in the disk.

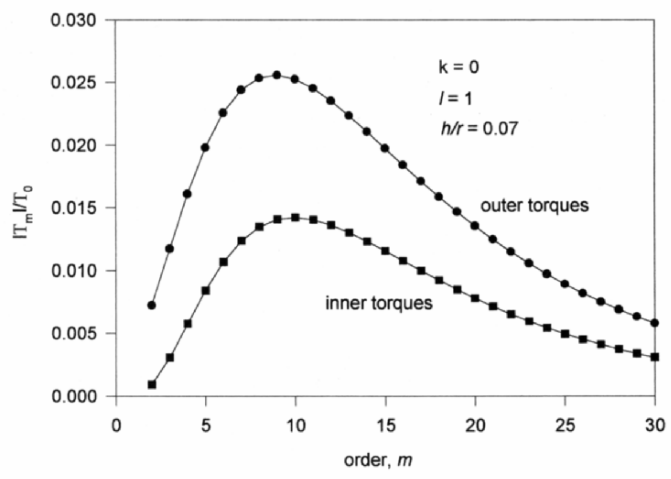

Fig. 34. Normalized torque as a function of mode number m, for typical values of disk parameters. In this figure, k refers to the power law index for the base state surface density (i.e., $k = 0$ for constant surface density) and l is the (negative) power law index for the base state temperature. Outer torques are systematically higher than inner torques, an asymmetry which produces inward planetary migration. (Figure from Ward 1997a)

Ward (1997a) calculated T_m from (20) for a typical disk configuration and obtained the results shown in Fig. 34. A maximum value of T_m occurs for intermediate values of m. Thus the highest order Lindblad resonances are of diminishing importance and there exists a "torque cut-off" which limits the total torque even when there is no gap. Most importantly, note that the torques associated with outer Lindblad resonances systematically exceed those of the inner Lindblad resonances. This imbalance is due to intrinsic characteristics of Keplerian disks; outer resonances are slightly closer to the planet than their inner counterparts, and are less diminished by pressure effects, among other factors (Ward 1997a). The result is that the dominance of the outer resonances produces a net negative torque on the planet, which induces inward migration.

The total torque can be found by the following trick. Calculate the *torque density* by assuming that the torque is distributed more or less smoothly between Lindblad resonances:

$$\frac{\mathrm{d}T}{\mathrm{d}r} = \frac{T_m}{\Delta r_{Lm}} \ .$$

The distance r_{Lm} is the separation of successive resonances; see (12). The total torque is then found as before, from the integral

$$T = \int_{\Delta r}^{\infty} \frac{\mathrm{d}T}{\mathrm{d}r}\,\mathrm{d}r \ .$$

The trick is rigorously justified if the waves excited by the resonant interaction are damped immediately, i.e., at or very close to the resonance. Even if they are not, the *total* torque is usually accurately represented, because the effect of propagating waves is simply to redistribute the torque deposited near the resonance. If the terms proportional to $mc_s/r\Omega$ are ignored, a formula essentially the same as (19) is found, except with the coefficient 0.84 instead of $8/27 = 0.296$ (Goldreich and Tremaine 1980). Including the torque cut-off effect yields a different coefficient, which depends on the details of the disk base state.

7.2 Rates of Orbital Evolution

What do the results described above predict for the rate of planetary migration? To answer this question, one must calculate both the total torque *and* the response of the disk, in a self-consistent manner. Note that the torque is proportional to the local value of the surface density (19), which changes as the disk material is redistributed in response to the torque. Thus one cannot simply use the base state disk conditions to calculate the torque on the planet. Ward (1997a) found an illuminating solution to this problem, in which the planet migrates inward at constant radial velocity, accompanied by a steady, wave-like disturbance in the disk density distribution (Fig. 35). In essence,

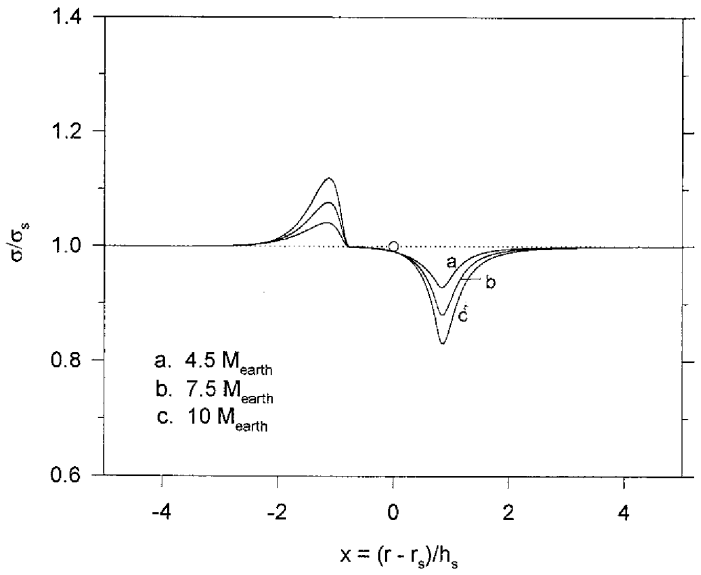

Fig. 35. The surface density perturbation produced by a migrating planet that is not large enough to clear a gap (Type I drift), for three planetary masses and typical disk parameters. The horizontal axis is the distance from the planet (normalized by the scale height) in a frame moving with the planet. (Figure from Ward 1997b)

