CHRIS ORMEL

PLANET FORMATION

PUBLISHED BY

TUFTE-LATEX.GITHUB.IO/TUFTE-LATEX/

First printing, March 2016

Contents

1
Planet Assembly

1.1 Introduction: Planetary size ladder

Planet formation takes place over a variety of size scales. In the core accretion model for planet formation it is postulated that solid particles drive this process, starting from (sub)micron sized grains and ending with cores that can contain $\sim$10 Earth masses onto which the gas of the disk will finally accrete. This transformation corresponds to an astonishing increase of a factor 10^{40} in terms of mass.* Because of the range in scales there are quite a number of physical processes that operate during planet formation. Perhaps the most important two are surface forces (*e.g.* van der Waal's forces), which bind material at small scales, and gravity.

* For comparison, a dust grain is made up of $\sim 10^{10}$ atoms and there are "only" $\sim 10^{11}$ stars in the galaxy.

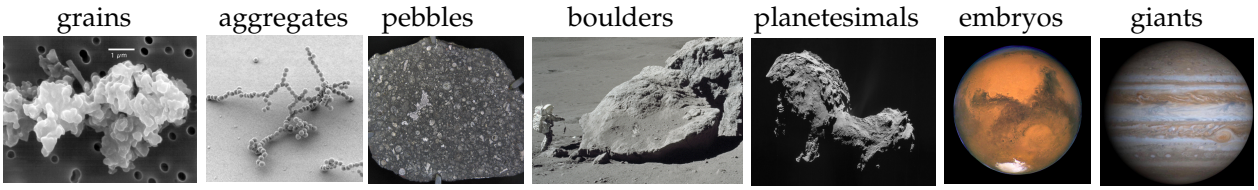

Figure 1.1: The planetary size ladder. From left to right: (i) porous chondrite IDP; (ii) laboratory dust aggregate (Blum et al. 1998); (iii) Allende meteorite, showing chondrules; (iv) boulders on the Moon – NASA photograph AS17-140-21496; (v) Comet 67P on 19 September 2014 – NavCam mosaic by ESA/Rosetta/NAVCAM; (vi) Mars on 23 August 2003 – The Hubble Heritage Team (STScI/AURA); (vii) Jupiter in true color from NASA's Cassini spacecraft on December 7, 2000.

Let us briefly review some of the terminology used in this chapter:

- Grains – up to 1 μm. These are either inherited from the collapse of the parent molecular cloud or have been condensed from the cooling protoplanetary nebula.
- Aggregates/fractals – $\sim\mu$m up to a unspecified upper size. The first product of dust coagulation. They can be very porous.
- Pebbles – sub-mm to meter. We simply mean compact objects larger than grains but smaller than boulders.† Chondrules ($\sim$100 μm) and Calcium-Aluminium Inclusions (CAIs; $\approx$cm), both found in meteorites, are examples.
- Boulders – meter size. In the historically popular "Minimum-Mass Solar Nebula" model, the meter-size scale is approximately the dividing line between particles that are strongly influenced by the gas and particles that travel in Kepler orbits.
- Planetesimals – sub-km to several $\sim 10^2$ km. These are bodies that start to bind material by their gravity. They are traditionally

† This is a broad size range that does not correspond to the geologist definition of a pebble.

considered to be the planetary 'building blocks'. Nowadays their remnants are better known as Asteroids, Comets and Kuiper Belt Objects (KBOs).

- (planetary) Embryos – 10^3 km to $\sim 0.1\, M_\oplus$. Bodies that emerge out of a planetesimal population due to a runaway growth coagulation process. At these masses, bodies start to bind the nebular gas, creating a first atmosphere.* Dwarf planets (Ceres, Pluto) and even Mars belong to this category.
- Rocky planets (not shown in Figure 1.1) – up to several Earth masses. They are believed to have formed by collisions among planetary embryo's – events known as giant impacts. One such giant impact could have resulted in the Earth-moon system.
- Gas giants – $\gtrsim 100\, M_\oplus$. Consist mostly of gas, but (in the core accretion model) thought to have formed only after the creation of a $\approx$10 Earth mass solid core.

* "Protoplanet" and "Oligarchs" are frequently-used synonyms for embryo.

The formation of these bodies takes place in a disk that is dominated by gas. The typical assumption is that the dust-to-gas ratio (or metallicity) of the gas is 'inherited' from the ISM: $Z \approx 0.01$. Understanding planet formation is therefore impossible without understanding the gas and in particular the interaction of these *solids*† with the gas. Gas entirely determines the (relative) motions of small dust particles, but it also damps the random motions (eccentricities and inclinations) of more massive planetesimals and it gravitationally interacts with planets, resulting in their migration.

† solids is a more general name for the refractory material (dust), which also includes larger bodies.

In this chapter we will often assume power-laws for the gas surface density (Σ) and temperature (T) profiles; for example:

$$\Sigma_{\rm gas} = \Sigma_1 \left(\frac{r}{\rm AU}\right)^{-p}, \qquad T_{\rm gas} = T_1 \left(\frac{r}{\rm AU}\right)^{-q} \tag{1.1}$$

where Σ_1 and T_1 are the values at 1 AU. When the disk is isothermal in the vertical direction and its self-gravity can be neglected compared to the Sun, we have seen that the disk *midplane* density is $\rho_{\rm gas}(z=0) = \Sigma_{\rm gas}/h_{\rm gas}\sqrt{2\pi}$ with $h_{\rm gas} = c_s/\Omega$ the gas scaleheight, $c_s = \sqrt{k_B T/\mu}$ the isothermal sound speed, and Ω the local orbital frequency.

A particularly well-known example is that of the so-called *MMSN*: the minimum-mass solar nebula. The MMSN profile is obtained from smearing out the masses of the solar system planets – correcting for the (unknown) dust-to-gas ratio where needed – and (courageously) fitting a power-law to these data points. This results in $p = 1.5$ and $\Sigma_1 = 1{,}700\,{\rm g\,cm^{-2}}$, see Figure 1.2 (Weidenschilling 1977; Hayashi et al. 1985). Although the MMSN is (by convention) a useful beacon, it should be emphasized that its construction assumes a 100% efficiency in converting solids to planets and ignores any radial transport of solids. As we will see, there is no good reason that any of these assumptions should hold.

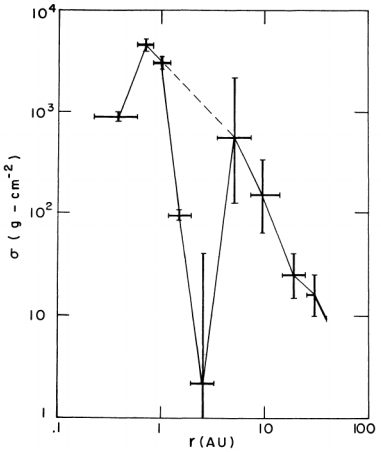

Figure 1.2: Minimum-mass solar nebula model (MMSN): surface density obtained from smearing out the mass of the solar system planets. Note that Mercury, Mars, and the asteroid belt are notable negative deviations to the 'fit'. From Weidenschilling (1977).

As we have seen in the previous chapters, disks are partially pressure-supported with the pressure gradient usually pointing inward‡. As a result the gas will orbit the star at a velocity slightly less than Ke-

‡ At least on large scales. But locally, pressure reversals are a possibility, in which case the "headwind" will become a "tailwind" (negative η).

plerian, *i.e.* $v_{\phi,\mathrm{gas}} = (1-\eta)v_K$ with $v_K = \sqrt{GM_\star/r}$ the Keplerian velocity and $\eta \ll 1$ a dimensionless parameter:

$$\eta \equiv \frac{1}{2\Omega^2 r \rho_{\mathrm{gas}}} \frac{dP}{dr}, \tag{1.2}$$

or, in words: half the ratio of the pressure gradient force to the solar gravity. For the power-law profiles in Equation (1.1) the magnitude of the lag in velocity (a.k.a. the *headwind* velocity) is $\eta v_K \simeq 50\,\mathrm{m\,s^{-1}}$, independent of r.

1.2 *Particle aerodynamics and velocities*

When solids have a relative motion with respect to the gas, they experience a *drag force*:

$$\boldsymbol{F}_D = -\frac{1}{2} C_D A \rho_{\mathrm{gas}} v \boldsymbol{v}, \tag{1.3}$$

where A is the projected surface area of the particle, ρ_{gas} the gas density, $\mathbf{v}$ the relative velocity with respect to the gas, with v the absolute value, and C_D the drag coefficient. The (dimensionless) value of the drag coefficient depends on the particle Reynolds number $\mathrm{Re}_p = 2sv/\nu_{\mathrm{mol}}$ where s is the particle radius.* For spheres:

$$C_D = \begin{cases} 24\mathrm{Re}_p^{-1} & (\text{Stokes:} \quad \mathrm{Re}_p < 1) \\ 24\mathrm{Re}_p^{-0.6} & (\text{Transition:} \quad 1 \leq \mathrm{Re}_p \lesssim 800) \\ \approx 0.5 & (\text{Newton:} \quad \mathrm{Re}_p \gtrsim 800) \end{cases} \tag{1.5}$$

The first, small particle, regime is known as Stokes drag law; the third regime is referred to as Newton's drag and the intermediate Reynolds numbers describe a transition regime. It should be noted that the above values are valid for spheres; bodies of different shapes will have different drag coefficients.

An important exception to these laws is when particles are so small that a fluid description is no longer valid. This occurs when the mean-free path between the gas molecules l_{mfp} becomes greater than the particle size. For such particles, the drag force law follows from a particle description instead. This is referred to as the *Epstein drag* regime, where

$$\mathbf{F}_D = -\frac{4}{3}\pi s^2 \rho_{\mathrm{gas}} v_{\mathrm{th}} \mathbf{v}, \qquad (s < \frac{9}{4} l_{\mathrm{mfp}}) \tag{1.6}$$

with v_{th} the mean thermal speed of the gas molecules.

In the following, we express the strength of the aerodynamic drag in terms of a *stopping time* t_{stop}. This is simply the time after which any initial momentum of a particle is lost due to friction:

$$t_{\mathrm{stop}} = \frac{mv}{F_D}. \tag{1.7}$$

An advantage of this definition is that in the Epstein and Stokes regimes t_{stop} becomes independent of the relative velocity v, such that t_{stop} can be regarded as a property of the particle. For example, $t_{\mathrm{stop}} = \rho_\bullet s / v_{\mathrm{th}} \rho_{\mathrm{gas}}$ in the Epstein regime.†

* Here ν_{mol} is the kinematic molecular viscosity, which is approximately the product of the mean thermal speed and the mean free path of a molecule, $\nu_{\mathrm{mol}} \approx \frac{1}{2} l_{\mathrm{mfp}} v_{\mathrm{th}}$ where l_{mfp} is the (molecular) mean free path, where:

$$v_{\mathrm{th}} = \sqrt{\frac{8k_B T}{\pi\mu}}, \qquad l_{\mathrm{mfp}} = \frac{\mu}{\sqrt{2}\rho_{\mathrm{gas}}\sigma_{\mathrm{mol}}} \tag{1.4}$$

are the mean thermal speed and the mean free path of gas molecules, respectively. σ_{mol} is the molecular cross section.

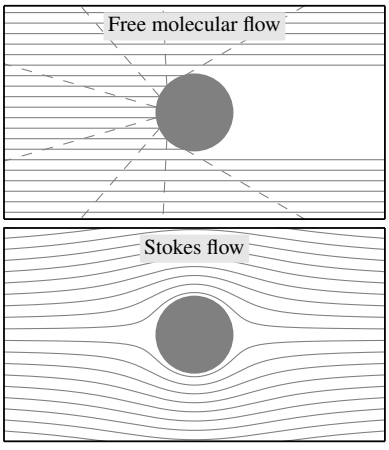

Figure 1.3: Gas streamlines in the free molecular flow, where gas particles bounce off, and Stokes flow ($\mathrm{Re}_p \ll 1$).

Exercise 1.1: Show that the expression for $t_{\rm stop}$ in the Stokes regime is quadratic with the particle size and independent of the gas density.

The stopping time is often nondimensionalized: $\tau_p = t_{\rm stop}\Omega_K$ where Ω_K is the local orbital frequency for circular (Keplerian) motion. Hence, particles with $\tau_p < 1$ decelerate within approximately 1 orbit, whereas particle of $\tau_p \gg 1$ are only marginally affected by gas over 1 orbit. To first order $\tau_p \gg 1$ particles move in Kepler orbits, with the gas providing a small perturbation. On the other hand, orbits of $\tau_p \ll 1$ particles cannot be described by Keplerian orbital elements.

As we have seen, the radial pressure support in the disk causes the gas azimuthal velocity to lag by an amount of $\eta v_K \sim 50$ m s^{-1}. For a solid, however, this pressure gradient force is negligible.* Therefore, solids move with respect to the gas and the excess gravitational acceleration δg_r is balanced by gas drag: $v_{\rm drift}/t_{\rm stop} = \delta g$, resulting in a drift velocity $v_{\rm drift}$. Examples of drift velocities are:

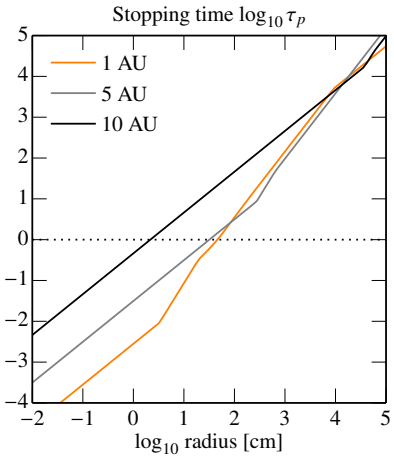

Figure 1.4: Dimensionless stopping time τ_p as function of particle radius at three locations in the disk for the MMSN model.

* To see this, consider the force that arises due to the pressure gradient in the gas: $\sim(dP/dz) \times 2a \times \pi a^2$. This corresponds to an acceleration of $\sim(dP/dz)/\rho_\bullet$ – a factor $\rho_{\rm gas}/\rho_\bullet$ smaller than experienced by the gas.

- *settling*: the gravitational stellar acceleration in the vertical direction $g_{\star z} \approx -z/r \times GM_\odot/r^2 = -\Omega_K^2 z$. Equating this to the drag force, $g_d = v_z/t_{\rm stop}$ we obtain that particles sediment towards the midplane at a velocity $v_z = -\Omega_K^2 z t_{\rm stop}$. This means that dust particles settle to the midplane on a timescale of $t_{\rm settl} \approx \Omega_K^{-1}/\tau_p$ (the *settling time*).

Exercise 1.2 particle scaleheight: When disks are turbulent, particles will diffuse along with the gas. A popular model for the turbulent viscosity is the Shakura & Sunyaev (1973) *alpha-model*, which parameterizes the turbulent viscosity as $\nu_T = \alpha c_s h_{\rm gas}$. When particles are small ($\tau_p < 1$) this is also the particles diffusivity D_p . The definition of D_p is such that particles move a distance $\sqrt{D_p t}$ in time t.

(a) Find the equilibrium distance h_p: the height where the settling timescale (from h_p to the midplane) equals the time to diffuse the particles from the miplane to h_p. Express the result in terms of α and τ_p.

(b) Naively, when $\tau_p \to 0$, one obtains $h_p > h_{\rm gas}$. Why is this result incorrect?

- *Radial drift*. As solids are not affected by the gas pressure gradient, they tend to move towards Keplerian velocities, faster than the gas. Solids therefore experiences a gas drag force in the opposite direction of their motion. Consequently, the corresponding torque removes angular momentum, causing the orbits to decay towards the star.† Thus, solids have both an azimuthal (v_ϕ) and a radial velocity (v_r) with respect to the gas. Balancing forces in the radial direction gives:

$$\frac{(u_{\rm gas}(r) + v_\phi)^2}{r} - \frac{v_r}{t_{\rm stop}} + g_\star = 0, \tag{1.8}$$

where the first term is the centrifugal force, the second term the drag force due to the radial drift, and the third term the stellar

† *i.e.* $d\boldsymbol{l}/dt = \Gamma = \boldsymbol{r} \times \boldsymbol{F}$ where Γ is the torque and $\boldsymbol{l}$ the (specific) angular momentum vector.

* Note the signs: $g_\star$ is negative (pointing inwards) and v_r will also be negative for inward drift.

gravitational force.* Equating the torque to the rate of angular momentum loss gives, per unit mass:

$$\frac{dl_z}{dt} = \frac{d}{dt}((u_{\rm gas} + v_\phi)r) = -r\frac{v_\phi}{t_{\rm stop}}, \tag{1.9}$$

where l_z is the specific angular momentum of a particle and the RHS is the gas drag torque.

Exercise 1.3– individual drift velocities: These two equations allow us to solve for the two unknowns (v_r and v_ϕ). But we can greatly simplify the procedure by using that v_r, v_ϕ, and the disk headwind ηv_K (see Equation (1.2)) are small with respect to the Keplerian velocity v_K. Expressions as $(u_{\rm gas} + v_\phi)^2$ can then be approximated as $u_{\rm gas}^2 + 2u_{\rm gas}v_\phi$. In the same gist, $u_{\rm gas}^2 = (1-\eta)^2 v_K^2 \approx (1-2\eta)v_K^2$, and $d/dt(u_{\rm gas} + v_\phi) \approx dv_K/dt$. This linearization allows the expressions to be put in matrix form:

$$\mathrm{A}\begin{pmatrix} v_r \\ v_\phi \end{pmatrix} = \boldsymbol{b}. \tag{1.10}$$

where A is a 2x2 matrix and $\boldsymbol{b}$ a 2x1 vector. Inverting this system of equations, show that the radial drift velocity becomes:

$$v_r = -2\eta v_K \frac{\tau_p}{1+\tau_p^2} \tag{1.11}$$

and the azimuthal velocity:

$$v_\phi = \eta v_K \frac{\tau_p^2}{1+\tau_p^2}. \tag{1.12}$$

(remark again that v_r, v_ϕ are with respect to the gas velocity.)

These velocity components are plotted in Figure 1.5. Radial drift velocities peak at $\tau_p = 1$ (or $t_{\rm stop} = \Omega_K^{-1}$) and have a magnitude of ηv_K – several tens of meters per second. This is the velocity at which these particles spiral towards the star. The timescale of this orbital decay is $\sim r/\eta v_K$: typically several hundreds of orbital periods for $\tau_p = 1$ particles. At 1 AU $\tau_p \approx 1$ approximately corresponds to meter-size boulders. These particles therefore disappear in a mere 100 years; but even mm–cm size particles should decay relatively quick. The orbital decay problem for $\tau_p \sim 1$ solids is known as the *meter-size problem*, although $\tau_p = 1$ does not always correspond to meter-sized bodies. It is one of the hardest problems in planet formation: how to retain the very solids which the planet formation process relies on.

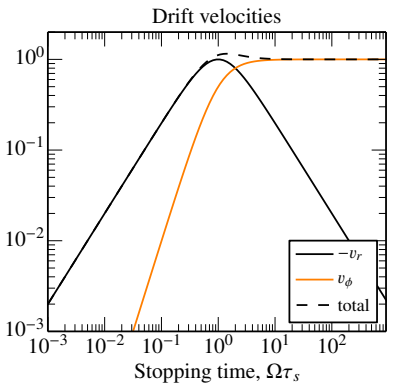

Figure 1.5: Radial and azimuthal drift velocities (in units of ηv_K) as function of dimensionless stopping time.

Drift motions are systematic motions. To obtain the *relative velocity* between two particles, we simply take the difference: $\Delta v = |v_r(t_{\rm stop,1}) - v_r(t_{\rm stop,2})|$. Hence, particles of exactly the same stopping time do not collide. Other velocities are random. For example, the molecules in a gas move in random directions at a mean velocity $v_{\rm th}$. For a hydrogen molecule in a ~300 K gas, the mean thermal motion amounts to $v_{\rm th} \approx 1$ km s. Because of energy equipartition heavier molecules (like H_2O) move slower. Likewise, dust particles also share in the equipartition, in which case the corresponding random motions are referred to as *Brownian motion*:†

† The relative velocity between two particles of different masses can be obtained by replacing m in Equation (1.13) by the *reduced mass* $m_\mu = m_1 m_2/(m_1+m_2)$.

$$v_{\rm BM} = \sqrt{\frac{8k_BT}{\pi m}} \sim 1\ \rho_\bullet^{-1/2} \left(\frac{T}{300\ {\rm K}}\right)^{1/2} \left(\frac{s}{\mu{\rm m}}\right)^{-3/2}\ {\rm mm\ s^{-1}} \quad (1.13)$$

Because of the mass dependence, the significance of Brownian motion quickly diminishes for particles larger than micron. Nevertheless, Brownian motion is important to initiate the dust coagulation process, as relative velocities of drift and turbulence motions are even smaller for (sub)micron sized particles because of their strong coupling.

Finally, turbulence also provides particles with a relative motion. However, the collision velocity between particles is generally much smaller than the motions of the turbulent gas, as the approaching particles are entrained in the same eddy and therefore tend to move in the same direction. In other words, turbulence-induced relative motions can be highly correlated.

To proceed, consider fully developed turbulence* where the energy dissipation rate $\epsilon = v_\ell^2/t_\ell$ is constant across all scales l in the inertial range. Here v_e and t_ℓ are respectively the eddy velocity and the turnover time of scale ℓ. The cascade commences at the largest scales (L, v_L) and proceeds downwards until the local Reynolds number $v_\ell l/\nu_{\rm mol}$ becomes unity. This corresponds to a scale of $\ell_{\rm Kol} = (\nu_{\rm mol}^3/\epsilon)^{1/4}$: the *Kolmogorov scale*. Defining the flow Reynolds number as ${\rm Re} = Lv_L/\nu_{\rm mol}$ we then find $\ell_{\rm Kol} = Re^{-3/4}L$. Correspondingly, the eddy velocity and turnover time become:

$$v_\ell \propto \ell^{1/3} \qquad (v_{\rm Kol} \le v_\ell \le v_L); \qquad v_{\rm Kol} = {\rm Re}^{-1/4} v_L \quad (1.14a)$$

$$t_\ell \propto \ell^{2/3} \qquad (t_{\rm Kol} \le t_\ell \le t_L); \qquad t_{\rm Kol} = {\rm Re}^{-1/2} t_L \quad (1.14b)$$

For a particle of stopping time $t_{\rm stop}$ the eddy spectrum can be split in two classes: those that exist long enough to align the motion of the particle (large eddies) and those with a turn-over time less than the stopping time (small eddies). Hence, if $t_{\rm stop} < t_{\rm Kol}$ all eddies are big and if $t_{\rm stop} > t_L$ all eddies are small. Otherwise, eddies of $t_{\rm Kol} \le t_\ell \le t_{\rm stop}$ are small and $t_{\rm stop} \le t_\ell \le t_L$ are large.

For eddies, we may write the particle forcing in terms of an acceleration of magnitude $g_\ell \sim v_\ell/t_\ell$.† Therefore, particles entrained in big eddies develop a "systematic" (drift) velocity of $v_{\rm sys} \sim v_\ell t_{\rm stop}/t_\ell$ (this velocity is thus relative to the eddy). Small eddies, on the other hand, deliver random kicks to the particle's motion. The velocity change associated with a single kick is of magnitude $\sim v_\ell t_\ell/t_{\rm stop}$. Because of the random nature of the kicks the velocity will increase by $\sqrt{N}$ where N is the number of kicks. As the velocity will saturate after a stopping time, there are $N = t_{\rm stop}/t_\ell$ of these kicks, resulting in a random velocity of $v_{\rm ran} \sim v_\ell \sqrt{t_\ell/t_{\rm stop}}$.

The situation is sketched in Figure 1.6 for a small particle (middle and bottom panel: all eddies are big), a particle of intermediate stopping time (all panels) and a large particle (top panel: all eddies small). The top panel shows the random component of the particle's motion, while the other two show the systematic contribution, with respect to the eddy and space, respectively. For example, eddies of $t_\ell \sim t_{\rm stop}$ dominate the particle's random motion

* see *e.g.* §33 of Landau & Lifshitz (1959)

† This argument goes back to Weidenschilling (1984). Note that in a Kolmogorov cascade, the smallest eddies have the largest accelerations.

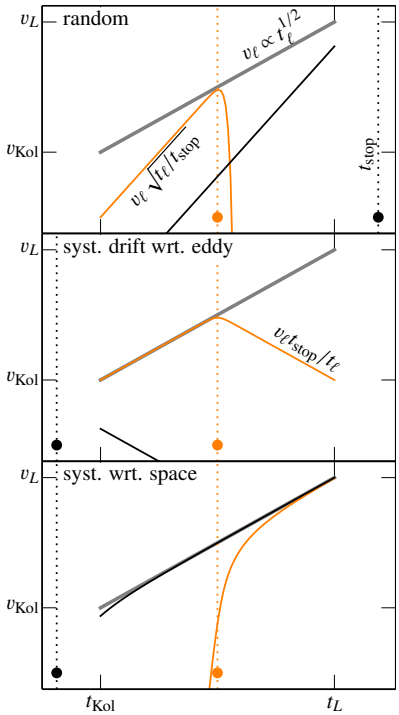

Figure 1.6: Sketch of turbulent excitation of inertial particles. Particles of different stopping times (indicated by dots and vertical dashed lines) couple differently to the turbulent eddy spectrum (gray line). Only eddies of turnover time t_ℓ less than the particle stopping time $t_{\rm stop}$ provide a random component to the particle motion (top). Eddies with turn-over timescales t_ℓ longer than $t_{\rm stop}$ are large and result in a systematic drift of the particle with respect to the eddy (middle). The bottom panel shows the drift with respect to the inertial frame.

when $t_{\rm Kol} \leq t_{\rm stop} \leq t_L$. This toy model can be used to calculate *relative* velocities (Δv) between two particles by adding the (uncorrelated) random components and subtracting the (correlated) systematic component should be subtracted. In particular, when the two stopping times are identical only the random component will contribute towards Δv. This model recovers most of the features of more detailed studies* (Völk et al. 1980; Markiewicz et al. 1991; Ormel & Cuzzi 2007; Pan & Padoan 2010), but is precise only at the order-of-magnitude level.

* An exception is the so-called Saffman-Turner effect (Saffman & Turner 1956), where velocities arise due to shear in the turbulent flow. This mechanism is however of minor importance in astrophysical environments, as the physical size of particles are small with respect to the eddies.

Exercise 1.4 turbulent velocities: Consider driving scales of $t_L = 1$ yr, $c_s = 1$ km s^{-1} and a turbulence Mach number of $\approx$0.1, so that $v_L = 0.1 c_s$. Take a Reynolds number of Re $= 10^8$.

(a) What are the values at the inertial scale, $\ell_{\rm Kol}$, $t_{\rm Kol}$, and $v_{\rm Kol}$?

(b) Given the toy model for the velocity excitation of particles above, as summarized in Figure 1.6, we can derive expressions for the relative velocities of particles. For example, for two particles of stopping times $t_{s1} \leq t_{s2} \leq t_{\rm Kol}$ (where $t_{s1} = t_{\rm stop}$ of particle #1 and t_{s2} of #2) all eddies are large (top panel). In that case, argue that the relative velocity becomes $\Delta v \sim |t_{s2} - t_{s1}| v_{\rm Kol}/t_{\rm Kol}$. (The minus sign is important: the velocity will vanish for $t_{s1} = t_{s2}$. Why?).

(c) In the case where the largest particle (2) has a stopping time in the inertial range, $t_{\rm Kol} \leq t_{s2} \leq t_L$, argue that the relative velocity becomes $\Delta v \sim v_L \sqrt{t_{s2}/t_L}$. Why does this expression not depend on t_{s1}?

(d) If both particles are large, $t_L < t_{s1} < t_{s2}$, argue that it is the particle of the shortest stopping time that determines the relative motions and give Δv.

(e) Between a very small particle ($t_{s1} < t_{\rm Kol}$) and a very big one ($t_{s2} > t_L$) the relative collision velocity is $\Delta v \sim v_L$. Why?

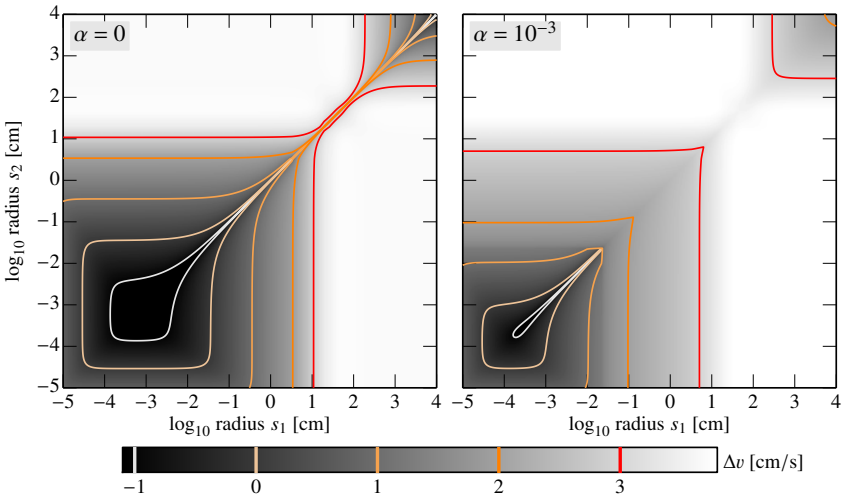

Figure 1.7: Contour plot of the total relative velocity between two particles, including Brownian motion (Eq. [1.13]), radial and azimuthal drift velocities (Eqs. [1.11, 1.12]), and turbulence velocities. For turbulence we adopted the expressions of Ormel & Cuzzi (2007). The relative velocity of each mechanism has been added in quadrature. Left: no turbulence; right: moderate turbulence. Velocities are calculated for MMSN nebular conditions corresponding to 1 AU. Other parameters are: $\rho_\bullet = 3$ g cm^{-3}, Re $= 10^8$, and $v_{\rm hw} = 5 \times 10^3$ cm s^{-1}.

Figure 1.7 presents the relative velocities between two particles for a *laminar* (no turbulence, $\alpha = 0$) and a turbulent disk of $\alpha = 10^{-3}$. In the laminar case, velocities almost vanish along the diagonal $s_1 = s_2$ as drift velocities for identical particles vanish. If one of the particles is very small, Δv is determined by Brownian motion. Otherwise drift motions determine the relative velocities. Between a small and

a large particle, the velocity is determined by the disk headwind ηv_K. Adding turbulence especially boosts Δv when the stopping times of both particles are the inertial regime (*i.e.* $t_\eta < t_{\rm stop} < \Omega_K^{-1}$), or when both particles are massive. Turbulence becomes even more significant for higher α, when the turbulent gas velocities ($\sim\alpha^{1/2}c_s$) start to exceed the disk headwind.

1.3 *Pre-planetesimal growth: modelling dust coagulation*

The densities in the protoplanetary disk are high enough that ISM-size dust particles will collide on a *collision time*:

$$t_{\rm coll} = (n_{\rm dust}\sigma_{\rm col}\Delta v)^{-1} = \frac{s\rho_\bullet}{3Z\rho_{\rm gas}\Delta v}$$
$$\simeq 1\ {\rm yr}\ \frac{0.01\rho_\bullet}{Z}\left(\frac{\rho_{\rm gas}}{10^{-10}\ {\rm g\ cm^{-3}}}\right)^{-1}\frac{a}{\mu{\rm m}}\left(\frac{\Delta v}{{\rm cm\ s^{-1}}}\right)^{-1} \quad (1.15)$$

where the collisional cross section $\sigma_{\rm col} = 4\pi s^2$ (for equal-sized particles; Figure 1.8), $n_{\rm dust} = Z\rho_{\rm gas}/m_{\rm dust}$, Z the dust-to-gas ratio by mass and $m_{\rm dust} = \frac{4}{3}\pi\rho_\bullet s^3$. Equation (1.15) shows that, unless Δv is extremely small, micron-sized particles will collide within the lifetime of the protoplanetary disk, initiating dust coagulation.

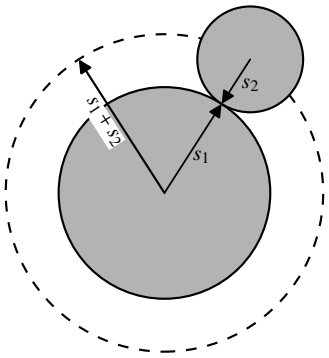

Figure 1.8: Assuming spheres, the collisional cross section (dashed circle) is $\sigma_{\rm col} = \pi(s_1+s_2)^2$.

The particle distribution function $f(m)$ is defined such that $f(m)dm$ gives the number density of particles in mass interval $[m, m+dm]$. Integration over all the masses gives the total number density of particles and

$$\rho = \int f(m)m\ dm \quad (1.16)$$

gives the total (mass) density. Due to sticky collisions, $f(m)$ changes with time at a rate given by the Smoluchowski *coagulation equation*: *

$$\frac{\partial}{\partial t}f(m,t) = \frac{1}{2}\int^m f(m')f(m-m')K(m',m-m')dm'$$
$$- f(m)\int K(m',m)f(m')dm' \quad (1.17)$$

where t is time and $K(m',m)$ the *collision kernel* – the rate at which two particles of mass m and m' collide. The collision kernel is of course simply the product of the collisional cross section and the relative velocity among two bodies, $K = \sigma_{\rm col}\Delta v$.

The Smoluchowski equation is an example of book-keeping: the term on the LHS gives the change in particles of mass m, the first term on the RHS corresponds to new particles of mass m created from the collision of two lower-mass particles and the second term on the RHS to the removal of particles of mass m by collisions with any other particle. Equation (1.17) only accounts for coagulation (perfect sticking). For a more complete picture, other collisional outcomes like (catastrophic) fragmentation, erosion, and bouncing should be added to Equation (1.17). Also remark that Equation (1.17) does not contain a position dependence; it must be combined with a transport

* after M. von Smoluchowski. Equation (1.17) is not to be confused with the (more familiar) equation governing the evolution of the probability density distribution of particles as used in statistical mechanics studies, which also carries his name (Smoluchowski 1916). The discrete form of the coagulation equation reads: $\frac{\partial f_k}{\partial t} = \frac{1}{2}\sum_{i+j=k}K_{ij}f_if_j - f_k\sum_i f_iK_{ik}$

equation to additionally obtain the distribution function as function of position.

But even in its "coagulation-only" form, Equation (1.17), an integro-differential equation, is very hard to solve analytically. There are analytical solutions for only three classes of kernels:

- constant kernels, *e.g.* $K = 1$;
- additive kernels, *e.g.* $K = \frac{1}{2}(m_1 + m_2)$;
- multiplicative kernels, *e.g.* $K = m_1 m_2$.

Solutions to these kernels are given in Figure 1.9, starting, at $t = 0$, from a population of particles at unity mass. These curves are plotted at different times for the three kernels. As can be seen both the constant and additive kernel reach a *self-similar* distribution where the shape of the distribution stays the same. Note that by plotting $m^2 f(m)$ on a logarithmic scale reveals the mass density of the distribution. The situation is different for the product kernel, which evolves to a $f(m) \propto m^{-5/2}$ power-law as $t \to 1$. Thereafter the analytical solution no longer conserves mass and becomes invalid. The physical interpretation is the formation of a runaway body (Wetherill 1990).

Instead of solving (analytically or numerically) for the mass distribution, key information of the solution can alternatively be obtained from the moments of the distribution:

$$M_p(t) = \int f(m', t) m'^p dm'. \tag{1.18}$$

Expressions are listed in Table 1.1. From these moments we can define the mean mass of the distribution as $\langle m \rangle = M_1/M_0$ and the "mass-weighted" mean mass (or peak mass) as $m_p = M_2/M_1$. The latter quantity corresponds to the peak of the distributions in Figure 1.9, except for the product kernel. From the moment expressions of the multiplicative kernel it is also clear that "something happens" at $t = 1$, as M_2 becomes infinite. Mathematically, this is known as *gelation*. The physical interpretation, in this context, is the formation of a runaway body.

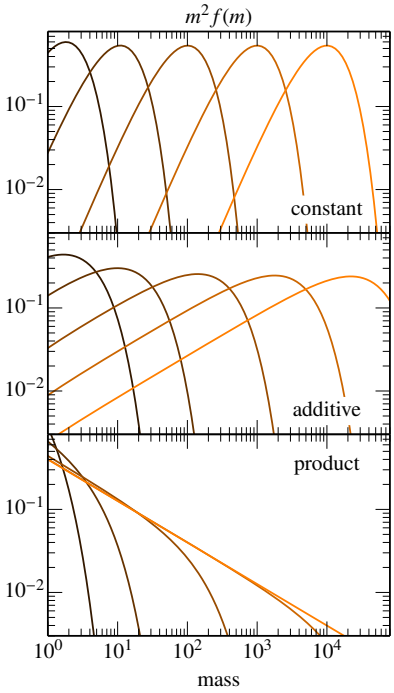

Figure 1.9: Evolution of the size distribution of the constant, additive, and product kernel. Distributions are plotted at times 1, 10, 10^2, 10^4, and 10^5 for the constant kernel; 1, 2.5, 5, 7.5, and 10 for the additive kernel; and 0.1, 0.5, 0.9, 0.99 and 1.0 for the product kernel, where in each case at time $t = 0$ only $m = 1$ particles are present. The distribution function is multiplied by m^2 to display the mass of the distribution on a log-log scale.

Case	Kernel	M_0	M_1	M_2
constant	$K = 1$	$1/(1 + \frac{1}{2}t)$	1	$1 + t$
additive	$K = \frac{1}{2}(m_1 + m_2)$	$\exp(-\frac{1}{2}t)$	1	$\exp[t]$
multiplicative	$K = m_1 m_2$	$1 - \frac{1}{2}t$	1	$1/(1 - t)$

Table 1.1: Zeroth, first, and second moments of the mass size distribution for the three kernel cases that have analytical solutions.