the solution represents a situation in which the disk surface density is locally disturbed in just the way necessary for the planet-induced torque to produce a constant mass flux in the frame of the radially migrating planet. For an inviscid disk, the planet's radial velocity is given approximately by

$$v_p = .25 c_s \frac{m_p}{M_* Q} \frac{r^3}{h^3},$$

where Q is Toomre's stability parameter (11). For typical parameter values ($c_s = 1\,\mathrm{km/sec}$, $h/r = 0.1$, $Q = 10$, $M_* = 2 \times 10^{33}\,\mathrm{gm}$), one finds $v_p = 7.5\,(m_p/\mathrm{M}_\oplus)$ cm/sec. Thus a $1\,\mathrm{M}_\oplus$ planet would migrate from 1 AU to the Sun in about 6×10^4 years and a $10\,\mathrm{M}_\oplus$ planetary core would migrate from 5 AU to the Sun in about 3×10^4 years. This mobility (called Type I migration, when the planet is not massive enough to clear a gap around itself) can therefore result in large radial displacements of planetary objects in times comparable to or less than their formation times.

It is found, however, that there is a critical mass, m_c, above which a gap in the disk forms around the planet and Type I migration ceases. Because viscosity acts to spread material into the gap, the critical mass increases with disk viscosity; it is given by Ward (1997b)

$$m_c \approx (\text{constant})\, M_* \alpha \left(\frac{M_*}{\Sigma r^2}\right) \left(\frac{c_s}{r\Omega}\right)^2,$$

where the constant depends on specifics of the disk properties, and is generally of order unity. For masses exceeding m_c, radial flow past the planet is inhibited and the planet becomes "trapped" in the gap. The planet and gap then move with the local disk gas, and drift at the rate

$$v_p \approx \frac{\nu}{r} \approx \frac{\alpha c_s h}{r} = \frac{\alpha c_s^2}{r\Omega}\ .$$

This drift is called Type II migration, and is slower than Type I migration, although it is clearly significant on timescales relevant to the evolution of the disk. Note that Type II migration could be outward, if the planet was located in the expanding part of the disk. Ultimately, however, most of the disk moves inward, and the planet would be carried with it toward the star. Figure 36 summarizes the drift rates for Type I and II migration, as a function of planet mass, for a typical disk configuration (Ward 1997a).

Although there is no obvious evidence for tidal drift in the Solar System, it is a likely for the Jovian-sized planets found very close to other stars. That is, it seems implausible that these planets formed so close to their parent stars (Bodenheimer et al. 2000), so one might suppose that they formed further away and then migrated inward to their present locations (Lin et al. 1996). Why might some systems suffer extensive tidal drift while others exhibit no effects? Trilling et al. (1998) addressed this question by calculating the orbital

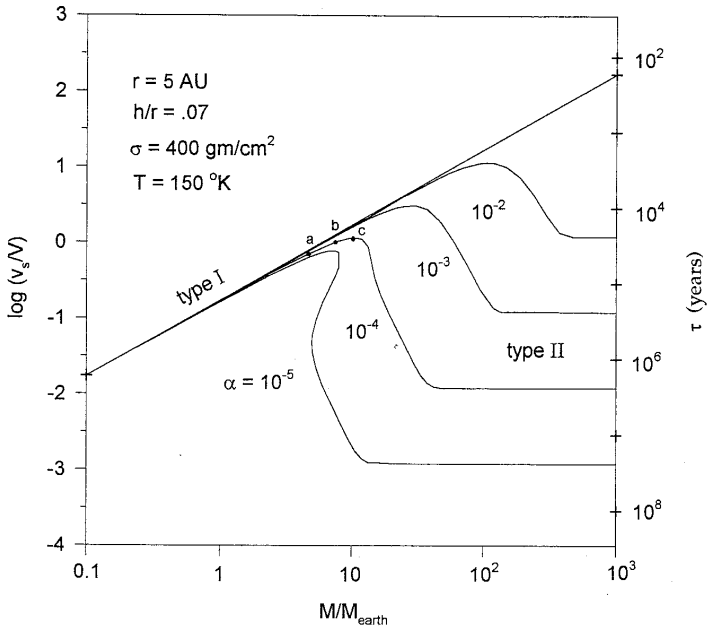

Fig. 36. Inward migration velocity as a function of planetary mass and turbulent viscosity parameter α, showing the transition from Type I to Type II drift. The velocity is normalized by $2\,(M_\oplus/M_*)\left(\pi\, r^2\, \Sigma/M_*\right)(V_K/c_s)^3\, V_K$. The time shown on the right-hand scale is the characteristic orbital decay time from 5 AU. (Figure from Ward 1997b)