Exercise 1.5– moments:

(a) Show that Equation (1.17) and Equation (1.18) combine into

$$\frac{dM_p}{dt} = \frac{1}{2} \int f(m') f(m'') K(m', m'') [(m' + m'')^p - m'^p - m''^p] \, dm' dm''. \tag{1.19}$$

Hint: rewrite the first term on the RHS of Equation (1.17) as

$$\frac{1}{2} \int f(m') f(m'') K(m', m'') \delta(m - m' - m'') dm' dm'' \tag{1.20}$$

where $\delta(x)$ is the Dirac-δ function.

(b) Clearly, $dM_1/dt = 0$. What is expressed by this?

(c) Write down equations for the zeroth, first and second moments of the constant, additive, and multiplicative kernels. For example, for the constant kernel ($K = 1$), you will find:

$$\frac{dM_0}{dt} = -\frac{1}{2}M_0^2; \qquad \frac{dM_2}{dt} = M_1^2. \tag{1.21}$$

Continue to derive the expressions listed in Table 1.1. Assume that initially (at $t = 0$) $M_0 = M_1 = M_2 = 1$.

(d) For the constant kernel the peak mass m_p and the average mass $\langle m$ are the same within a factor of 2. Explain that this is consistent with Figure 1.9. (Hint: measure the power-law slope of the low-m tail of the distribution.)

(e) For the additive kernel, explain that most particles are small (even for $t \gg 1$).

Are these three analytical kernels of any use for dust coagulation in disks? In reality, K, the product of the collisional cross section and the relative velocity, will never attain the precise forms discussed above. Nevertheless, situations do resemble the three cases. Brownian motion-induced velocity have a $\sim m^{-1/2}$ dependence, which gives rise to an $m^{-1/6}$ – nearly constant – dependence on the kernel. Brownian motion-driven growth will therefore result in very narrow size distributions (Figure 1.9). For drift-induced relative motions, on the other hand, relative velocity typically increases with the particle size, which gives (approximately) a linear dependence, as in the additive kernel. In that case, growth is exponential but most of the particles at any time remain small. This means that the difficult to observe large particles are 'buried' in a sea of small dust grains. Nevertheless, if coagulation is the only process, even the small grains deplete uncomfortably quickly. It is therefore believed that either growth stalls, because sticking fails or because collisions become so energetic that they will replenish the small dust population.

The multiplicative kernel is an example of a non-linear kernel, where $K \propto (\text{mass})^\kappa$ with $\kappa > 1$. In these kernels a runaway body will separate from the (continuous) distribution after a finite time*. In nature, nonlinear kernels occur when the collisional cross section is enhanced by gravitational focusing (see §1.6). In that case runaway bodies, which separate from the distribution, appear.

* In many nonlinear kernels, the runaway time $t_{\rm run}$ approaches zero if the system size is infinitely large (Malyshkin & Goodman 2001; Ormel & Spaans 2008).

1.4 *Pre-planetesimal growth and planetesimal formation*

Before gravity becomes effective to bind material, the outcome of growth depends on (attractive) surface forces – Van der Waals' and hydrogen bonds. These forces are necessary to overcome the repulsive force due to the elasticity of the material. This elasticity is quantified by Young's modulus: the relation between stress (pressure or force) and strain (contraction). Hertz (1882) showed that two particles brought together under an (attractive) force F form a contact

with a surface of radius a given by

$$a = \left(\frac{3sF}{4\mathcal{E}^*}\right)^{1/3} \tag{1.22}$$

where $\mathcal{E}^*$ is the combined elastic modulus * of the material(s) and a the local radius of curvature, which is related to the grain radii as $s^{-1} = s_1^{-1} + s_2^{-1}$.

* Defined $\mathcal{E}^{*-1} = (1-\nu_1^2)/E_1 + (1-\nu_2^2)/E_2$ where E_i and ν_i are, respectively, the Young moduli and Poisson rations of the two spheres.

Exercise 1.6: While the derivation of Equation (1.22) is not trivial, an order-of-magnitude estimate can be obtained as follows. Consider two spheres of radius s in contact that have been indented by a distance δ over a surface area πa^2 due to the applied force (F).

(a) Show that $a^2 = \delta s$. Let $r \gg s$ be the distance from the contact into the sphere. Then argue that the resulting stress σ_r within the grain can be reasoned (*e.g.* by dimensional arguments) to be $\sigma_r \simeq F/r^2$. Along the symmetry axis (z) the stress-strain relation reads:

$$\frac{d\delta}{dz} = \frac{\sigma_r}{\mathcal{E}^*} = \frac{F/2}{\mathcal{E}^* z^2} \tag{1.23}$$

(Hint: Hooke's law for a uniform rod reads $\sigma_r = \mathcal{E}^* \Delta L/L$ with ΔL the indentation and L the length of the rod.)

(b) Choosing a suitable lower cut-off for z, show that integration of Equation (1.23) gives Equation (1.22) barring the numerical factor.

(c) For general (s, δ, a), it can be shown that the potential energy associated with the formation of the contact is (Muller et al. 1980):

$$U_c = \mathcal{E}^* s^3 \left(\frac{a\delta^2}{s^3} - \frac{2a^3\delta}{3s^4} - \frac{a^5}{5s^5}\right) \tag{1.24}$$

Show that F as defined by Equation (1.22) follows from Equation (1.24) when $a^2 = \delta s$.

(d) Why is U_c positive?

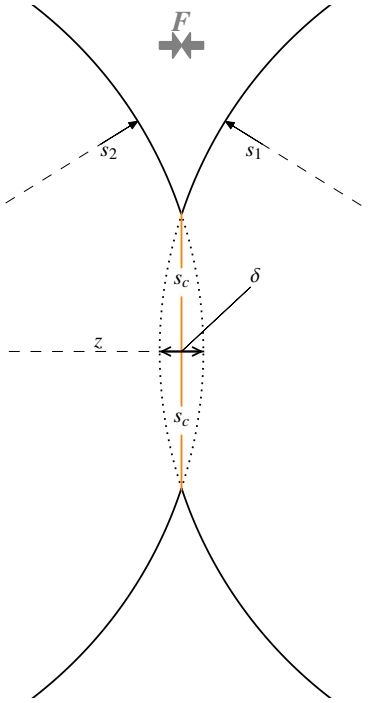

Figure 1.10: Geometry of the contact area. (The contact size has been exaggerated for illustration purposes).

This Hertzian contact model only includes elastic forces. Grains therefore will not stick, unless there is an externally applied force.

The situation changes, however, if we account for the surface energy stored in the contact area:

$$U_{\rm JKR} = U_c + U_s; \qquad U_s = -\pi a^2 \gamma, \tag{1.25}$$

where γ, the energy surface density, is a material constant. The name "JKR" refers to the authors who first constructed this theory†. Minimizing the total energy for a gives a relation between δ and a: $\delta = a^2/s - \sqrt{2\pi\gamma a/\mathcal{E}^*}$. The force thus experienced becomes:

$$F_{\rm JKR} = \frac{4\mathcal{E}^* a^3}{3s} - \sqrt{8\pi\gamma\mathcal{E}^* a^3}. \tag{1.26}$$

Importantly, the equilibrium contact area, where $F_{\rm JKR}$ vanishes is:

$$s_{\rm eq} = \left(\frac{9\pi\gamma s^2}{2\mathcal{E}^*}\right)^{1/3}. \tag{1.27}$$

The situation is illustrated in Figure 1.11. The potential $U_{\rm JKR}$ has a minimum at $a = s_{\rm eq}$ where the force vanishes. The contact is therefore stable. Note that the potential is still negative for $\delta < 0$: grains are bound by virtue of the attractive surface forces.

† Johnson, Kendall, & Roberts (Johnson et al. 1971).

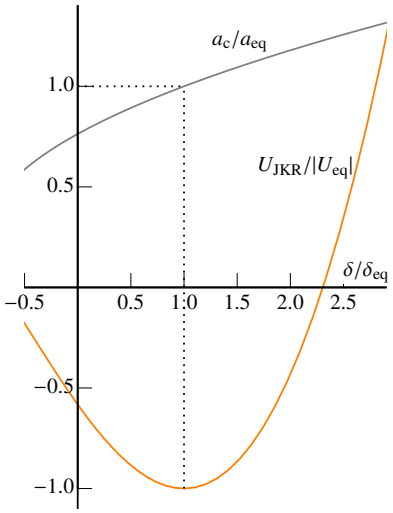

Figure 1.11: The JKR potential. $\delta_{\rm eq}$ and $U_{\rm eq}$ are the offset and energy corresponding to the equilibrium contact size $s_{\rm eq}$ (Eq. [1.27]). The equilibrium point is denoted by the dotted lines.

Exercise 1.7: Show that an energy of $E_{\rm break} \simeq 10\gamma^{5/3}s^{4/3}/(\mathcal{E}^*)^{2/3}$ is needed to break a contact.

The above calculations do not include energy dissipation: an incoming grain should rebound elastically. In practice, however, dissipation mechanisms* are present, which ensures particles stick as long as their impact energy† is less than $\sim E_{\rm break}$. This can be cast in a critical (upper) threshold for sticking:

$$v_{\rm stick} \simeq \frac{4.5\gamma^{5/6}s^{2/3}}{m_\mu^{1/2}(\mathcal{E}^*)^{1/3}} \approx 3 \text{ cm s}^{-1}\, \rho_\bullet^{-1/2} \left(\frac{\mathcal{E}^*}{10^{12}}\right)^{-1/2} \left(\frac{\gamma}{10}\right)^{5/6} \left(\frac{s}{\mu\text{m}}\right)^{-5/6}, \quad (1.28)$$

where m_μ is the reduced mass, $\mathcal{E}^*$ and γ are in cgs-units and the values roughly correspond to silicates (see Table 1.2). Clearly, $v_{\rm stick}$ decreases with grain radius s. As seen above, particle relative velocities typically increase with size. Sticking therefore becomes harder as particles grow. Intuitively, this is clear: meter-size rocks do not just 'stick'. Through $\mathcal{E}^*$ and γ the sticking threshold does depend on material properties. Ice (or ice-coated grains) are stickier than silicates, because they have a higher surface tension and are softer (lower $\mathcal{E}^*$, Table 1.2).

In the early stages when relative velocities are still low, dust coagulation is believed to lead to the formation of open structures – *aggregates* – where the individual grains are connected through their contacts. Examples are shown in Figure 1.12. Initially, when collisional energies are low, aggregates can stick with the creation of one (or two, three) new contacts, largely preserving their internal structure. This mode is referred to as *hit-and-stick* growth. Hit-and-stick growth obeys a fractal law:

$$\text{mass} \propto R_g^{D_f} \quad (1.29)$$

where R_g is a characteristic radius‡ of the aggregate and D_f the fractal dimension. When the internal density ($\rho_\bullet$) stays constant, $D_f = 3$. A linear chain is represented by $D_f \approx 1$ and a pancake geometry by $D_f \approx 2$. Hit-and-stick growth results in fractal dimensions of $D_f \approx 2$.

Exercise 1.8: Fractal growth can have dramatic consequences when the fractal dimension is low. Similar to the radius of the aggregate, we can define a fractal dimension for the evolution of the surface area A: mass $\propto A^{d_A}$. For homogeneous spheres, $d_A = 3/2$.

(a) Argue that the lowest possible value of d_A is 1.

(b) In that case, show that the stopping time is independent of particle size (or mass) as long as particles are in the Epstein regime. Consequently, drift-induced relative velocity does not increase with particle size.

Of course, the collisional energy does increase with particle growth and will at some point exceed the so-called rolling energy $E_{\rm roll}$. Similar to the breaking energy, this is a critical energy but now associated

* For example: vibrations (Chokshi et al. 1993) and visco-elasticity (Krijt et al. 2013).

† The impact or collisional energy is $\frac{1}{2}m_\mu(\Delta v)^2$.

Material	Elasticity $\mathcal{E}$ [10^{10} erg cm^{-3}]	Surface energy γ
Quartz	54	25
Water ice	7	370

Table 1.2: Material constants.

Figure 1.12: Simulations of aggregates formation by hit-and-stick processes. Top: a porous (but homogeneous: $D_f = 3$) aggregate. Bottom: an aggregate of low fractal dimension. From Seizinger et al. (2013).

‡ For fractals there are many definition of 'radius'. R_g is often taken to be the gyration radius.

with rolling of the contact area. Due to the rolling, aggregates will restructure (Dominik & Tielens 1997) – meaning compaction – increasing their stopping times, which results in ever more violent collisions. Ultimately, aggregates may compact to filling factors ≈ 0.1–1. How effective compaction operates is, however, still debated. The two main viewpoints are:

1. Compaction is effective enough for growth to stall at mm/cm sizes, at which stage sticking ceases.* Beyond this size, and with increasing relative velocities, particles are prone to fragmentation, replenishing the micron-sized grains, or particles simply bounce upon collision. As a result, a steady-state grain distribution will emerge, in which the largest particles contain most of the mass, but where the small grains are abundant enough to render the disk optically thick (at optical and near-IR wavelengths). In this scenario, kilometer-sized bodies (planetesimals) can form from a gravitational instability of a dense concentration of mm/cm-size particles (§1.5).
2. Collisional compaction is ineffective; although collisions surely cause restructuring, they will preserve a fractal relation between their mass and surface area. Consequently, the porosity of aggregates increases, facilitating growth. Kilometer-sized aggregates (still fluffy) form, until the point where they are ultimately compacted by gas drag (Kataoka et al. 2013). In this scenario, planetesimals form quickly through a continuous growth process (incremental growth).

* A key exception are collisions among particles of different size, as the sticking threshold is determined by the properties of the smallest particle.

1.5 *Collective effects and gravitational instabilities*

When growth by sticking fails, gravitational instabilities may yet result in the formation of planetesimals and larger bodies. A key requirement is that the governing equations of motions are unstable against (gravitational) perturbations. Suppose that the steady state solutions are given by $\rho_0, \mathbf{v}_0$, *etc.*. To analyze stability, one considers perturbations to these quantities of the form:

$$\delta\rho \propto \exp\left[i(\mathbf{k}\cdot\mathbf{r} - \omega t)\right], \tag{1.30}$$

where $\mathbf{k}$, the wave-number, is chosen real, whereas ω is a complex number. We then solve the original equations of motions by inserting $\rho = \rho_0 + \delta\rho$ *etc.*, employing that the perturbations are small: $\delta\rho \ll \rho_0$. The goal is to find a relation between ω and $\mathbf{k}$ – the *dispersion relation*. In particular, when $\omega(\mathbf{k}) = a + ib$ and b is positive, the perturbation associated with spatial frequency $\mathbf{k}$ will amplify with time. The disk is then unstable against perturbation on scale $1/\mathbf{k}$.

For thin disks (Σ instead of ρ) and axisymmetric perturbations ($\mathbf{k} = k_r = k$), the dispersion relation reads:†

$$\omega^2 = \kappa^2 - 2\pi G\Sigma k + k^2 c_s^2, \tag{1.31}$$

where $\kappa^2 = r(d\Omega^2/dr) + 4\Omega^2$ is the *epicyclic frequency* and c_s the sound speed. In Keplerian disks, $\kappa = \Omega_K$. Clearly, large κ and large

† See *e.g.* Binney & Tremaine (2008). The derivations is straightforward, but lengthy. A further approximation is that the resulting spiral-wave pattern is tightly wound – the so-called WKB-approximation.

c_s promote stability as ω becomes positive. Physically, pressure stabilizes small-scale (high k) perturbations, whereas rotation stabilizes large scales (small k). On the other hand, Σ causes ω^2 to become negative, especially at intermediate scales.

Exercise 1.9:

(a) From Equation (1.31), show that the scale most 'vulnerable' to instability is $\lambda_c = 2c_s^2/G\Sigma$ (where we switched to a spatial scale $\lambda = 2\pi/k$) and that instability is triggered when

$$Q_T \equiv \frac{c_s \Omega}{\pi G \Sigma} < 1, \tag{1.32}$$

where Q_T is the *Toomre-Q parameter*.

(b) A physically-intuitive way to obtain Q_T approximately is to compare the total internal, rotational, and gravitational energies. When

$$\left|\frac{E_{\rm therm}}{E_{\rm grav}}\right| \times \left|\frac{E_{\rm rot}}{E_{\rm grav}}\right| < 1 \tag{1.33}$$

the gravitational energy dominates over the combined rotational and thermal energies, leading to instability. Show that the above estimate results in Q_T, barring a factor of unity.

(c) Take a disk with a solar-mass star and, sound speed $c_s = 1$ km s$^{-1} \times r_{\rm AU}^{-1/4}$ and $\Sigma = 10^3$ g cm$^{-2} \times r_{\rm AU}^{-1}$. Where does the disk become unstable and what is the corresponding disk mass?

(d) What is the mass associated with the scale λ_c?

Disk instability – the gravitational instability of the gas disk is one of the two main models to form gas giant planets. It is invoked especially for explaining those planets seen far away from the star, for example HR8799 (Figure 1.13), because it is believed that to form a planet by core accretion (see Figure 1.18) takes too long. It should be noted, however, that $Q_T < 1$ is a necessary but not a sufficient condition for collapse. To proceed towards a planet, the fragment should be able to cool effectively to ensure that Q_T stays <1 *during* the collapse.

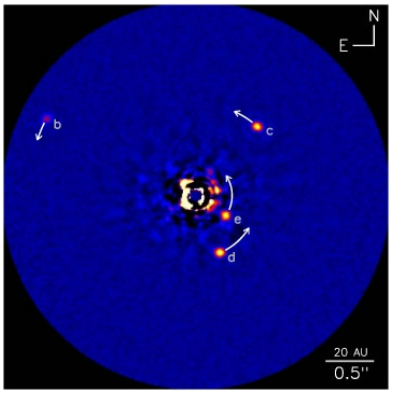

Figure 1.13: Four giant planets in the HR8799 system, observed by direct imaging (Marois et al. 2010).

Gaseous disk thus need to be massive and be able to cool quickly in order to spawn (giant) planets. However, it is also possible for the dust to collapse gravitationally, when it has sedimented into a thin sub-disk. Because dust forms a pressureless fluid, the thermal motions in Equation (1.31) can be taken 0: $c_s = 0$. This means that all scales less than $\lambda_c = 4\pi^2 G\Sigma_s/\Omega^2$ are unstable, where Σ_s is the surface density in solids. Clearly, the critical wavelength λ_c in the case where the dust collapses is much lower than the λ_c derived above for the gaseous case. The corresponding mass, $\lambda_c^2\Sigma_s$ already corresponds to planetesimals of radius $\sim$10 km (it can be lower due to fragmentation during the collapse). This (rather simple) idea of forming the first generation of planetesimals is known as the Goldreich & Ward (1973) mechanism, after its discovers.