evolutions of planets subject to Type II migration in a circumstellar disk. The planets were considered to be fully formed at the beginning of the calculation, and located at a specified distance from a solar mass star, usually about 5 AU. Subsequent orbital migration due to disk tides was calculated using the impulse approximation. If and when a planet approached the star, it suffered additional torques due to mass exchange with the star (Roche lobe overflow) and tidal interaction with the star, both of which tend to increase the orbital angular momentum of the planet and thereafter oppose the disk tidal torques. These torques depend on the size and internal structure of the planet, which were also calculated. Finally, it was assumed that the disk disappeared after a time interval of, say, 10^7, after which tidal interaction with the disk ceased.

Representative results from Trilling et al. (1998), for planets with initial masses between 1 and 5 times that of Jupiter, are shown in Fig. 37. For the case shown (disk mass $1.1 \times 10^{-2}\,\mathrm{M}_\odot$, $\alpha = 1 \times 10^{-3}$), planets with initial masses less than 3.36 Jupiter masses migrate toward the star and ultimately lose most or all of their mass by Roche lobe overflow. Planets with initial masses greater than 3.41 Jupiter masses lose no mass, and the most massive do not migrate far from their initial location before the disk disappears.

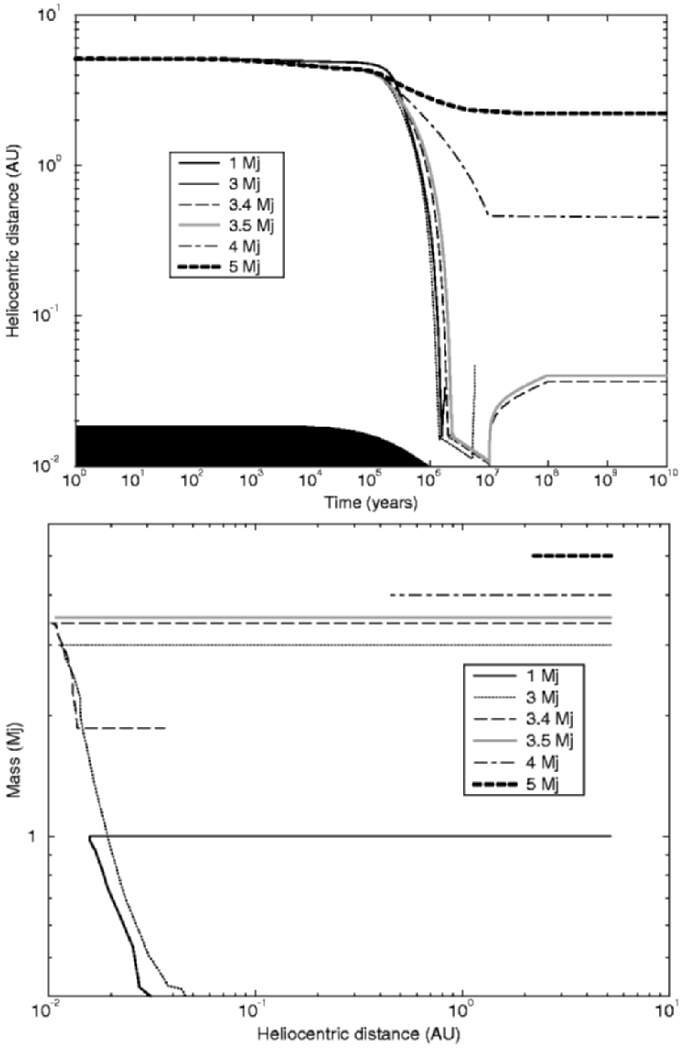

Fig. 37. (*upper*) Orbital radius as a function of time for planets of various initial masses, starting at 5 AU in a typical disk, and (*lower*) planetary mass versus orbital radius for the same planets. Masses are in units of Jupiter's mass. The dark region in the *upper* figure represents the radial extent of the central star. Outward motion is caused by mass loss and tidal interaction with the star. (Figure from Trilling et al. 1998)

Planets with initial masses between 3.36 and 3.41 Jupiter masses migrate to distances at which they lose mass to the star, but are saved by the disappearance of the disk. Trilling et al. (1998) suggest that the few percent of the planets that fall into the last category may account for the extrasolar

giant planets found close to other stars, although the relevant mass range would vary with model parameters (disk mass and lifetime, initial planetary location, etc.).

Note that the analysis of Trilling et al. (1998) is not consistent, in detail, with the solutions found by Ward (1997a) First, Type I migration, which is consequential during the planet building stage, being more rapid than Type II, cannot be rigorously treated in the impulse approximation and is therefore ignored in Trilling et al. (1998). Also, according to Ward's analysis, Type II migration rates are mass-independent, because gap formation in a viscously evolving disk produces density perturbations which adjust the torque imbalance on the planet to that which causes the planet migration to match the movement of disk material. In contrast, according to Fig. 37, larger planets migrate slower than smaller ones, a result attributed by Trilling et al. (1998) to the larger gaps cleared by the larger planets. In fact, the final word on planet migration has yet to be spoken. Not all of the potentially relevant feedback effects have been examined (e.g., the effects of waves on thermal state, turbulent viscosity, dissipation, etc.). The interactions of multiple migrating planets have yet to be fully assessed. For instance, differential Type I migration of planets of differing masses could affect planetary growth rates (Ward and Hahn 1995). Numerical simulations by Artymowicz and Lubow (1996) show that gap formation may not be complete, in the sense that material can be accreted through the gap, which, in turn, would affect the migration rate. See Ward (1997a) and Lin et al. (2000) and Ward and Hahn (2001) for discussions of some of these issues.

7.3 Modeling the Formation of the Solar System

The Solar System will continue to provide the main test case for theories of planet formation until the orbital configurations of other systems have been determined to a much greater degree of completeness than they are presently known. Thus our quantitative theories of particle accumulation, collisional growth, giant planet formation, planetary migration, etc. must, in the end, satisfactorily explain the dynamical state of the Solar System, which we know to high accuracy. As discussed in these lectures, much progress has been made, but the problem is of such complexity that a completely satisfactory model does not yet exist, and may indeed be unattainable without further specific constraints.

In fact, aside from dynamical factors, there exists a wealth of information about the composition and physical state of the Solar System's individual members, which must also be consistent with our theories. For instance, the outstanding characteristic of Solar System planetary composition is the extreme volatile depletion of the terrestrial planets compared to the gas giants. The former lack not only their cosmic share of the light gases hydrogen and helium and noble gases, but are also vastly deficient in water and carbon- and nitrogen-containing compounds. This state is commonly attributed to

the relatively small masses of the terrestrial planets and the higher temperatures they have experienced by virtue of their proximity to the Sun, but a quantitative integration of this idea with a general theory of planet formation is lacking.

Among the most intriguing objects of the Solar System are the meteorites: rocky fragments of asteroids, as well as some from the Moon and Mars, whose orbits have brought them to Earth. Of the many kinds of meteorites, the most common (the "chondritic" meteorites) are primitive, in the sense that they have escaped the equilibrating processes (sustained high temperatures and pressures) that are characteristic of terrestrial rocks. Thus they are composed of an unequilibrated mix of pebbles and fine-grained material. The outstanding features of chondritic meteorites are the following:

1. They are as old as the Solar System. The ages of individual components, as determined by nucleochronological methods, are typically within several million years of 4.56 Gyr (Wadhwa and Russell 2000, and references therein). The oldest known components (calcium-aluminum rich inclusions, or CAIs) are commonly taken to be the first solid objects formed in the Solar System, with ages as great as 4.566 ± 0.002 Gyr. Despite the ambiguities encountered in interpreting radiometric ages, there is no doubt that the chondritic meteorites have retained their essential character since their formation in earliest Solar System history.
2. Their composition, as measured by the relative abundances of elements, is the same as the Sun's, within a factor of two or so, for all but the most volatile elements. Although there are modest systematic differences in composition, these meteorites formed from "solar material", presumably the protoplanetary disk itself, and have experienced little compositional modification since then.
3. They are composed largely of igneous (melted and solidified) pebbles, the "chondrules" from which the chondritic meteorites derive their name. These chondrules exist within a matrix of fine-grained material to form the bulk rock, which, laboratory studies have shown, could not have experienced the high temperatures required to melt the individual chondrules. Moreover, there also exist within the meteorites, material with isotopic compositions which identify them as *pre-solar*, and which apparently survived the formation of the Solar System intact, to be incorporated and preserved in these primitive rocks. For the detailed properties of chondritic meteorites and pre-solar material, see Bernatowicz and Zinner (1997), Kerridge and Matthews (1988), and Hewins et al. (1996).
4. They contain unequivocal evidence for the existence of short-lived radionuclides (with half-lives less than a few million years) in the early Solar System (Goswami and Vanhala 2000). These radionuclides were either made in stellar sources or the interstellar medium shortly before the Solar System formed, or they were produced by energetic events within the Solar System itself.