A key requirement of the GW-model is, however, that particles settle into a thin layer of scaleheight $h_p < \lambda_c$, as otherwise the 2D assumption would not be justified. At 1 AU λ_c corresponds to approximately 0.1% of the gas scaleheight, implying that the dust sub-

disk will be dominated by solids ($\rho_p > \rho_g$). Because it now is the most dominant component by mass, it is the dust that start to dictate the (gas) dynamics. This is simply a consequence of Newton's third law as momentum conservation dictates that $\mathbf{F}_{\rm drag}\rho_p = -\mathbf{F}_{\rm back}\rho_g$ the "back-reaction" force $\mathbf{F}_{\rm back}$ – which we hitherto ignored – becomes important when the total dust density becomes large. Therefore, the motion of a single dust particle – and indeed that of the gas – can no longer be studied in isolation; it has become an *collective effect*.

Here, we follow Nakagawa et al. (1986), who solved for the steady solution of the dust *and* gas. Consider a frame rotating with the *local* Keplerian freqency $\Omega_K(r)$, such that the (outward) centrifugal force cancels the (inward) gravitational force. In such a frame the equations of motion for the gas ($\boldsymbol{u}$) and the dust ($\boldsymbol{v}$) read:*

$$\frac{D\boldsymbol{v}}{Dt} = -\frac{\boldsymbol{v}-\boldsymbol{u}}{t_{\rm stop}} - 2\boldsymbol{\Omega}_K \times \boldsymbol{v} + \boldsymbol{F}_{\rm Euler-dust} \tag{1.34a}$$

$$\frac{D\boldsymbol{u}}{Dt} = \frac{\rho_p}{\rho_g}\frac{\boldsymbol{v}-\boldsymbol{u}}{t_{\rm stop}} - 2\boldsymbol{\Omega}_K \times \boldsymbol{u} + \boldsymbol{F}_{\rm Euler-gas} + \boldsymbol{F}_{\rm pres} \tag{1.34b}$$

where the terms on the RHS are the gas drag, Coriolis, and Euler term, respectively. The latter is, like the Coriolis force, a fiducial force that enters the equation of motion because the angular velocity of the rotating frame depends on r: $\mathbf{F}_{\rm Euler} = -(d\boldsymbol{\Omega}_K/dt) \times \mathbf{r}$. The gas also experiences a pressure force of $F_{\rm pres} = 2\eta v_K \Omega_K$ in the radial direction. Setting the LHS o $(D/Dt = 0)$,† we can solve for $\mathbf{v}$ and $\mathbf{u}$. In the radial and azimuthal directions these read:

$$v_r = -\frac{2\tau_p}{\tau_p^2 + (1+Z)^2}\eta v_K \tag{1.35a}$$

$$v_\phi = -\frac{1+Z}{\tau_p^2 + (1+Z)^2}\eta v_K \tag{1.35b}$$

for the solids and

$$u_r = \frac{2Z\tau_p}{\tau_p^2 + (1+Z)^2}\eta v_K \tag{1.35c}$$

$$u_\phi = -\frac{1+Z+\tau_p^2}{\tau_p^2 + (1+Z)^2}\eta v_K \tag{1.35d}$$

for the gas, where $\tau_p = t_{\rm stop}\Omega_K$ is the dimensionless stopping time and $Z = \rho_p/\rho_g$ the "metallicity" of the gas.

Exercise 1.10: It is interesting to consider the limiting expressions of Equation (1.35). Verify that:

(a) $\tau_p \gg 1$ (big rocks): $v_r = u_r = v_\phi = 0$ and $u_\phi = -\eta v_K$.

(b) $Z = 0$ (negligible dust): solutions are the same as the individual solutions of Equations (1.11) and (1.12)

(c) $Z \gg 1$ (dust-dominated): $v_r = v_\phi = u_r = u_\phi = 0$

(d) $\tau_p \ll 1$ and $Z \ll 1$ (tracer dust): $v_r = u_r = 0$ and a reduced gas headwind velocity with an "effective" $\eta \to \eta/(1+Z)$.

Explain these limits physically.

Hence, when $Z \gg 1$ and $\tau_p \ll 1$ the dust forces the gas to move at Keplerian orbits. In Figure 1.14 this is illustrated for the azimuthal

* In these equations, we assume that the particles are of the same size and have a single stopping time $t_{\rm stop}$. The expressions by Nakagawa et al. (1986) can be generalized towards a size distribution of particles (Tanaka et al. 2005).

† this is an assumption. We wish to solve for steady state, meaning $\partial/\partial t = 0$ (velocities are time-independent). We therefore have also assumed that the gradient in the velocity disappears, which strictly is only valid when τ_p and ηv_K are constant. Nevertheless the neglected terms are second order in velocity and therefore small.

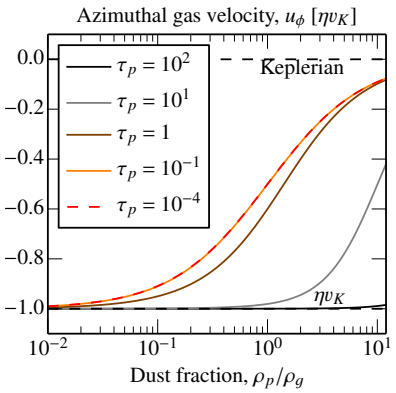

Figure 1.14: Azimuthal velocities of the gas (u_ϕ) with respect to Keplerian as function of the dust-to-gas ratio.

drift velocity of the gas (u_ϕ) as function of the dust-to-gas ratio. If the dust has sedimented to the midplane, a steep gradient of u_ϕ with height z will develop; the azimuthal velocity then ranges from $u_\phi \approx 0$ (Keplerian) to $u_\phi = -\eta v_K \sim 10$ m s^{-1} over a small range in heigh z. Such shear flows are not necessarily stable. Specifically, the *Kelvin-Helmholtz instability* will develop when the *Richardson number*

$$\mathrm{Ri} = \frac{-(g_z/\rho)(\partial\rho/\partial z)}{(\partial u_\phi/\partial z)^2} < \mathrm{Ri}_{\mathrm{crit}} \tag{1.36}$$

at some point z, where ρ can be taken as the dust+gas density. Typically, the critical Richardson number $\mathrm{Ri}_{\mathrm{crit}} \approx \frac{1}{4}$ (Chandrasekhar 1961). In Equation (1.36) the nominator is a measure of the (stabilizing) buoyancy, which causes lighter parts of the fluid to lie on top of denser parts (Archimedes principle). Buoyancy therefore promotes the density stratification.* Nevertheless, when the vertical shear $(\partial u_\phi/\partial z)$ is strong, instability will ensue.

* $N \equiv \sqrt{-(g_z/\rho)(\partial\rho/\partial z)}$ is known as the Brunt-Väisälä frequency.

Exercise 1.11: Give an order-of-magnitude expression for Ri, by evaluating Equation (1.36) at $z = h_p$, the particle scaleheight. Assume that the dust has settled such that dust dominates ρ up to $\sim h_p$. For g_z consider both the self-gravity limit (*i.e.* g_z is determined by the dust) and the stellar gravity limit ($g_z = g_{\star,z}$). Taking a critical Richardson number of $\mathrm{Ri}_{\mathrm{crit}} = \frac{1}{4}$, what is the smallest scaleheight into which the dust can settle? How does this height compare to the critical wavelength λ_c of the GW-model? (Note that $h_p < \lambda_c$ is required for the GW-instability.)

For these reasons, it will be difficult for small particles ($\tau_p \ll 1$) to undergo gravitational instability, unless their concentration is extremely high.

However, carrying out a linear stability analysis of Equation (1.34) it can be shown that the steady solutions (Eq. [1.35]) of the dust-gas mixture are unstable (Youdin & Goodman 2005; Jacquet et al. 2011). Generally, the amount of clumping is quite weak, but it results in spatial fluctuation of the drift velocities and pressure. The clumping accelerates greatly, however, when the particles approach $\tau_p \sim 0.1$ and when the dust-to-gas ratio in the gas is large (Johansen et al. 2009). In that case the *streaming instability* will turn nonlinear and amplify greatly. As clumps grow larger their drift stalls and they collect the inward-moving material. This phase of the instability can best be compared to how cyclists and migrating geese travel together to protect themselves from the headwind of the atmosphere†. The streaming instability is thus a robust mechanism capable of producing the first generation of planetesimals, provided that particles grow to aerodynamical sizes $\tau_p \gtrsim 10^{-2}$ and settle into a relatively thin layer. Exactly under which conditions the gas-dust mixture becomes prone to the (nonlinear) streaming instability and what the outcome of the instability is in terms of planetesimal sizes, are areas of active research.

† See homepage A. Johansen

1.6 *Runaway and Oligarchic Growth*

In the next phase of planet formation, we assume that planetesimal formation has been completed and that the bulk of the solid material of the disk resides in bodies of $\sim 1 - 10^2$ km in size. This situation is referrred to as the *planetesimal hypothesis*. To first order, such bodies move in Keplerian orbits ($\tau_p \gg 1$). How quick can these bodies merge to form a planet?

Exercise 1.12: Consider a test body of mass M immersed in a sea of smaller bodies of mass m. Assume that the M-body is on a circular orbit at semi-major axis a and orbital frequency Ω_K, while the m-bodies are in Kepler orbits with eccentricity e and inclinations $i \simeq e$.

(a) Ignoring (for the moment) gravitational focusing, show that the growth timescale of the M-body is:

$$t_{\rm growth} \equiv \frac{M}{dM/dt} \simeq \frac{R\rho_\bullet}{\Sigma_m \Omega_K} \qquad \text{(no focusing)} \tag{1.37}$$

where R is the radius corresponding to M, $\rho_\bullet$ the internal density of the material and Σ_m the surface density in m-bodies.

(b) How long would it take to form an Earth-mass planet at: 0.1, 1, and 10 AU?

(c) How much shorter are growth times in the 2D case ($i = 0$)?

For growth that relies on geometrical sweepup – with the collisional cross section $\sigma_{\rm col} \sim \pi R^2 \propto M^{2/3}$ and relative velocities constant – the mass e-folding time increases with mass M. This is referred to as *ordinary growth*: growth slows down with time.* *Gravitational focusing*, however, changes the picture.

Exercise 1.13 gravitational focusing: Consider a stationary massive body of mass M and radius R and a test particle. The test particle approaches the body at an impact parameter b at velocity v_∞. Show by invoking conservation of energy *and* angular momentum, that the largest impact parameter leading to collision is

$$b_{\rm col} = R\sqrt{1 + \left(\frac{v_{\rm esc}}{v_\infty}\right)^2} \equiv R\sqrt{1 + 2\Theta}, \tag{1.38}$$

where $\Theta = v_{\rm esc}^2 / 2v_\infty^2$ is the gravitational focusing factor and $v_{\rm esc} = 2GM/R$ the (surface) *escape velocity*. Θ is also known as the *Safronov number* after Victor Safronov, who pioneered many of the concepts discussed in this chapter (Safronov 1969).

Therefore, the collisional cross section scales superlinearly with the mass: $\sigma_{\rm col} \propto M^{4/3}$. When the relative velocities is constant the M-body experiences *runaway growth*, in which the growth timescale decreases. The timescale for runaway growth is therefore at most the initial collision timescale among the planetesimal bodies.

This make planet formation look easy. But there is one important catch: *collisionless encounters*. †

* In Figure 1.9 the constant kernel is an example of ordinary growth, the linear of *neutral growth* (constant $t_{\rm growth}$) and the multiplicative of *runaway growth* (decreasing $t_{\rm growth}$).

† The derivation in this exercise is for a *hyperbolic* (unbound) encounter. But, the equations of motions are identical to those of a bound Kepler orbit. The influence of the stellar force is also ignored.

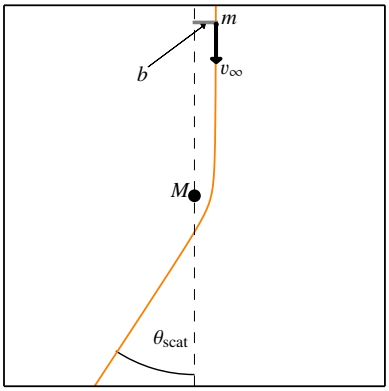

Figure 1.15: Scattering geometry

Exercise 1.14 gravitational scattering: Calculate the trajectory of a gravitational scattering. Consider polar coordinates (r, θ) with $\theta = 0$ the *direction* of the unperturbed velocity and $r(\theta = -\pi) = \infty$ initially (see Figure 1.15). Let primes denote derivatives towards θ;

(a) show that energy conservation implies $v_r v_r' + v_\theta v_\theta' + GMr'/r^2 = 0$ and that conservation of angular momentum gives $v_\theta = bv_\infty/r$ with b the impact parameter and v_∞ the initial velocity.

(b) Show that $v_r = bv_\infty r'/r^2$ and retrieve the following ODE for $r(\theta)$:

$$rr'' - 2(r')^2 - r^2 + r^3 \frac{b_{90}}{b^2} = 0 \tag{1.39}$$

where $b_{90} \equiv GM/v_\infty^2 (= \Theta R)$. This equation can be simplified by substituting $u = 1/r$:

$$u'' + u = \frac{b_{90}}{b^2}. \tag{1.40}$$

which has the solution $u = A\cos\theta + B\sin\theta + b_{90}/b^2$ where A and B are integration constants.

(c) Determine these to find:

$$r(\theta) = \frac{b}{b_{90}(1+\cos\theta)/b - \sin(\theta)}. \tag{1.41}$$

(d) Find the angle θ corresponding to the collisional focusing impact parameter of Equation (1.38), *i.e.* the location where the particle impacts the big body.

(e) Finally, derive the scattering angle – the direction the test particle is heading to after the scattering:

$$\theta_{\text{scat}} = \arcsin\left(\frac{2bb_{90}}{b^2 + b_{90}^2}\right) \tag{1.42}$$

and explain the meaning of "b_{90}".

The impact parameter b_{90} is therefore a measure for the range of close (collisionless) encounters. Since $b_{90} = \Theta R$ it increases faster with Θ than the collisional cross section (Eq. [1.38]), see Figure 1.16. Hence, for higher focusing factors (larger Θ) collisionless scatterings become more important relative to collisional encounters.

This is important because scatterings lead to an increase of the eccentricity of the population. To see this, consider the *guiding center approximation** for a Keplerian potential:

$$r \simeq a - ea\cos(\Omega_a \tilde{t}) + \mathcal{O}(e^2) \tag{1.43a}$$

$$\phi - \phi_0 \simeq \Omega_a \tilde{t} + 2e\sin(\Omega_a \tilde{t}) + \mathcal{O}(e^2) \tag{1.43b}$$

where e is the eccentricity, assumed $\ll 1$, a the semi-major axis, Ω_a the orbital frequency corresponding to a, t the time since periapsis passage†, and ϕ_0 a constant. Since $e \ll 1$ the terms of order e represents a small correction to the circular, uniform motion of the guiding center. Differentiating Equation (1.43) with respect to time gives the velocities in the guiding center approximation:

$$v_r \simeq ev_K \sin\Omega_a \tilde{t} \tag{1.44a}$$

$$v_\phi \simeq v_K + ev_K \cos\Omega_a \tilde{t} \tag{1.44b}$$

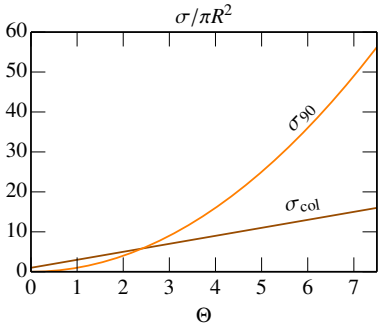

Figure 1.16: Collisional and scattering cross section

* See Exercise 2.1 and *e.g.* Binney & Tremaine (2008) for a derivation valid for general radial potentials, $\Phi(r)$.

† Hence, $\tilde{t} = 0$ corresponds to periapses. We could have written explicitly $\tilde{t} = t - t_0$ with t_0 the time of periapsis passage

again valid up to second order in eccentricity. The difference with respect to the circular motion, ev_K, is also known as the *random velocity* $v_{\rm ran}$. The motions are "random" because for a swarm of planetesimals the time of periapsis passage (t_0) all differ. Most naturally, these are uniformly distributed. Therefore, the total velocity is well approximated by a superposition of the constant guiding center motion v_K pointing in the azimuthal direction and $v_{\rm ran}$ pointing in an arbitrary direction.

In Equation (1.46) v_K is the Keplerian velocity corresponding to semi-major axis a: $v_K = \sqrt{Gm_\star/a}$. We can also express the motion with respect to the *local* Keplerian velocity at the instantaneous radius r *i.e.* $v_c = \sqrt{Gm_\star/r}$. Since $v_K \approx v_c - \frac{1}{2}ev_c \cos\Omega_a\tilde{t}$* the velocity difference with respect to the local Keplerian motion reads:

$$\delta v_r \simeq ev_c \sin\Omega_a\tilde{t} \tag{1.46a}$$

$$\delta v_\phi = v_\phi - v_c \simeq \frac{1}{2}ev_c \cos\Omega_a\tilde{t} \tag{1.46b}$$

Expressed with respect to the local Keplerian motion, the amplitude of the radial velocity is therefore a factor of two larger than the azimuthal, see Figure 1.17.

How can the eccentricity of a population increase if the gravitational scattering does not change the magnitude of the relative velocity vector? The reason is the asymmetry of the random velocity vector, as explained in Figure 1.17, which plots $\delta\mathbf{v}$ (Eq. [1.46]). For example, at periapsis 'P' ($M = 0$) the planetesimal is closest to the star and moving faster than the local Keplerian motion by $\delta v_\phi = \frac{1}{2}ea\Omega_a$, whereas at apoapsis 'A' ($M = \pi$) it is moving slower (negative δv_ϕ). Mid-way, at quadrature 'Q' the azimuthal excess motions are zero, but there is now a radial velocity as the body is moving towards either periapsis or apoapsis. As remarked, the radial velocity amplitude is twice that of the azimuthal velocities (Eq. [1.46]).

Now consider an interaction between the eccentric planetesimal and a (more massive) perturber on a circular orbit in the dispersion-dominated regime. In a scattering event, energy conservation does not alter the magnitude of the relative velocity magnitude, but it will change the orientation. See Figure 1.17 (middle and bottom panel), where the black dot indicates the point where the scattering occurs and the dashed circle the point in velocity space where it ends up after, say, a 90 degree scattering. The distances of the points to the center, corresponding to $|\delta\mathbf{v}|$ are the same. Therefore, a body experiencing a scattering at quadrature can find itself at apoapsis (or periapsis) and vice versa. Consider first the A/P→Q case. The particle ends up with the same $|\delta\mathbf{v}|$ it had before the scattering, but has its eccentricity decreased by a factor two as it is now at periapsis. Conversely, in the Q→A/P the eccentricity will double.

The two cases do not cancel. Scattering from/to quadrature are approximately equally likely; the average change in eccentricity is positive:

$$\langle\Delta e\rangle = \frac{1}{2}\left(-\frac{1}{2}e + e\right) = +\frac{1}{4}e. \tag{1.47}$$

* This follows from expanding the solution for the Kepler orbit (Eq. [2.1]) towards e:

$$\frac{r}{a} = \frac{(1-e^2)}{1+e\cos\nu} \simeq 1 - e\cos M + \mathcal{O}(e^2) \tag{1.45}$$

where $M = \tilde{t}\Omega_a$ is the mean anomaly and ν is the true anomaly.