The first three properties listed above identify the chondritic meteorites as early products of the planet-building epoch. They have therefore been regarded as key elements in understanding planet formation in the Solar System. However, attempts to explain their properties have led to widely divergent hypotheses. I will mention three approaches to the problem, to illustrate the diversity of ideas regarding the origin of the primitive meteorites and their radically different consequences for planetary formation theories.

A traditional approach has been to relate the bulk compositions of planets and chondritic meteorites to the thermal state of the protoplanetary disk at the time of their formation (e.g., Lewis (1974), Cameron (1978), Wasson (1985) and Cassen (2001)). In this case, it is assumed that thermal gradients reflect, directly or indirectly, position in the disk, and that a memory of these gradients has been retained in the final product. This approach focuses on bulk properties and usually has little to say regarding the details of physical state, chondrule formation, or isotopic properties.

A contrasting theory has been extensively developed by Shu et al. (1996, 2001), which postulates extensive redistribution of planetary material by inward drift within the disk and subsequent outward transport by the protosolar wind. The theory takes advantage of the energetic environment near the Sun (within 0.1 AU) to account for the observed thermal processing of chondrules and CAIs (as well as the production of short-lived radionuclides). Because of the high abundance of chondrules in chondritic meteorites (up to 80% by volume), a large amount of material must be recycled from close to the Sun back to the terrestrial planet region and the asteroid belt. Therefore, in this theory, one would expect that the final bulk compositions of planetary material retained little memory of its initial distribution in the protoplanetary disk.

Yet another idea (currently unpopular among meteoriticists) is that the chondritic meteorites are the products of vaporizing collisions among massive (perhaps lunar-sized), volatile-rich planetesimals of chondritic composition. Collisional production of chondrules has been dismissed for various reasons (Grossman 1988; Bos 1996), but I am partial to it because (1) the chondritic meteorites reflect a range of physical and chemical conditions (pressures, temperatures and oxidation states) which might be expected in such collisions, and which are not particularly characteristic of those we believe prevailed in the solar disk, (2) the time scales inferred for chondrule heating and cooling (minutes to days) (Jones et al. 2000) are commensurate with those expected to result from such large collisions, (3) large amounts of material might be locally processed and rapidly re-accumulated, thus preserving the individual characteristics of various meteorite classes and (4) such collisions are predicted by planetary formation dynamics. A consequence of this hypothesis would be that the detailed properties of the chondritic meteorites would have little to do with disk properties or any astronomically observable phenomena. But until a rigorous model for the remnants of such collisions is developed, the hypothesis will remain in speculative limbo.

7.4 Concluding Comments

I began these lectures by noting that, despite the unequivocal genetic association of circumstellar disks with planetary systems, there was reason to be skeptical of the premise that the nature of a planetary system could be predicted (or even modeled from) the properties of its progenitor disk. This skepticism might be provoked just by the existence of planets with highly eccentric orbits. Indeed, an argument can be made that final planetary configurations are determined only by the dynamical laws governing many-body systems, regardless of the origins of the bodies. I mentioned that the distribution of inferred extrasolar planet eccentricities appears to be similar to the distribution stellar binary eccentricities (Stepinski and Black 2001), which could be interpreted to mean that dynamics alone determines the orbits of gravitating objects, regardless of origins. Also, it must be remembered that the first extrasolar planets discovered were two Earth-sized objects in regular orbits about a pulsar (Wolszczan and Frail 1992), a planetary system of orthodox mass and orbital configuration, but surely with an origin and history dramatically different from other systems. One might speculate that we will eventually recognize classes of planetary systems, distinguished by distinct (or a continuous distribution of) dynamical histories, solely determined by the range and variety of stable (i.e., long-lived) states.

The idea that any connection with a primordial disk is effectively obscured is supported by numerous theoretical concepts, besides the gravitational scattering of large objects: the radial mobility of small solid objects in disks due to gas drag, the orbital evolution of planets due to tidal interactions with a gaseous disk, the potential recycling of material by a stellar wind and so forth. And yet the dynamical regularity found in our own Solar System, as well as its enigmatic but systematic compositional regularities, seem to provide sound evidence for inheritance from a disk. It still seems useful, therefore, to pursue the idea that the properties of individual planets, as well as the system they comprise, can be traced to their disk origins.

Acknowledgments

I thank Didier Queloz for inviting me to give these lectures, which I found challenging beyond all expectations. My thanks also to Didier, Michel Mayor, Willy Benz, Stephane Udry and Elisbeth Teichmann for being such gracious hosts and expert organizers, and making our stay in Grimentz so enjoyable. And thanks to Dotty Woolum, who provided invaluable support throughout.

References

P. Artymowicz: ApJ. **419**, 155 (1993)
P. Artymowicz, S. Lubow: ApJLett. **467**, L77 (1996)

S.A. Balbus, J.F. Hawley: ApJ **376**, 214 (1991)
S.A. Balbus, J.F. Hawley: Rev. Mod. Phys. **70**, (1998)
S.A. Balbus, J.F. Hawley: *From Dust to Terrestrial Planets*, (W. Benz, R. Kallenbach, and G.W. Lugmair, Kluwer Academic Publishers 39, 2000)
S.A. Balbus, J.F. Hawley, J.M. Stone: ApJ **467**, 76 (1996)
S.V.W. Beckwith, A.I. Sargent, R.S. Chini, R. Güsten: AJ **99**, 924 (1990)
S.V.W. Beckwith, Th. Henning, Y. Nakagawa:*Protostars and Planets IV*, (V. Mannings, A.P. Boss and S.S. Russell, University of Arizona Press Tucson, 533, 2000)
K.R. Bell: ApJ, **526**, 411, (1999)
K.R. Bell, D.N. C. Lin: ApJ **427**, 987, (1994)
K.R. Bell, P. Cassen, H.H. Klahr, Th. Henning: ApJ. **486**, 372, (1997)
J. Blum: *From Dust to Terrestrial Planets*, (W. Benz, R. Kallenbach, and G. W. Lugmair, Kluwer Academic Publishers, 265, 2000)
P. Bodenheimer, O. Hubickyj, J.J. Lissauer: Icarus **143**, 2, (2000)
A.P. Bos: *Chondrules and the Protoplanetary Disk*, (R.H. Hewins, R. H. Jones and E.R. D. Scott, Cambridge, 257, 1996)
A.P. Boss: Science **276**, 1836, (1997)
A.P. Boss: ApJ **503**, 923, (1998)
A.P. Boss: ApJLett. **536**, L101, (2000)
N. Calvet, G. Gullbring: ApJ, **509**, 802, (1998)
N. Calvet, L. Hartmann, S.E. Strom: *Protostars and Planets IV*, (eds. V. Mannings, A.P. Boss and S.S. Russell, University of Arizona Press, Tucson, 453, 2000)
A.G.W. Cameron: Moon and Planet, 18,5, (1978)
P. Cassen: *LPSC XXIV*, 261, (1993)
P. Cassen: Met. and Planet. Sci. **36**, 671, (2001)
J.E. Chambers: Icarus **152**, 205, (2001)
J.E. Chambers, G.W. Wetherill: Icarus **136**, 304, (1998)
J.E. Chambers: MNRAS **304**, 793, (1999)
E I. Chiang, P. Goldreich: Astrophys. J. **490**, 368, (1997)
J.N. Cuzzi, A.R. Dobrovolskis, R.C. Hogan: *Chondrules and the Protoplanetary Disk*, (R.H. Hewins, R.H. Jones, & E.R. D. Scott, Cambridge Univ. Press, 35, 1996)
J.N. Cuzzi, R.C. Hogan, J.M. Paque, A.R. Dobrovolskis: ApJ **546**, 496, (2001)
P. D'Alessio, J. Cantó , N. Calvet, S. Lizano: ApJ **500**, 411, (1998)
P. D'Alessio, N. Calvet, L. Hartmann, S. Lizano, J. Cantó: ApJ **523**, 893, (1999)
P. D'Alessio, N. Calvet, L. Hartmann: ApJ **553**, 321, (2001)
F. D'Antona, I. Mazzitelli: ApJS **90**, 467, (1994)
C. Dominik, A.G. G. Tielens: ApJ **480**, 647, (1997)
B. Dubrulle, G. Morfill, M. Sterzik: Icarus, **114**, 237, (1995)
M.J. Duncan, H.F. Levison, M.H. Lee: ApJ **116**, 2067, (1998)
P. Goldreich, S. Tremaine: ApJ **233**, 857, (1979)
P. Goldreich, S. Tremaine: ApJ **241**, 425, (1980)
P. Goldreich, W.R. Ward: ApJ **183**, 1051, (1973)