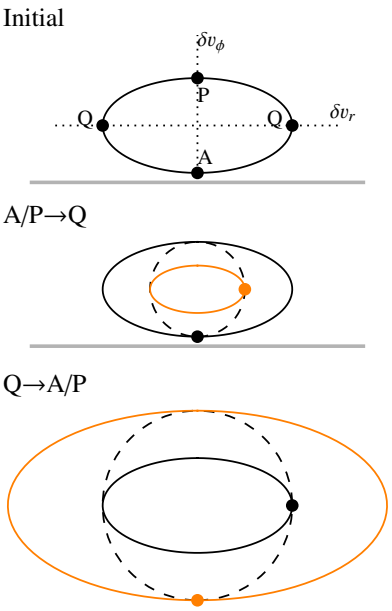

Figure 1.17: Viscous stirring. A close encounter will randomize the phase angle of a Kepler orbit, while preserving its magnitude (the eccentricity). Scattering at periapsis (P) or apoapsis (A) therefore tend to halve the eccentricity, whereas scatterings at quadrature (Q) increase the random velocity by a factor two. The net effect is positive and known as viscous stirring.

The net increase in eccentricity due to re-orientation of the phase angle is known as *viscous stirring*. Viscous stirring therefore increases the eccentricity of the population of planetesimals, which in turn decreases Θ. This negative feedback (the more massive the body, the lower Θ) lies at the heart of the transition from the runaway growth to oligarchic growth stage.

Interactions also equipartition the random energy (eccentricity) of the population. This means that the eccentricity of large bodies is kept much smaller than the low-mass planetesimals. This exchange of random energy between bodies of small and high mass is referred to as *dynamical friction*, and differs from viscous stirring as it does not change the total random energy. By the same reasoning, dynamical friction equilibrates the planer (eccentric) and vertical (inclination) motion. As there is one degree of freedom more in the plane, we can expect a 1:2 ratio between inclinations and eccentricity: $i/e \simeq 0.5$.

The negative feedback from viscous stirring causes runaway growth (one body dominates) to give way to *oligarchic growth*: growth ruled by multiple bodies of similar mass. How the transitions proceeds is illustrated in Figure 1.18. We consider a population of low-mass planetesimals and a couple of runaway bodies. During runaway growth, the most massive body is growing fastest as illustrated by the length of the arrows (top panel). Viscous stirring, however, excites the eccentricities of the planetesimals in its feeding zone (middle panel). Consequently, the growth of this embryo slows down. Neighbouring embryos are also affected when their feeding zone (partially) overlaps with a more massive body, suppressing their growth. But embryos that accrete from dynamically cold feeding zones still grow fast. Therefore, they catch up with neighboring embryos (lower panel). Hence, in oligarchy:

- $d(M_1/M_2)/dt > 0$ between embryos that share their feeding zone, where M_1 is the mass of the largest body and M_2 a runner-up.
- $d(M_1/M_2)/dt < 0$ between embryos of independent feeding zones.

As a result oligarchy will proceed towards a two component system with spatially separated and dynamically cold big bodies (embryos of mass M) and low mass planetesimals of mass m, that still dominate the total surface density Σ.

The emergence of a two-component system facilitates the development of analytical toy models that address how fast planet formation proceeds:

Exercise 1.15 oligarchy: Assume that the separation between embryos is a couple of mutual Hill radii, $\Delta a \sim \tilde{b} R_{\rm Hill}$. Then, it may be argued that the (effective) number density of embryos seen by a planetesimal is $n_M \simeq [(2\pi a) \times (\tilde{b} R_{\rm Hill}) \times (2ia)]^{-1}$.

(a) Argue that viscous stirring increases the eccentricity of the plan-

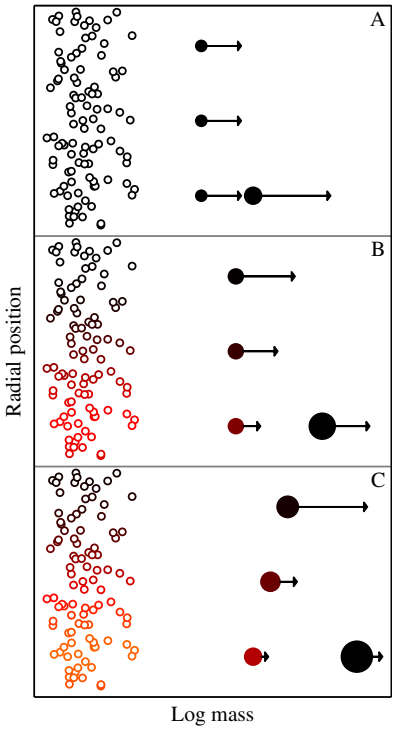

Figure 1.18: Emergence of oligarchy. Colors are a measure of the dynamical excitation (eccentricity). Arrows denote growth speed.

The Hill radius is defined:

$$R_{\rm Hill} = a \left(\frac{M}{3m_\star} \right)^{1/3} \simeq a q_p^{1/3} \qquad (1.48)$$

(Eq. [2.14])where $q_p = M/M_\star$.

etesimals at a rate:

$$\left.\frac{de}{dt}\right)_{\rm vs} \sim (n_M \sigma_{90} \Delta v)\Delta e \sim \frac{q_p^{5/3}\Omega_K}{e^3 \tilde{b}} \tag{1.49}$$

where $q_p = M/m_\star$. Gas friction, on the other hand, damps the eccentricity of the planetesimals. For planetesimals, the gas drag law depends quadratically on velocity/eccentricity. Let us therefore define:

$$\left.\frac{de}{dt}\right)_{\rm gas} \equiv -e^2 \Omega_K \mathcal{F}_{\rm aero} \tag{1.50}$$

where the dimensionless $\mathcal{F}_{\rm aero}$ depends on the aerodynamical properties of gas and planetesimals.

(b) Give the expression for $\mathcal{F}_{\rm aero}$. Balancing viscous stirring and gas damping gives therefore an *equilibrium eccentricity* of $e \sim q_p^{1/3}/(\tilde{b}\mathcal{F}_{\rm aero})^{1/5}$.

(c) Show that for these eccentricities, the Safronov numbers are constant. How far out in the disk can 10 Earth-size planets form within 10 Myr?

Oligarchy relies on the assumption that planetary embryos traverse on circular orbits so that they remain separated from each other. This assumption will break down when the embryos are no longer sufficiently damped, which happens when the gas in the disk disperses *and* when the total mass in embryos dominates that of the planetesimals. In that case the growth of embryos also stalls. The celebrated *isolation mass* is reached when the embryos have swallowed all the solids; *i.e.* when $M_{\rm iso} = (2\pi a) \times (\tilde{b} R_{\rm Hill}[M_{\rm iso}]) \times \Sigma$ or:

$$M_{\rm iso} = \frac{(2\pi\tilde{b}\Sigma r^2)^{3/2}}{(3M_\oplus)^{1/2}} \simeq 0.25\, M_\oplus \left(\frac{\tilde{b}}{10}\right)^{3/2} \left(\frac{\Sigma}{10}\right)^{-3/2} \left(\frac{a}{\rm AU}\right)^3 \tag{1.51}$$

For most disk models Σr^2 increases with r: the isolation mass is larger in the outer disk.

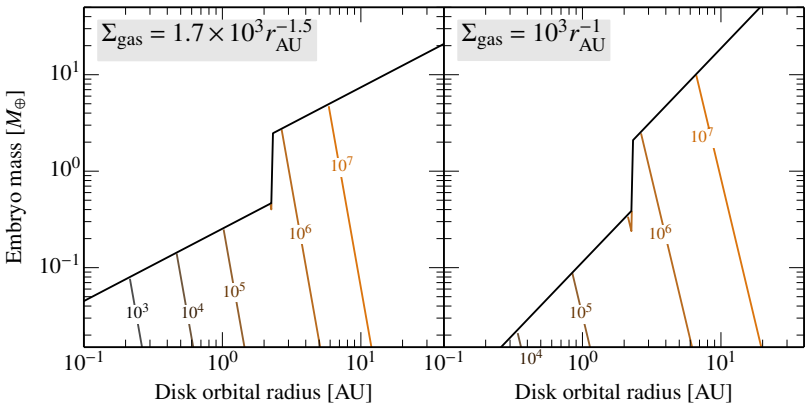

Figure 1.19: Embryo growth timescales (in years) during oligarchy for two disk surface density profiles: the MMSN (left) and a shallow $\propto r^{-1.5}$ profile. For the gravitational focusing factor, which enters the growth timescale, the equilibrium eccentricity as calculated in Exercise 1.15 has been used. Growth cannot proceed past the isolation mass (thick solid line). Note the presence of the ice-line at ≈3 AU, which raises the density in solids due to condensation of H_2O. Other fixed parameters are: a planetesimal radius of 10 km and a oligarchic spacing ($\tilde{b}$) of 10 Hill radius.

Figure 1.19 summarizes the outcome of the oligarchic model. Here, we show contours of the embryos growth timescales, $t_{\rm growth} = t_{\rm gr,geom}/(1+2\Theta)$ where Θ is calculated according to Exercise 1.15. Albeit it is fast, embryo growth in the inner solar system stalls at low masses. In the outer disk, growth timescales are much longer because of the reduced densities and the fact that gas damping is less effective. There,

embryos may not reach their isolation mass before the gas disk disperses. The region where oligarchic growth reaches the largest mass within the lifetime of the gas-rich disk ($\approx 10^7$ yr), is thus at ∼5–10 AU. An advantage of the oligarchic model is therefore that it explains the presence of giant planets in the "middle" regions, as we see in the solar system.

There are (at least) two criticisms of the oligarchic growth model as sketched above:

1. It assumes material doesn't move: planetesimals and embryos stay where they are. For planetesimals this may seem a plausible assumption, as their drift times are small but pebbles of course drift in from the outer disk regions
2. Fragmentation of planetesimals. The problem is that planetesimal bodies are relatively weak. With increasing embryo masses, eccentricities increase and collisions among planetesimals proceed at increasingly higher energies.* Especially small bodies are quite fragile. Therefore, planetesimals may grind themselves down to ∼1–10 m sizes – a size range which makes them very vulnerable to radial drift

In addition, while the oligarchic growth model broadly explains the architecture of the solar system, it may be less relevant to exoplanetary systems where the distinction between rocky planets and gas giants is (as of now) less clear.

* Consider the equilibrium eccentricity, calculated in Exercise 1.15. Corresponding impact velocities among planetesimales $\sim e_{\rm eq} v_K$ can reach ∼100 m s^{-1} – (for km-size planetesimals) much larger than their escape velocity.

1.7 *Pebble accretion*

Figure 1.19 and the arguments above cast some doubt on the viability of the planetesimal-driven planet formation scenario – particularly in the outer disk. Perhaps much smaller particles are more effectively accreted. In particular, protoplanetary disks are often bright at radio (∼mm) wavelengths, suggesting that a sizeable fraction of the solid mass resides in pebble-sized† particles.

Can these particles be accreted efficiently by a protoplanet? The key difference with the planetesimal encounters discussed in Exercise 1.14 is that for pebbles gas drag plays a decisive role *during the encounter*. This means that assumptions used above to calculate the scattering trajectory (conservation of energy and angular momentum) are no longer applicable. Rather, aerodynamical concepts now regulate the encounters. In particular, particles will settle towards the planet at a drift velocity $v_{\rm settl} = g_M t_{\rm stop}$.

Three timescales determin the fate of the interaction (see Figure 1.20):

- The stopping time $t_{\rm stop}$, which contains the aerodynamical properties of the pebble;
- The settling timescale: the time needed for a pebble to sediment to the embryo: $t_{\rm settl} \sim b/v_{\rm settl} = b/g_M t_{\rm stop} \sim b^3/GMt_{\rm stop}$;
- The encounter timescale: the duration of the (unperturbed) encounter: $t_{\rm enc} \sim 2b/v_\infty$.

† In this context "pebble" is a loose term for any particle larger than a dust grain and much smaller than a planetesimal (see Figure 1.1).

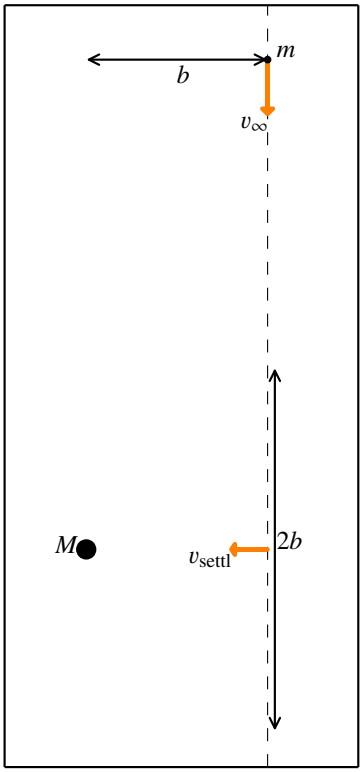

Figure 1.20: Collision geometry during pebble accretion. Here v_∞ is the approach velocity, b the impact parameter, and $v_{\rm settl} = g_M t_{\rm stop}$ the terminal settling velocity, which causes the particle to deviate from the unperturbed motion (dashed line).

Clearly, $t_{\rm enc} > t_{\rm stop}$ for gas drag to be effective during the encounter. Assuming this holds, we equate $t_{\rm settl}$ and $t_{\rm enc}$ to get the largest impact parameter b where encounters will result in accretion. We can distinguish two cases:

1. Small impact parameters. Relative motions are determined by the disk headwind ($v_\infty = \eta v_K$), resulting in $b_{\rm PA} \sim \sqrt{2GMt_{\rm stop}/\eta v_K}$.
2. Large impact parameters. Relative motions are determined by the Keplerian shear, * $v_\infty = \frac{3}{2}\Omega_K b$, resulting in $b_{\rm PA} \simeq \left(GMt_{\rm stop}/\Omega_K\right)^{1/3}$.

* See Exercise 2.3. When particles travel on circular orbits, they approach each other with a relative velocity governed by the (Keplerian) shear in the disk. This is known as the shear-dominated limit.

Exercise 1.16 pebble accretion:

(a) In contrast to gravitational scatterings, collision rates in pebble accretion are independent of the physical radius of the embryo (they only depend on the mass). Why?

(b) Considering the 2nd regime, show that $b_{\rm PA}$ expressed in terms of the embryo's Hill radius reads $b_{\rm PA} \simeq \tau_p^{1/3} R_{\rm Hill}$. This expression is valid for pebbles of dimensionless stopping time $\eta^3/q_p < \tau_p < 1$. Why?

(c) Correspondingly, the gravitational focusing factor for pebble accretion is: $\Theta_{\rm PA} = \tau_p^{2/3}(R_{\rm Hill}/R)^2$. Give an expression for $R_{\rm Hill}/R$ in terms of the disk orbital radius r.

(d) It is instructive to compare $\Theta_{\rm PA}$ to the corresponding expression of Safronov focusing, $\Theta_{\rm Saf} = v_{\rm esc}^2/2v_\infty^2$ (see Exercise 1.13). If we consider the zero eccentricity case (which maximizes $\Theta_{\rm Saf}$) planetesimals enter the Hill sphere at a typical velocity of $v_\infty \approx R_h\Omega_K$ (the Hill velocity). In that case, what is $\Theta_{\rm Saf}$? How much faster is PA compared to planetesimal accretion?

Apart from the very large accretion cross sections (up to the Hill radius!), pebble accretion does not suffer from the negative feedback effect of the planetesimal accretion model. It therefore virtually eradicates the timescale problem that characterizes the oligarchic model (see Figure 1.18), especially in the outer disk. Nevertheless, caveats remain:

- Pebble accretion is optimal for a relatively narrow particle size range, $10^{-2} \lesssim \tau_p \lesssim 1$ and also requires a sufficiently massive embryo.
- In a turbulent disk, pebbles will be distributed over a thick height (as calculated in Exercise 1.2), reducing the local pebble density in the vicinity of the embryo.
- Pebbles drift and (when they are not accreted) will move out of the embryo's feeding zone.

Exercise 1.17: When pebbles have sedimented to the midplane, the pebble accretion rate is $\dot{M} \simeq 2b_{\rm PA}\Sigma\eta v_K$. Assume that these pebbles drift to the star at a velocity given by Equation (1.11). Show that the fraction of pebbles that are accreted by the embryo is:

$$P \simeq \frac{1}{2\pi}\left(\frac{q_p}{\tau_p^2}\right)^{1/3} \qquad \text{for} \quad \eta^3/q_p \lesssim \tau_p \lesssim 1 \tag{1.52}$$

where $q_p = M/M_\odot$ (Hint: use $GM_p = q_p a_0^3 \Omega_K^2$).

1.8 *Gas accretion onto planets and giant planet formation*

The next phase of planet formation starts when embryos become massive enough to bind the gas of the disk. This occurs when the surface escape velocity $v_{\rm esc}$ equals the thermal speed of the gas molecules. Ignoring numerical factors of unity, atmosphere formation starts when the protoplanet's *Bondi radius*:

$$R_{\rm Bondi} = \frac{GM}{c_s^2} \tag{1.53}$$

becomes larger than the physical radius (R). In Equation (1.53) M is the combined mass of the "rocky" core (M_c) and gaseous atmosphere ($M_{\rm atm}$), but initially $M \approx M_c$. From Equation (1.53) atmosphere formation thus commences at $R \approx 10^3$ km slightly depending on the temperature of the gas.

From Equation (1.53) it is clear that with further growth of the propoplanet, $R_{\rm Bondi}$ quickly becomes much larger than R. Consequently, the planet can acquire a – by volume – huge atmosphere. But how massive is it?

Exercise 1.18 isothermal atmosphere: consider an isothermal atmosphere where the temperature is the same as that of the protoplanetary nebula.

(a) Invoking hydrostatic balance and neglecting for the moment the contribution of the atmosphere mass towards the gravitational potential show that the density of the gas is given by:

$$\rho(r) = \rho_{\rm neb} \exp\left[\frac{R_{\rm Bondi}}{r}\right], \tag{1.54}$$

where $\rho_{\rm neb}$ is the gas density in the nebula and r the distance from the planet's center. The pressure scaleheight corresponding to Equation (1.54) is $H_r = r^2/R_{\rm Bondi}$.

(b) Argue that for $r/R_{\rm Bondi} \ll 1$ most of the mass of the atmosphere resides near the core radius R_c and that the atmosphere mass can be approximated as

$$M_{\rm atm} \approx 4\pi\rho_{\rm neb} R_{\rm Bondi}^3 \left(\frac{R_c}{R_{\rm Bondi}}\right)^4 \exp\left[\frac{R_{\rm Bondi}}{R_c}\right] \tag{1.55}$$

(c) Also, give an expression of the atmosphere-to-core mass ratio, in terms of the density of the nebula-to-core $\rho_{\rm neb}/\rho_c$. As function of disk orbital radius, how massive should planets become before their isothermal atmospheres reach parity ($M_{\rm atm} = M_c$)?

Figure 1.21 plots the atmosphere-to-core mass ratio obtained from the exact atmosphere mass and from Equation (1.55). Clearly, the atmosphere mass under the isothermal assumption increases steeply with $R_{\rm Bondi}/R_c$ (appreciate again that the Bondi radius is proportional to core mass!). What stops the growth of the atmosphere? Perhaps:

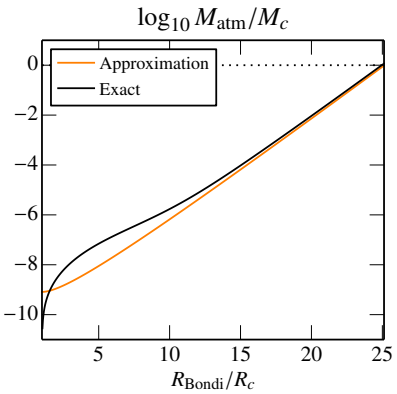

Figure 1.21: Atmosphere-mass-to core ratio as function of $R_{\rm Bondi}/R_c$ for a density ratio of $\rho_{\rm neb}/\rho_c = 10^{-10}$.

- The atmosphere does not extend all the way to $R_{\rm Bondi}$. Certainly at $r = R_{\rm Hill}$ (stellar) tidal forces become important and it is inconceivable that planets can bind material beyond $R_{\rm Hill}$. This will limit the atmosphere size for large planets, once $R_{\rm Bondi}$ has exceeded $R_{\rm Hill}$. Nevertheless, as $R_{\rm Hill}/R \gg 1$ protoplanetary atmospheres that fill the Hill sphere are still enormous.
- Centrifugal forces (rotation) can becomes important. When the atmosphere is supported by rotation, it no longer requires such steep density gradient.
- The atmosphere is not isothermal. Nebular gas falling (adiabatically) into the potential heats up. The high temperature gas can more effectively provide the pressure gradients necessary to balance the gravitational force.

Exercise 1.19 adiabatic atmosphere: assume an adiabatic equation of state

$$P = K\rho^\gamma \tag{1.56}$$

with K constant and $\gamma > 1$. (Note that $\gamma = 1$ corresponds to the isothermal case above. Why?). Repeating the above exercise, derive:

$$\rho = \rho_{\rm neb}\left[1 + \frac{(\gamma-1)R_{\rm Bondi}}{r}\right]^{1/(\gamma-1)} \tag{1.57}$$

and demonstrate that $M_{\rm atm}$ never exceeds $\sim\rho_{\rm neb}R_{\rm Bondi}^3$ as long as $\gamma > \frac{4}{3}$.

The exercises above illustrate that the fate of the planet's primordial atmospheres is primarily determined by its thermal evolution. Of course an isothermal or adiabatic EOS are merely approximations. In essence, the problem is identical to solving the well-known stellar structure equations: *

$$\frac{\partial M_{<r}}{\partial r} = 4\pi G r^2 \rho \tag{1.58a}$$

$$\frac{\partial P}{\partial r} = -\rho\frac{GM_{<r}}{r^2} \tag{1.58b}$$

$$\frac{\partial T}{\partial r} = \frac{\partial P}{\partial r}\frac{T}{P}\nabla \tag{1.58c}$$

$$\frac{\partial L}{\partial r} = 4\pi r^2\rho\left(\epsilon + T\frac{dS}{dt}\right). \tag{1.58d}$$

In stars ϵ is given by nuclear fusion, but in planets a similar role is played by the accretion of solid material (planetesimals), if present.† An *equation of state* between P, S, ρ and T completes the above equations‡ Finally, the actual gradient ∇ is often taken to be the minimum of the radiative and adiabatic temperature gradients:

$$\nabla = \min(\nabla_{\rm ad}, \nabla_{\rm rad}) \tag{1.59}$$

(*Schwarzschild criterion*) where $\nabla_{\rm rad} = 3\kappa LP/64\pi\sigma_{\rm sb}GM_{<r}T^4$ is the gradient in the radiative case, with $\sigma_{\rm sb}$ Stefan-Boltzmann constant.

* symbols denote: $M_{<r}$ total mass inside radius r; $\nabla \equiv d\log T/d\log P$ actual temperature gradient; S entropy; ϵ energy production rate.

† If they make it to the core, they will provide an accretion luminosity of $L_{\rm acc} \approx \frac{GM_c}{2R_c}$ per unit mass.

‡ The simplest is the ideal EOS, $P = k_B\rho T/\mu$, where k_B is Boltzmann's constant and μ the mean molecular mass. This can be expanded by including phase transitions (dissociation, ionization) of, *e.g.*, hydrogen. At very high densities, non-ideal effects will result in a non-ideal equation of state (*e.g.* the metallic interiors of the giant planets).

Exercise 1.20 Radiative zero solution: Assume the following:

- κ and L are constant and define $W \equiv 3\kappa L/64\pi\sigma_{\rm sb}$ (also constant);
- The gravitational mass interior to r, $GM_{<r}$ can be approximated by the total mass of the planet+atmosphere, $GM_{\rm tot}$, which is a constant;
- The atmosphere is radiatively supported: $\nabla = \nabla_{\rm rad} = WP/GM_rT^4$;
- An ideal EOS, $P = k_B\rho T/\mu$.

(a) Under these assumptions, show that Equation (1.58c) gives:

$$P = \frac{GM_{\rm tot}T^4}{4W}; \qquad \rho = \frac{GM_{\rm tot}\mu T^3}{4k_BW} \tag{1.60}$$

where we neglected the boundary condition (the solutions are valid only in the "deep" atmosphere).

(b) Continue, by invoking Equations (1.58b) and (1.58c), to derive the atmosphere temperature and density profiles:

$$T(r) \simeq \frac{GM\mu}{4k_B}\frac{1}{r}; \qquad \rho(r) \simeq \frac{1}{W}\left(\frac{GM\mu}{4k_B}\right)^4\frac{1}{r^3}. \tag{1.61}$$

Integrating these gives the mass of the atmosphere:

$$M_{\rm atm} = \frac{4\pi}{W}\left(\frac{GM_{\rm tot}\mu}{4k_B}\right)^4\Lambda, \tag{1.62}$$

where $\Lambda = \log(r_{\rm out}/r_c)$ is a function of the outer radius of the atmosphere. Note that due to the logarithmic dependence, Λ can be considered a constant.

(c) Equation (1.62) can be recast in a relation between the core mass $M_c = M_{\rm tot} - M_{\rm atm}$ and the total mass $M_{\rm tot}$, see Figure 1.22. Clearly, for small mass $M_c \approx M_{\rm tot}$. However, at some point $M_c(M_{\rm tot})$ is seen to peak. Give an expression for the core mass $M_{\rm crit}$, corresponding to the peak. What is atmosphere mass fraction at this point?

(d) Give a physical explanation for the behaviour.

Although this exercise is still very crude, it contains the key features of the general problem. Specifically:

1. The atmosphere mass $M_{\rm atm}$ is a strong function of the (grain) opacity of the gas and of the planetesimal accretion rate. Paradoxically, a high accretion rate – a high L – decreases the atmosphere mass.*
2. When $M_{\rm atm} \approx M_c$ (crossover) a hydrostatic solution is no longer possible. The corresponding core mass is referred to as the *critical core mass*.

For pre-planetary atmospheres that have breached the critical core mass, pressure gradients are insufficient against gravitational forces. Such planets can accrete unlimited quantities of gas from the circumstellar disk; more precisely, the amount of gas they accrete is fully determined by the supply of gas the disk can deliver. High-mass planets will open gaps in the disk (§2.3), thereby halting (to some extent) the accretion of nebular gas.†

After the primordial atmosphere has collapsed, gas moves in at free-fall velocities or is funnelled through a circumplanetary disk. Either way, the material that (eventually) falls onto the planet radiates away a large amount of energy, decreasing the entropy of

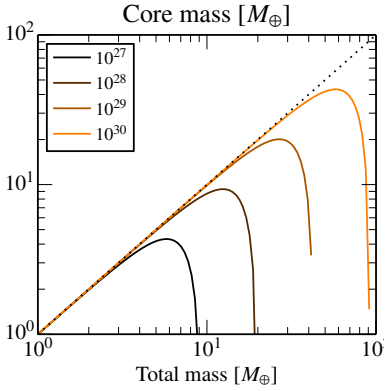

Figure 1.22: Relation between the core mass and total mass for several values of W (labelled). See Exercise 1.20 for the definition of W.

* When there no planetesimals left to accrete, a luminosity is nonetheless generated by *Kelvin-Helmholtz contraction*, in a similar way as stars evolve before the onset of nuclear fusion.

† How much gas can be accreted and the mechanisms that halt gas accretion are still active research fields.

the gas. This contrast with the giant planet formation in the disk-instability scenario, where a self-gravitating clump adiabatically collapses, largely preserving the entropy. In due time, these clumps do also cool, but without the shocks that accompany gas accretion of the core accretion model it takes them much longer. These different entropy states are referred to as the *cold start* and *hot start* and are loosely connected to the *core accretion* and *disk instability* scenarios.

1.9 Giant impacts

In Figure 1.18 we saw that planetary embryos in the inner disk stayed small because of the low isolation masses – typically around the mass of Mars. These bodies are tightly packed with mutual spacings of ∼0.01 AU. The embryos gravitationally perturb each other, exciting their eccentricity. But in a gas-rich environment this effect is suppressed because of efficient eccentricity damping due to disk-planet interactions (see §2.3). Embryos therefore remain on circular orbit, isolated from each other.

The situation changes once the gas has disappeared. Secular* perturbations gradually increase the eccentricity. At some point orbits start to cross, meaning that the apoapsis of the interior body exceeds the periapsis of the exterior body. Once *orbit crossing* has been achieved embryos have a chance to collide with each other – a period known as the *giant impact* phase of (terrestrial) planet formation. This reduces the number of bodies in a system. The process continues until the orbit crossing timescale has become longer than the system timescale.†

The situation can best be investigated numerically, through N-body simulations. Many studies have found that the *crossing time* depends exponentially on the mutual spacing among the embryos:

$$\log_{10} \frac{t_{\rm cross}}{t_{\rm K}} = b\Delta + c \tag{1.63}$$

where Δ is the distance among the embryos expressed in mutual Hill radii‡, t_K the orbital period and b and c are constants. N-body simulations find $b \approx 0.76$–0.8 and $c \approx -(0.2$–$0.5)$ (Chambers et al. 1996; Iwasaki et al. 2001; Zhou et al. 2007). For example, when $\Delta \approx 10$, the initial configuration of planetary embryos will become unstable after $\sim 10^7$ orbits.

Exercise 1.21: How fast do embryos collide, once they are on crossing orbits?

(a) Verify that:

- in the planar case (zero inclination), where the embryo's orbits intersect at two points, the probability of a collision during the orbit is $\approx R_{12}/\pi r$, where $R_{12} = R_1 + R_2$ is the linear cross section.
- in the inclined case (inclination i) the corresponding 3D collision probability amounts to $P \sim \sigma_{\rm geo}/ia^2$.

(b) For spacings of $\Delta \approx 10$ mutual Hill radii and $0.1 M_\oplus$ embryos at $r \approx 1$ AU, give the expression for the collision timescale in the geometrical

* A secular process means that the process takes place on a timescale (much) longer than the dynamical time.

† An exception is the Pluto-Neptune system, where orbits do cross (Pluto moves interior to Neptune). Here, orbit crossing does not result in close encounter, however, as the planets are in a 3:2 resonance.

‡ defined as:

$$R_{\rm Hill} = \frac{a_1 + a_2}{2} \left(\frac{M_1 + M_2}{3 M_\odot} \right)^{1/3}. \tag{1.64}$$

limit.

(c) Same, but allowing for gravitational focusing.

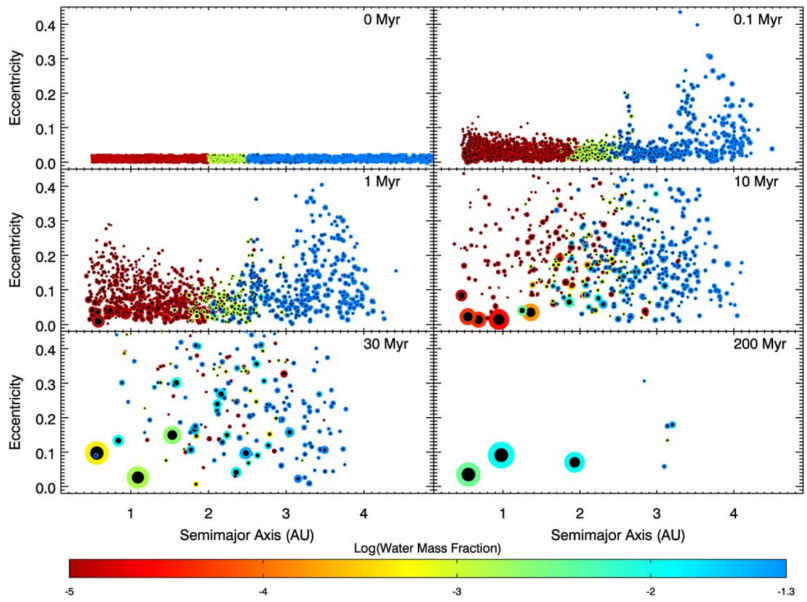

Figure 1.23: Outcome of an N-body simulation on how planet formation completes in the inner disk. Initially ≈1900 embryos are present of mass $\approx 10^{-3}$–10^{-2} $M_\oplus$. Colors indicate the water fraction. A Jupiter-mass giant planet resides at 5.5 AU. From Raymond et al. (2006).

Figure 1.23 presents an example of a computer simulation on how the giant impact phase may proceed. These studies are usually conducted using N-body techniques, in which the equations of motion of each of the N particles are integrated. The advantage of this approach is that the compositions of the material can be followed directly; in Figure 1.23 this is done for the water contents. Here, the initial dynamically quiescent system ($e \approx 0$) becomes unstable, quickly reaching high eccentricities of ≈0.1.* Due to the high density of the system – the initial mutual distance is far less than $\Delta = 10$ in this particular simulation – material efficiently scatters from the outer to the inner regions and vice versa.

On long timescales (>100 Myr) a configuration similar to that in the solar system emerges. Note the size of the Mars analogue at 2 AU, which is too large compared to Mars. This is a very common feature of these types of simulations: starting from a broad initial distribution of embryos (as predicted by the oligarchic growth model), the Mars-size body is too large. Simulations in which the initial populations of embryos are clustered together in a band centered at ≈1 AU seem to give a better result concerning the *small Mars problem*.

* Orbit crossing is accelerated by the (unseen) presence of a Jupiter at 5.5, which also perturbs the bodies, especially at resonances.

2
Planet dynamics

2.1 *Review of the 2-body problem*

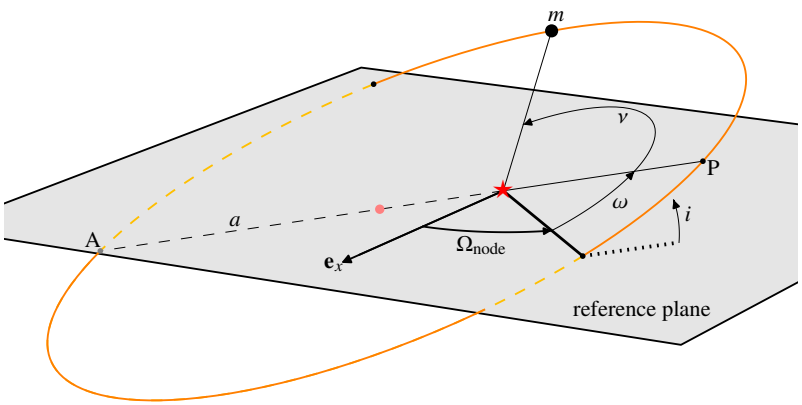

Figure 2.1: Orientation of a Kepler orbit in space (orange). Lines above the $z = 0$ reference plane are solid; below are dashed, while lines in the reference plane are bold. The five angles are: longitude of the ascending node Ω_{node}; argument of periapsis ω; true anomaly ν; inclination i. Points on the major axis of the ellipse show the periapsis (P), apoapsis (A), the focus ($\star$), and center ($\bullet$) of the ellipse.

In the 2-body problem the relative motion can be solved exactly. There are two conserved integrals of motion: the angular momentum vector* $\boldsymbol{l} = \boldsymbol{r} \times \dot{\boldsymbol{r}}$ and the total energy $E = \frac{1}{2}\dot{r}^2 - G(m_\star + m)/r$ where $m_\star$ is the mass of the central body and m the mass of the satellite. Solving the equation of motion in the orbital plane gives the solution:†

$$r = \frac{a(1-e^2)}{1+e\cos\nu}; \qquad r^2\dot{\nu} = l = |\boldsymbol{l}| = \sqrt{G(m_\star+m)a(1-e^2)} \tag{2.1}$$

with integration constants:

- The semi-major axis, a, which defines the energy of the system, $E = -G(m_\star + m)/2a$;
- The eccentricity e;
- The true anomaly ν (not constant); the angle in the orbital plane between periapsis and the position of the body (Figure 2.1). The true anomaly is not a linear function of time as the particle moves faster at periapsis than at apoapsis.

In 3D, the orbital plane is in general tilted with respect to a (fixed) reference plane – say the XY-plane – defining a reference direction – say the unit vector $\boldsymbol{e}_x$. The motion in space is supplemented by the following constants:

* From which Kepler's law follows: $r^2\dot{\theta}$ is constant where θ is the angle in the orbital plane.

† The equation of motion is the same as in Exc. 1.14. The integration constants differ.

- The inclination i, which gives the tilt of the orbital planet with respect to a plane of reference;
- The longitude of the ascending node, Ω_{node}, which denotes the angle in the reference plane between a reference direction ($\boldsymbol{e}_x$) and the point where the orbit ascends through the reference plane;
- The argument of periapsis ω, indicating the angle between the periapsis and the line of nodes.

Together $\{a, e, i, \omega, \Omega_{\text{node}}, \nu\}$ constitute the *orbital elements*. Together with the time of periapsis passage (t_0), they fully specify the Kepler orbit.

Other auxiliary angles can be construed from the orbtial elements. The longitude of periapsis, $\varpi = \omega + \Omega_{\text{node}}$ is often used instead of ω.* Also, instead of ν, which is a nonlinear function of time (ν proceeds faster at periapsis than at apoapsis) the position is expressed in terms of a mean anomaly M, which proceeds linear with time: $M = n(t - t_0)$ where $n = \sqrt{G(m_1 + m_2)/a^3}$ is the *mean motion*†. See Figure 2.2. Note that M has no geometrical interpretation; it is introduced merely as a mathematical convenience. Similarly, the mean longitude $\lambda = M + \varpi$ measures the advance of the particle in a fashion linear with time with respect to the reference frame.