J. Goodman, B. Pindor: Icarus **148**, 537, (2000)
J.N. Goswami, H.A. T. Vanhala: *Protostars and Planets IV* (V. Mannings, A.P. Boss and S.S. Russell University of Arizona Press, Tucson, 963,2000)
Y. Greenzweig, J.J. Lissauer: Icarus, **87**, 40, (1990)
J.N. Grossman: *Meteorites and the Early Solar System*, (J.F. Kerridge and M.S. Matthews, University of Arizona Press, Tucson, 680, 1988)
T. Guillot, D. Gautier, W.B. Hubbard: Icarus **130**, 534, (1997)
E. Gullbring, L. Hartmann, C. Briceño, N. Calvet: ApJ, **492**, 323, (1998)
L. Hartmann: *Accretion processes in Star Formation*, (Cambridge University Press 1998)
P. Hartigan, L. Hartmann, S.J. Kenyon, R. Hewett, J. Stauffer: ApJS **70**, 899, (1998)
J.F. Hawley, S.A. Balbus, W.F. Winters: ApJ **518**, 394, (1999)
J.F. Hawley, C. Gammie, S.A. Balbus: ApJ **440**, 742, (1995)
Th. Henning, R. Stognienko: A&A **11**, 291-303, (1996)
S. Inaba, H. Tanaka, K. Nakazawa, G.W. Wetherill, E. Kokubo: Icarus **149**, 235, (2001)
R.H. Jones, T. Lee, H.C. Connally Jr., S.G. Love, H. Shang: *Protostars and Planets IV* (eds. V. Mannings, A.P. Boss and S.S. Russell University of Arizona Press, Tucson, 927, 2000)
S. Kenyon, L. Hartmann: ApJs **101**, 117, (1985)
H.H. Klahr, P. Bodenheimer: ' Disks, Planetesimals and Planets'. In: *ASP Conference Series* , eds. F.Garzn, C. Eiroa, D. de Winter, T.J. Mahoney, (Astronomical Society of the Pacifc, 2000)
E. Kokubo, S. Ida: Icarus, **131**, 171, (1998)
D.G. Korycansky, J.E. Pringle: MNRAS **272**, 618, (1995)
J.S. Lewis: Science **186**, 440, (1974)
H. Li, S.A. Colgate, B. Wendroff, R. Liska: ApJ **551**, 874, (2001)
C.C. Lin, F.H. Shu: ApJ **140**, 646, 1964
D.N. C. Lin, J. Papaloizou: MNRAS **186**, 799, (1979)
D.N. C. Lin, J. Papaloizou: MNRAS, **191**, 37, (1980)
D.N. C. Lin, P. Bodenheimer, D.C. Richardson: Nature, **380**, 606, (1996)
D.N. C. Lin, J.C. B. Papaloizou, C. Terquem, G. Bryden, S. Ida: *Protostars and Planets IV*, (V. Mannings, A.P. Boss, and S.S. Russell, University of Arizona Press, Tucson, 1111, 2000)
J.J. Lissauer: Ann. Rev. Astron. Astrophys. **31**, 129, (1993)
J.J. Lissauer, G.R. Stewart: *Protostars and Planets III*, (E.H. Levy and J. Lunine, University of Arizona Press, Tucson, 1061, 1993)
R.V. E. Lovelace, H. Li, S.A. Colgate, A.F. Nelson: ApJ **513**, 805, (1999)
S.H. Lubow, P. Artymowicz: *Protostars and Planets IV*, (V. Mannings, A. P. Boss, and S.S. Russell, University of Arizona Press, Tucson, 731, 2000)
S.H. Lubow, G.I. Ogilvie: ApJ **504**, 983, (1998)
S.H. Lubow, J.E. Pringle: ApJ **409**, 360, (1993)
R. Lüst: Z. Naturforsch. **7a**, 87, (1952)
D. Lynden-Bell, J.E. Pringle: MNRAS **168**, 603, (1974)
J.M. Makino, T. Taiji, T. Ebisuzaki, D. Sugimoto: ApJ **480**, 432, (1997)