Several concluding remarks on the 2-body problem:

- the Kepler solution has been expressed in a non-inertial coordinate frame (a consequence of using relative coordinates).‡
- except for ν, the orbital elements are constant in time, which is specific to the 2-body problem. For more than two bodies the mutual gravitational perturbations will cause the orbital elements to vary in time. They are then referred to as *osculating orbital elements*.
- there is no closed-form solution between time (or mean anomaly M) and the position in the orbit (the true anomaly ν). The relation can be found numerically, through *Kepler's equation*.§

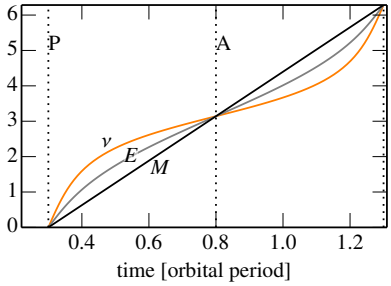

Figure 2.2: True (ν), eccentric (E), and mean anomaly (M) as function of time for an eccentricity of $e = 0.5$.

* In particular for circular (planar) motion where ω (Ω_{node}) is undefined.

† The mean motion is the same as the orbital frequency Ω in Chapter 1. (We use the notation n for convention).

‡ In a barycentric system, the total angular momentum is $L = m_1 l_1 + m_2 l_2$ and the total energy is $E = -Gm_1 m_2/2a$.

§ the relation proceeds through an intermediate quantity known as the eccentric anomaly E, defined as $r = a(1 - e\cos E)$. Kepler's equation then relates E and M:

$$M = E - e\sin E \tag{2.2}$$

Given M (*i.e.* time), this transcendental equation must be solved numerically for E, which gives r and hence ν via Equation (2.1).

Exercise 2.1 Guiding center:

(a) Consider two bodies in Kepler orbits separated by Δa in semimajor axis where $\Delta a \ll a$ and a is the semimajor axis of one of the bodies. Show that the *synodical period*, which is the time between successive conjunctions (close encounters), is

$$P_{\text{syn}} = \frac{2P}{3}\left(\frac{a}{\Delta a}\right) \tag{2.3}$$

where P is the orbital period corresponding to a.

(b) Show that for $e \ll 1$ the equations of motions (Eq. [2.1]) can be approximated:

$$r - a \simeq -ae\cos(M) \tag{2.4a}$$
$$\nu - M \simeq 2ae\sin(M) \tag{2.4b}$$

which is the *guiding center approximation*. The Keplerian motion is approximated by a superposition of a circle and an ellipse.

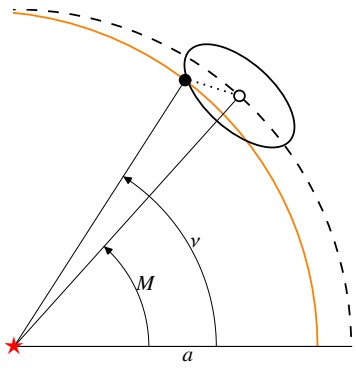

Figure 2.3: Epicycle approximation.

2.2 *The 3-body problem*

In the 3-body problem analytical (closed-form) solutions are no longer possible. A simplification of the 3-body problem is that of a massless particle being perturbed by a secondary (*e.g.* planet) that moves on a circular orbit around the primary (star). This is known as the circular, restricted 3-body problem *CR3BP*. We will focus exclusively on this problem.

The equation of motion in a frame of reference rotating with angular frequency ω is:

$$\ddot{\boldsymbol{r}} = -\nabla\Phi - 2\boldsymbol{\omega}\times\dot{\boldsymbol{r}} - \boldsymbol{\omega}\times(\boldsymbol{\omega}\times\boldsymbol{r}) \tag{2.5}$$

In the CR3BP we will of course choose $\omega = n_p\boldsymbol{e}_z$ such that Φ – the gravitational potential – is time-independent in the rotating frame. Equation (2.5) can be integrated to give an integration constant J:*

$$J = \frac{1}{2}\dot{\boldsymbol{r}}^2 + \Phi - \frac{1}{2}(\boldsymbol{\omega}\times\boldsymbol{r})^2 \tag{2.6}$$

which is the *Jacobi energy*. In the 3-body problem it is the only integral of motion.

Exercise 2.2 Jacobi integral:

(a) Converting Equation (2.6) back to the inertial frame, show that:

$$J = E - \boldsymbol{\omega}\cdot\boldsymbol{l} = E - n_p l_z \tag{2.7}$$

where E and $\boldsymbol{l}$ are the energy and angular momentum measured in the inertial frame. Hence, in the CR3BP interactions will exchange E and $\boldsymbol{l}$, while J stays constant.

(b) Express J in orbital elements:

$$J = -\frac{Gm_\star}{2a} - n_p\sqrt{Gm_\star(1-e^2)a}\cos i \tag{2.8}$$

where n_p is the mean motion of the secondary and the other symbols refer to the test particle. Written in the form of Equation (2.8) (or analogous) the Jacobi integral is called the *Tisserand relation*.

(c) Let $a = a_p + b$ with a_p the semimajor axis corresponding to n_p and consider the limits where $b/a_0 \ll 1$, $i \ll 1$ and $e \ll 1$. Show that in that case:

$$J \approx \frac{Gm_\star}{a_p}\left(-\frac{3}{8}\frac{b^2}{a_p^2} + \frac{e^2+i^2}{2}\right) \tag{2.9}$$

where we have discarded a constant term from J.

It is instructive to redefine the potential in Equation (2.5), incorporating the centrifugal term:†

$$\Phi_{\rm eff} \equiv \Phi_1 + \Phi_2 - \frac{1}{2}n_p^2 r^2 = -\left[\frac{Gm_\star}{r_1} + \frac{1}{2}n_p^2 r_1^2\right] + \Phi_2 \tag{2.10}$$

where we used the identity $\frac{1}{2}\nabla r^2 = \boldsymbol{r}$. Consider the motion of the test particle in the vicinity of m_2, see Figure 2.4, and express the potential in local coordinates (x, y) centered on m_2. This amounts to expanding the inverse distance $1/r_1$ in terms of (the small) x and y.‡

* To see this, multiply Equation (2.5) by $\dot{\boldsymbol{r}}$ and write all terms as time-differentials (d/dt): $\frac{d}{dt}(\dot{\boldsymbol{r}}^2/2) = \ddot{\boldsymbol{r}}\cdot\dot{\boldsymbol{r}}$, $\frac{d\Phi}{dt} = \dot{\boldsymbol{r}}\cdot\nabla\Phi$, and $\frac{d}{dt}(\boldsymbol{\omega}\times\boldsymbol{r})^2/2 = [\boldsymbol{\omega}\times\boldsymbol{\omega}\times\boldsymbol{r}]\cdot\dot{\boldsymbol{r}}$. Also, $\dot{\boldsymbol{r}}\cdot(2\boldsymbol{\omega}\times\dot{\boldsymbol{r}}) = 0$

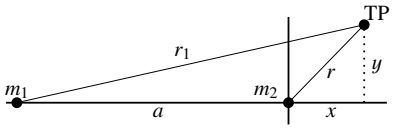

Figure 2.4: Definitions of x and y in Hill's approximation of the CR3BP.

† In celestial mechanics text books it is customary to define $\Phi_{\rm eff}$ with the opposite sign.

‡ this becomes $r_1^{-1} \approx 1/a_2 - x/a_2^2 + x^2/a_2^3 - \frac{1}{2}y^2/a_2^3 - \frac{1}{2}z^2/a_2^3$. The leading (constant) term of the expansion can be discarded from $\Phi_{\rm eff}$ as it is a potential function. In addition, in Equation (2.11) we assumed that $m_1 \gg m_2$ such that $Gm_1 \approx n_2^2 a_2^3$.

The result is (Hill's approximation):

$$\Phi_{\text{eff}} = -\frac{3}{2}n_2^2x^2 + \frac{1}{2}n_2^2z^2 - \frac{Gm_2}{r} \tag{2.11}$$

with which the Jacobi energy is written:

$$J = \frac{1}{2}\dot{r}^2 + \Phi_{\text{eff}} \tag{2.12}$$

Contours of $\Phi_{\text{eff}}(x, y)$ are known as *zero velocity curves*; they define the region where a particle of a certain J can move, since $\Phi_{\text{eff}} = J - \frac{1}{2}\dot{r}^2 \leq J$. Therefore, although the 3-body problem is not integrable, given J, we can constrain the regions where particles can be found. Figure 2.5 shows contours of constant Φ_{eff} with lighter contours having larger Φ_{eff}. The regions bounded by high Φ_{eff} (the darker contours) are therefore not accessible for low-energy particles (low J). In particular, the high Φ_{eff} zero velocity curves have a horseshoe shape and the corresponding orbits are referred to as *horseshoe orbits* as they make a U-turn. It must be emphasized however that in general particles do not follow the zero velocity contours as $\dot{\boldsymbol{r}}$ is a function of time. Figure 2.6 gives examples of particle trajectories obtained from integrating Hill's equation of motion. Three types of orbits can be seen:

- *Horseshoe orbits*, which make a U-turn (impact parameter $b \lesssim 1.7R_{\text{Hill}}$;
- *Hill-penetrating orbits*. They are strongly excited after they leave the Hill sphere ($1.7R_{\text{Hill}} \lesssim b \lesssim 2.5R_{\text{Hill}}$);
- *Circulating orbits*, which are only modestly excited. ($b \gtrsim 2.5R_{\text{Hill}}$).

Exercise 2.3 Hill's equations:

(a) Show that the equations of motion in Hill's approximation are:

$$\ddot{x} = -\frac{Gm_p}{r^3}x + 2n_p v_y + 3n_p^2 x \tag{2.13a}$$

$$\ddot{y} = -\frac{Gm_p}{r^3}y - 2n_p v_x \tag{2.13b}$$

where $r^2 = x^2 + y^2$ if we restrict the motion to the orbital plane.

(b) Show that zero eccentricity particles at distances far from the secondary obey $v_y = -\frac{3}{2}n_p x$ and $v_x = 0$. This (local) approximation of the Keplerian flow is known as the *shearing sheet*.

(c) Equilibrium points are points where $\ddot{\boldsymbol{r}} = \dot{\boldsymbol{r}} = 0$. Show that these *Lagrange points* are located at $(x, y) = (\pm R_{\text{Hill}}, 0)$ where R_{Hill} is the *Hill radius*:

$$R_{\text{Hill}} = a_p \left(\frac{m_p}{3m_\star}\right)^{1/3} \tag{2.14}$$

(d) Are these stable or unstable equilibrium points?

(e) What is the Jacobi constant at the Lagrange point (J_L)? And what is the Jacobi constant far from the perturber (J_∞), assuming $e = 0$. What is the half-width x_{hs} of the corresponding horseshoe orbit?

From this section it is clear that particles that enter the Hill sphere do so at a velocity $\sim R_{\text{Hill}}n_p$ – the *Hill velocity*. This is therefore the minimum (relative) velocity at which the gravitational scattering takes

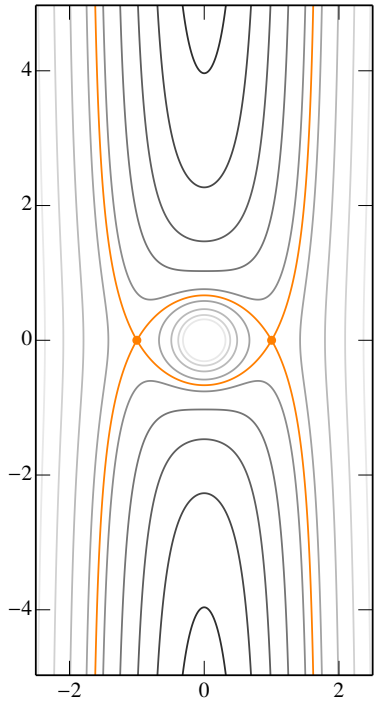

Figure 2.5: Zero velocity curves (contours of Φ_{eff}) in the $z = 0$ plane. Contours of larger Φ_{eff} are darker. The Lagrange equilibrium points L1 and L2 are indicated by circles. Distances are in units of Hill sphere. Curves are *not* orbits.

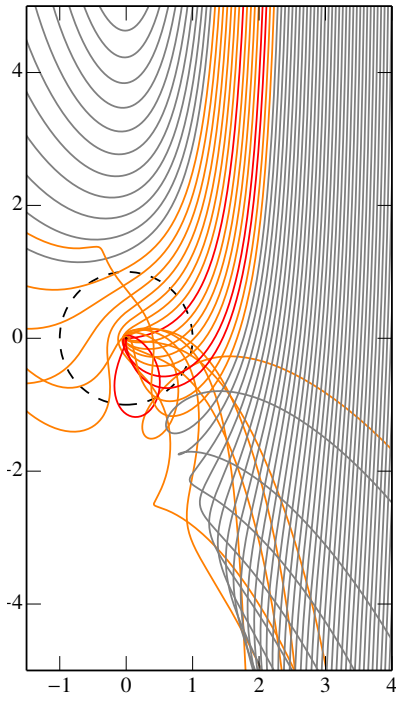

Figure 2.6: Examples of particle trajectories (initially on circulating orbits) in the CR3BP, obtained by integrating Equation (2.13). Particles that enter the Hill sphere (dashed circle) are highlighted. Red streams hit the planet ($R < R_p = 5 \times 10^{-3}R_{\text{Hill}}$).

place. Of course, the relative velocity will be larger as soon as the particle moves on an eccentric orbit. We can therefore distinguish the:

1. *shear-dominated limit*: $ev_K \lesssim R_{\rm Hill} n_p$. Encounter rates are given by the Keplerian shear.
2. *dispersion-dominated limit*: $ev_K \gtrsim R_{\rm Hill} n_p$. Encounter rates are determined by the random motions (eccentricities) of the bodies.

2.3 *Gap opening and planet migration*

We can use the Jacobi constant to obtain the migration rate of a planet for distant interactions. Distant interactions are interactions where the scattering angle $\theta_{\rm scat} \ll 1$ (see Eq. [1.42]) or, equivalently $b \gg b_{90}$. We assume that the initial motion of the TP is Keplerian, $v_\infty = |-\frac{3}{2} n_p b|$ in the $-y$ direction. After the scattering the particle will end up having a velocity in the x direction of $v_x = -\theta_{\rm scat} v_\infty$. The corresponding eccentricity is:*

$$e \sim \frac{|v_x|}{an} \approx \frac{4}{3} q_p \left(\frac{a}{b}\right)^2 \tag{2.15}$$

where we used that $n_p \approx n$ and $a_p \approx a$. Now that we have a change in eccentricity Equation (2.8) can be used to find the corresponding change in semimajor axis or, equivalently, impact parameter b. Calculating the Jabobi energy and equating J before (where $e = 0$) and after (where e is as in Eq. [2.15]) the encounter, while assuming that the changes $\delta a/a$ and δe^2 are small, one obtains:†

$$\frac{\delta b}{a_p} = \frac{4}{3} \left(\frac{a_p}{b}\right) \delta e^2. \tag{2.16}$$

Since $\delta e^2 > 0$, the attractive (gravitational) interaction has – perhaps counter-intuitively – pushed the particles further apart in semimajor axis. Distant encounters are therefore repulsive. These changes due to small-angle scattering are referred to as the *impulse approximation*. In a particulate disk, subsequent (inelastic) collisions can damp the eccentricity of the test particle, such that by the time of the next encounter the TP will again be on a circular orbit, at an impact parameter b incremented by δa. Since the interaction period is the synodical period of Equation (2.3), the TP moves away at a speed:

$$\frac{db}{dt} \sim \left(\frac{a_p}{b}\right)^4 q_p^2 n_p a_p. \tag{2.17}$$

Instead of particles, these expressions also apply to gaseous disks (Lin & Papaloizou 1979). Therefore, (massive) planets tend to open a gap in the disk, as they repel the gas. On the other hand, viscous diffusion tends to close the gap. When the timescale for particles to move over a distance b is equated with the viscous diffusion timescale over this length, $b^2/D_{\rm gas}$,we obtain the width of the gap, in steady state:

$$\left(\frac{b}{h_{\rm gas}}\right)^3 \sim \frac{q_p^2}{\alpha} \left(\frac{a}{h_{\rm gas}}\right)^5, \tag{2.18}$$

* This is a rather crude estimate as we have not accounted for the solar gravity during the encounter, which is especially pertinent in the $e \to 0$ limit. More precisely, the prefactor in Equation (2.15) will be $\simeq$2.24, approximately a factor of two larger.

† Here we used Equation (2.9) and the relation $\delta b^2 = 2b\delta b$, which is valid for distant encounters.

where we used the α-model for the viscosity. This model is very crude; for example, when $\alpha \to 0$ it predicts an infinitely large gap*. More importantly, for our essentially 2D arguments to be valid and for *gap opening* to commence, we must have that $b \gtrsim h_{\rm gas}$. This implies that gap opening commences at $q_p \simeq \alpha^{1/2}(h_{\rm gas}/a)^{5/2}$.†

Exercise 2.4 Planet migration:

(a) Relaxing the condition that the TP is massless, replace it by a narrow ring of particles of surface density Σ and width db. Argue that the *planet* migration rate is

$$\frac{da_p}{dt} \sim -\frac{2\pi a_p^2 \Sigma}{m_\star} q_p \left(\frac{a}{b}\right)^4 n_p db. \tag{2.19}$$

where $b = a - a_p \ll 1$ such that $a \approx a_p$.

(b) Consider b positive and add the contributions of many subsequent rings, assuming constant surface density Σ. Clearly the rings that are closest to the planet contribute the most. It can be argued that the minimum impact parameter for which the impulse approximation remains valid in the case of a gaseous disk is $b = b_{\rm min} \simeq h_{\rm gas}$ (This phenomenon is known as 'torque-cutoff'). In that case, show that the total migration rate becomes:

$$\frac{da_p}{dt} \sim -q_p \frac{a_p^2 \Sigma_p}{m_\star} \left(\frac{a_p}{h_{\rm gas}}\right)^3 a_p n_p \tag{2.20}$$

(c) Equation (2.20) is still incomplete, since it only considers the disk exterior to the planet. The disk interior to the planet will also exert a torque on the planet, giving rise to similar expression as Equation (2.20) but with one key difference. Which?

(d) The total torque (net migration rate) is therefore determined by the difference in the disk conditions (*e.g.* Σ and T) on both sides of the planet. Argue that the net migration rate in smooth disks (*i.e.* no gap opening) becomes:

$$\frac{da_p}{dt} \sim \gamma_a q_p \frac{a_p^2 \Sigma_{\rm gas}}{m_\star} \left(\frac{a_p}{h_{\rm gas}}\right)^2 n_p a_p \tag{2.21}$$

where $|\gamma_a|$ is of order unity. Hint: realize that the largest contributions arise from $\pm h_{\rm gas}$. Consider a Taylor-expansion of $\Sigma_{\rm gas}$ around $a = a_p$.

(e) Give an expression for the migration timescale $a_p/(da_p/dt)$. Evaluate this expression for an Earth-mass planet at 1 AU? Is Type-I migration important?