D. Mihalas: *Stellar Atmospheres* (Freeman, San Francisco 1978)
H. Mizuno: *Progr. Theor. Phys.* **64**, 544, (1980)
Y. Nakagawa, M. Sekiya, C. Hayashi: Icarus, **67**, 355, (1986)
A. Nelson, W. Benz, T.V. Ruzmaikina: ApJ **529**, 357, (2000)
A. Nelson, W. Benz, F.C. Adams, D. Arnett, D: ApJ **502**, 342, (1998)
G.I. Ogilvie: MNRAS **297**, 291, (1998)
G.I. Ogilvie, S.H. Lubow: ApJ **515**, 767, (1999)
B.K. Pickett, P. Cassen, R.H. Durisen, R. Link: ApJ **504**, 468, (1998)
B.K. Pickett, P. Cassen, R.H. Durisen, R. Link: ApJ **529**, 1034, (2000a)
B.K. Pickett, R.H. Durisen, P. Cassen, A.C. Mejia: ApJLett **540**, L95, (2000b)
J.B. Pollack, D. Hollenbach, D. Simonelli, S. Beckwith, T. Roush, W. Fong: ApJ **421**, 615, (1994)
J.B. Pollack, O. Hubickyj, P. Bodenheimer, J.J. Lissauer, M. Podolak, Y. Greenzweig: Icarus, **124**, 62, (1996)
V.S. Safronov: Ann. Astrophys. **23**, 979, (1960)
V.S. Safronov: *Evolution of the protoplanetary cloud and the formation of the Earth and planets*, (NASA TT F-67, chapter 4, 1969)
P. Saha, S. Tremaine: *AJ*, **104**, 1633, (1992)
P. Saha, S. Tremaine: AJ, **108**, 1962, (1994)
N.I. Shakura, R.A. Sunyaev: A&A **24**, 337, (1973)
D.P. Sheehan, S.S. Davis, J.N. Cuzzi, G.N. Estberg: Icarus **142**, 238, (1999)
F.H. Shu: *The Physics of Astrophysics. Vol. II: Gas Dynamics* (University Science Books, Mill Valley, CA., 1992)
F.H. Shu , D. Johnstone, D. Hollenbach: Icarus **106**, 92, (1993)
F.H. Shu, H. Shang, T. Lee: Science **271**, 1545, (1996)
F.H. Shu, H. Shang, M. Gounelle, A.E. Glassgold, T. Lee: ApJ **548**, 1029, (2001)
T.F. Stepinski, D.C. Black: A&A, **371**, 250, (2001)
S.W. Stahler: ApJ **274**, 822, (1983)
D.J. Stevenson: Planet. Space Sci. **30**, 755, (1982)
G.R. Stewart, W.M. Kaula: *Icarus* **44**, 154, (1980)
G.R. Stewart, S. Ida: *Icarus* **143**, 28, (2000)
D. Sugimoto, Y. Chikada, J. Makino, T. Ito, T. Ebisuzaki, and M. Umemura: Nature **345**, 33, (1990)
H. Tennekes, J.L. Lumley: *A First Course in Turbulence* (MIT Press, Cambridge, 1972)
E.W. Thommes, M.J. Duncan, H.F. Levison: Nature **402**, 635, (1999)
H.B. Throop, J. Bally, L.W. Esposito, M.J. McCaughrean: Science **292**, 1686, (2001)
A. Toomre: ApJ **139**, 1217, (1964)
D.E. Trilling, W. Benz, T. Guillot, J.I. Lunine, W.B. Hubbard, A. Burrows: ApJ **500**, 428, (1998)
H.J. Volk, F.C. Jones, G.E. Morfill, S. Roser: A&A **85**, 316, (1980)
M. Wadhwa, S.S. Russell: *Protostars and Planets IV*, (V. Mannings, A. P. Boss, and S.S. Russell, University of Arizona Press, Tucson, 995, 2000)

W.R. Ward: Icarus, **126**, 261, (1997a)
W.R. Ward: ApJLett, **482**, L211, (1997b)
W.R. Ward, J.M. Hahn, J: *ApJ* **440**, L25, (1995)
W.R. Ward, J.M. Hahn: *Protostars and Planets IV*, eds. (V. Mannings, A. P. Boss, and S.S. Russell, University of Arizona Press, Tucson, 1135, 2001)
J.T. Wasson: *Meteorites* (W.H. Freeman, New York, 155, (1985)
S.J. Weidenschilling: Astrophys. and Space Sci, **51**, 153, (1977)
S.J. Weidenschilling: MNRAS **180**, 57, (1977)
S.J. Weidenschilling, D. Spaute, D.R. Davis, F. Marzari, K. Ohtsuki: *Icarus* **128**, 429, (1997)
S.J. Weidenschilling, J.N. Cuzzi: *Protostars and Planets III*, eds. (E.H. Levy and J. Lunine, University of Arizona Press, Tucson, 1031, 1993)
G.W. Wetherill, G.R. Stewart: Icarus, **77**, 330, (1989)
G.W. Wetherill, G.R. Stewart: Icarus, **106**, 190, (1993)
G.W. Wetherill: Icarus **88**, 336, (1990)
G.W. Wetherill: Icarus **100**, 307, (1992)
G.W. Wetherill: Geochim. Cosmochim. Acta **58**, 4513, (1994)
G.W. Wetherill: Icarus **119**, 219, (1996)
J. Wisdom, M. Holman: AJ, **102**, 1528, (1991)
A. Wolszczan, D. Frail: Nature **255**, 145, (1992)
D.S. Woolum, P. Cassen: Met. and Planet. Sci. **34**, 897, (1999)
G. Wuchterl, T. Guillot, J.J. Lissauer: *Protostars and Planets IV*, eds. (V. Mannings, A.P. Boss, and S.S. Russell, University of Arizona Press, Tucson, 1081, 2000)
G. Wurm, J. Blum, J.E. Colwell: Icarus **151**, 318, (2001)
C. Yuan, Y. Cheng: ApJ **340**, 216, (1989)
Astrophysical Implications of the Laboratory Study of Presolar Materials. Edited by Thomas J. Bernatowicz and Ernst Zinner. (1997)
Meteorites and the Early Solar System, Ed. J.F. Kerridge and M.S. Matthews, (1988)
Chondrules and the Protoplanetary Disk, Ed. R.H. Hewins, R.H. Jones and E.R.D Scott, (1996)