A more detailed analysis will reveal that the net "distant encounter" torque – more commonly known as the *Lindblad torque* – is usually negative, resulting in inward migration. It is important to realize that the torque resulting in planet migration is the net effect of the near-cancellation of two opposing torques: one from the interior disk and one from the exterior disk.

The Lindblad torque is not the only disk-planet interaction that results in planet migration. For example, the material moving on horseshoe orbits‡ can give rise to a *horseshoe torque* as (under certain conditions) the density before and after the U-turn is slightly different (Ward 1991).§ While the magnitude of the additional torque is always similar to that of the Lindblad torque calculated above, additional torques can alter the sign of the total torque and, hence, the

* The fallacy is that Equation (2.18) assumes steady-state; such a large gap may be reached only after an infinite amount of time.

† In the literature an alternative calculation for the gap opening criterion amounts to equating the viscous torque with the torque from the impulse approximation at a critical distance b_* – essentially the distance below which the impulse approximation becomes invalid. When $b_* = h_{\rm gas}$ we obtain a similar answer.

‡ sometimes referred to as the coorbital region, as the material rotates with the planet

§ This is a subtle effect roughly related to the conservation of vortensity – a fluid dynamical concept. In the vicinity of the protoplanet the analogy with (pressureless) particles becomes inaccurate and we often need to solve the fluid dynamic equations of the gas (numerically).

direction of the migration.

In Equation (2.21) the migration rate scales linearly with planet mass, which is valid for relatively low-mass planets. Torques linear with q_p are collectively referred to as *Type-I migration*. Indeed, for higher mass planets, gap opening suppresses the migration. In that case the migration is regulated by the viscosity (diffusion) of the disk. The subsequent migration is known as *Type-II migration*.

Apart from reducing the angular momentum of planets, disk torques will also damp the eccentricity. Eccentricity damping is often even more effective, because the energy associated with the eccentric motion ($\sim(ean_p)^2$) is much smaller than the orbital energy. It can therefore be expected that planets embedded in gaseous disk move on near-circular orbits.

Planets can migrate in gas-free disk. Planetesimals can drive migration, especially when the planetesimals are dynamically cold (low eccentricity). In particular, the *Nice model* hypothesises that planetesimal-driven migration caused Jupiter and Saturn to cross the 2:1 mean-motion resonance, which strongly excited the eccentricities of these planets, causing a major "shakeup" of the nascent solar system.* Other example of planetesimal-driven migration are the motion of moonlets in Saturn's rings and the presumably outward migration of Neptune (into the nascent Kuiper belt).

* There are many attractive features in the Nice model. It provides a natural explanation for the so-called Late Heavy Bombardment of the inner solar system. It also provides a mechanism for the capture of Jupiter's Trojan's (Morbidelli et al. 2005) and explains many of the characteristics in the Kuiper belt.

Tides raised on planetary bodies can also result in migration. A classical example is the Earth-Moon system, where the moon moves outwards as the Earth spins down. In these systems tides lead to a net loss of the total energy, while preserving the total angular momentum. If the orbit is still eccentric, the condition $h_z = 0$ results in an eccentricity decay of

$$\dot{e} = -\frac{\dot{E}}{E}\frac{(1-e^2)}{e}. \tag{2.22}$$

where $\dot{E} < 0$ is the rate of energy loss due to tides. Equation (2.22) illustrates once more that eccentricity damping operates fast (unless $e \approx 1$). Planets on very eccentric orbits, such that their periapsis distance lies close enough to raise tides on the star can therefore undergo circularization, decreasing the eccentricity will preserving (approximately) their periapsis. *Tidal damping*† has been invoked to explain the presence of hot-Jupiters.

† Confusingly, planet-disk interactions are also referred to as 'tidal'.

2.4 *Resonant perturbations and resonant trapping*

In this section we consider a slightly more involved model for the evolution of the orbital elements of the test particle, which allows for resonant interactions. We nevertheless consider the simplest application of the circularly restricted 3-body problem, with the secondary moving on a circular orbit and perturbing a body exterior to it.‡

‡ Our analysis follows Chapters 6 and 8 of Murray & Dermott (1999). However, their key application concerns an exterior perturber (*e.g.* Jupiter's interaction with asteroidal bodies), while we consider an interior perturber.

The *disturbing function* $\mathcal{R}_i$ is defined $\ddot{\mathbf{r}}_i = \nabla_i(U_i + \mathcal{R}_i)$ where $U_i = G(m_\star + m_i)/r_i$ is the potential due to the star, $\mathcal{R}_i$ due to the other bodies in the system, and ∇_i highlights that the gradient is with

respect to the i-body. For a perturber at $\mathbf{r}_k$ of mass m_k the disturbing function reads:

$$\mathcal{R}_i = \frac{Gm_k}{|\boldsymbol{r}_k - \boldsymbol{r}_i|} - \frac{Gm_k(\boldsymbol{r}_i \cdot \boldsymbol{r}_k)}{r_k^3}. \tag{2.23}$$

Here, $|\boldsymbol{r}_k - \boldsymbol{r}_i|$ in the first term – the direct term – is simply the distance to the k-body and the second, indirect, term arises because the origin of the coordinate system lies at $m_\star$ and not at the barycentre.

Clearly, $\mathcal{R}_i$ is a rapidly varying function with time, because the position $\boldsymbol{r}_i$ continuously changes. On the other hand, provided the bodies are far enough apart, the orbital elements will change only slowly. It is of course more meaningful to consider changes in the (slowly varying) orbital elements than in the positions. To proceed, the disturbing function $\mathcal{R}_i$ must be expanded in terms of the orbital elements. We will then use Lagrange planetary equations (without derivation) to find the time-evolution of the orbital elements (*e.g.* $\dot{a}$, $\dot{e}$).

We will not detail how the expansion of $\mathcal{R}_i$ proceeds, as this is a very technical derivation. Using appropriate mathematical tools, the disturbing function can be written as a summation of cosine arguments (ϕ_n) and associated prefactors (S_n):

$$\mathcal{R}_i = Gm_k \sum_n S_n \cos(\phi_n); \qquad \phi_n = j_1\lambda_i + j_2\lambda_k + j_3\varpi_i + j_4\varpi_k + j_5\Omega_{\mathrm{node},i} + j_6\Omega_{\mathrm{node},k} \tag{2.24}$$

where S_n is a function of a, e, and inclination of both bodies and the $\{j_i\}$ are integers. The total number of these terms depends on the order of the expansion (*e.g.* 2nd order in eccentricity and inclination) but is already high for even low order. Fortunately, many of these terms are rapidly varying (because they contain the mean longitudes λ) and can hence be discarded when the long-term average is considered (we are interested here in times much longer than the orbital periods). Hence, we categorize:

- Highly variable terms. These are terms that contain the mean longitude in the cosine arguments, *e.g.* $\langle \cos\lambda_i \rangle$, and average-out over long periods.
- Slowly variable terms. These are terms in the expansion that do not contain λ_i and λ_k in the cosine arguments. Hence, they do not average out. They are said to give a *secular* contribution.
- Resonant terms. These are terms where $j_1\lambda_i + j_2\lambda_k$ are close to zero, such that they also do not average out on orbital periods.

To illustrate the theory, we consider a special case, which is essentially the circular, restricted 3-body problem with the interior body moving on a circular orbit and planar motion (no inclination). Furthermore, we include only j:j+1 resonances. In that case, the long-term average of the disturbing function reads (Murray & Dermott 1999):

$$\langle \mathcal{R} \rangle = \frac{Gm_p}{a_p}\left[e^2\mathcal{F}_{\mathrm{sec}}(\alpha) + e\mathcal{F}(\alpha)\cos\phi\right] \tag{2.25a}$$

$$\phi = (j+1)\lambda - j\lambda_p - \varpi \tag{2.25b}$$

where (fixed) quantities of the inner body are denoted by subscript p and the other quantities refer to the outer (perturbed) body and $\alpha = a_p/a < 1$. In Equation (2.25a) the first term on the RHS ($\mathcal{F}_{\rm sec}$) contains the secular terms of the disturbing function (those that do not contain mean longitudes in the cosine arguments).* When the system is far from resonance, these secular terms determine the evolution. But because the secular part is second order in eccentricity, they do not feature in the following. The second term in Equation (2.25a) describes the resonant interaction with ϕ the *resonant angle* and $\lambda = M + \varpi$ the mean longitude. †

Invoking the Lagrange planetary equations to the resonant part of the $\langle \mathcal{R} \rangle$, we obtain a system of equations involving a, e and ϕ:

$$\dot{n} \equiv -\frac{3}{a^2}\frac{\partial \langle R \rangle}{\partial \lambda} = \frac{G m_p}{a_p^3} 3(j+1)\alpha^2 \mathcal{F}(\alpha) e \sin\phi \tag{2.29a}$$

$$\dot{e} \equiv \frac{-1}{n a^2 e}\frac{\partial \langle R \rangle}{\partial \varpi} = -\frac{G m_p}{a_p^3}\alpha^2 \frac{\mathcal{F}(\alpha)\sin\phi}{n} \tag{2.29b}$$

$$\dot{\varpi} \equiv \frac{1}{n a^2 e}\frac{\partial \langle R \rangle}{\partial e} = \frac{G m_p}{a_p^3}\alpha^2 \frac{\mathcal{F}(\alpha)\cos\phi}{ne} \tag{2.29c}$$

Let us rewrite the last equation in terms of the resonant angle, $\dot{\phi} = (j+1)n - j n_p - \dot{\varpi} = ((j+1)\alpha^{3/2} - j) n_p - \dot{\varpi}$.‡ We then obtain a closed system of equations for e, a, and ϕ, that can be succinctly written as:

$$\dot{a}/a = -\dot{\alpha}/\alpha = -2(j+1) n_p C_q(\alpha) e \sin\phi \tag{2.30a}$$

$$\dot{e} = -n_p C_q(\alpha)\sin\phi \tag{2.30b}$$

$$\dot{\phi} = -\delta j_\alpha n_p - C_q(\alpha) n_p \frac{\cos\phi}{e} \tag{2.30c}$$

where we approximated $G m_p = q_p n_p^2 a_p^3$, defined $\delta j_\alpha = j - (j+1)\alpha^{3/2}$ as a "distance" from exact resonance and $C_q(\alpha) = \alpha^{1/2}\mathcal{F}(\alpha) q_p$ as a dimensionless parameter. Note that while we used $e \ll 1$ the expressions evidently break down when $e \to 0$. We will hence assume that $q_p/e \sim C_q(\alpha)/e$ remains small.

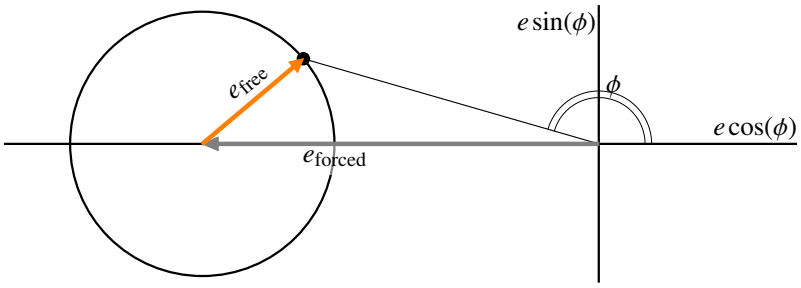

Figure 2.7: A resonant configuration. The eccentricity is composed of a "free" and "forced" (due to the resonant interaction) component. There is no damping.

Exercise 2.5 Resonances:

(a) Verify that Equation (2.30) does not have a steady solution.

(b) Assume the system is in resonance and that α is approximately constant, *i.e.* we ignore Equation (2.30a). Define $h = e\cos\phi$ and $k =$

* $$\mathcal{F}_{\rm sec.} = \frac{\alpha^2}{8}\left(2\frac{\partial}{\partial\alpha} + \alpha\frac{\partial^2}{\partial\alpha^2}\right) b_{1/2}^{(0)} \tag{2.26}$$

(Murray & Dermott 1999), where $b_{1/2}^{(n)}(\alpha)$ are the *Laplace coefficients*:

$$b_{1/2}^{(n)}(\alpha) \equiv \frac{1}{\pi}\int_0^{2\pi} \frac{\cos nx\, dx}{(1 - 2\alpha\cos x + \alpha^2)^s}. \tag{2.27}$$

† $$\mathcal{F}(\alpha) \equiv \frac{\alpha}{2}\left(2j + 1 + \alpha\frac{\partial}{\partial\alpha}\right) b_{1/2}^{(j)}(\alpha) - \frac{\delta_{j1}}{2\alpha} \tag{2.28}$$

(Murray & Dermott 1999), where δ_{j1} represents the indirect contribution to the perturbing function for $j = 1$.

‡ Formally, we should also have accounted for the drift in the ascending node (or, rather, the epoch as...) but this can be shown to have a $\mathcal{O}(q_0 e^0)$ dependence and is therefore negligible.

$e \sin\phi$, which linearizes the above equations:

$$\dot{h} = \omega_0 k \tag{2.31a}$$
$$\dot{k} = -\omega_0 h - C_q(\alpha) n_0 \tag{2.31b}$$

where $\omega_0 = \delta j_\alpha n_0$. Verify that the solution to this system of equations are circles with offset $-e_{\text{forced}} \equiv -C_{q,\alpha} n_0/\omega_0 = -C_{q,\alpha}/\delta j_\alpha$ and a radius that is a constant of integration, say, e_{free}. An example is plotted in Figure 2.7 (for positive δj_α). Explain the meaning of "free" and "forced".

(c) Explain that when $e_{\text{free}} \ll e_{\text{forced}}$ (as in the figure) ϕ librates (oscillates) around $\phi = \pi$, *i.e.* that the solution does not attain *all* values of ϕ. Explain that the system is then in resonance.

(d) Since ϕ librates around π resonances take place at apoapsis. Explain that this configuration promotes stability.

In the presence of a dissipative medium, eccentricities are damped (*i.e.* gas drag for planetesimals or tidal torques on planets). In that case a term $-e/t_e$ must be added to Equation (2.30b) where t_e is the damping timescale for eccentricity. The presence of a dissipative term allows equilibrium solutions. Assuming once more that α is constant, we obtain for the equilibrium values of the eccentricity and resonant angle:

$$e = \frac{\alpha^{1/2} q_p F(\alpha) n_p t_e}{\sqrt{1 + (n_p t_e \delta j_\alpha)^2}}; \qquad \sin\phi = \frac{-1}{\sqrt{1 + (n_p t_e \delta j_\alpha)^2}} \tag{2.32}$$

In Figure 2.8 the eccentricity is plotted as function of $\alpha^{-1} = a/a_0$.

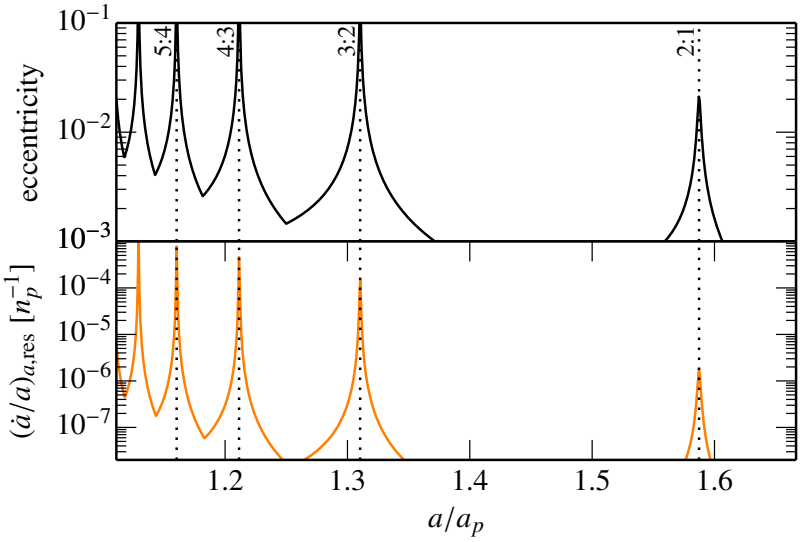

Figure 2.8: Top: equilibrium values for eccentricity, given by Equation (2.32), for $q_p = 10^{-4}$, $t_e n_p = 10^3$. Bottom: resonance-induced migration rate (outwards), corresponding to the solution. In evaluating Equation (2.32) the value of j that was closest to exact resonance was taken, *i.e.* δj_α can be both positive (exterior to the resonance) as well as negative (interior).

Using these equilibrium values for e and ϕ the migration rate $\dot{a}$ (Eq. [2.30a]) can be calculated, which is shown in the bottom panel of Figure 2.8. The perturbation in semi-major axis is positive. This forced (outward) migration rate can be contrasted with the (usually inward) migration rate (t_a) from *e.g.* Type-I migration or planetesimal orbital decay. Assuming inward migration, the resonant forcing can stop the migration when $(\dot{a}/a)_{\text{res}} > 1/t_a$. This phenomenon is known as *resonant trapping*. Figure 2.8 shows that the higher j-values

have stronger forcing. Therefore, when a body does not stop at –say– the 2:1, it can yet end up near a 3:2 or higher-j resonance.

Exercise 2.6:

(a) From Equation (2.30a) it may appear that inward migration will stop eventually, since, as $j \to \infty$, $\dot{a}$ can become arbitrarily large. What is the fallacy with this statement?

(b) Another reason why resonant trapping may fail is when the synodical timescale at resonance j becomes larger than the time to migrate to a_p. Why? Give the expression corresponding to this condition.

2.5 *Kozai-Lidov cycles*

Resonant configuration are not limited to $j : j+1$ mean motion resonances, discussed above. Higher order resonances, like $j : j+2$ may also play a role (but are much weaker). Resonances do not have to be of the commensurate type, however. These are referred to as secular resonances. In the planet formation context, the *Kozai-Lidov resonance*s are particularly important. These can occur in stellar binary systems, provided that the planet is sufficiently inclined with respect to the plane in which the two stars move. If it indeed exceeds a critical angle of $\arccos\sqrt{\frac{3}{5}} \approx 39^\circ$ a "runaway" growth of eccentricity takes place. At its peak the eccentricity will be near unity, such that the periapsis of the erstwhile far out planet will be very close to the star. Thereafter the eccentricity decreases, near-cicular motion returns, and the process repeats itself. The timescale associated with these Kozai-Lidov cycles is

$$t_{\rm KL} \sim \frac{m_\star}{m_2}\frac{P_2^2}{P^2}P, \tag{2.33}$$

where P_2 is the period of the binary star and P that of the planet.

As $t_{\rm KL}$ is usually much less than the lifetime of the system ($\sim$Gyr), there will be many cycles. However, the cycles do not necessarily conserve energy, because of tidal interactions with the star during periapsis. This removes energy from the system, causing the orbit to shrink (a decreases). This mechanism has been invoked as an explanation for the hot-Jupiters. In contrast to disk-migration mechanism outlined in §2.3, no gaseous disk is needed.

A
List of symbols

Δv	relative velocity
Φ	gravitational potential
Σ	surface density (solids)
Σ_{gas}	surface density (gas)
Θ	gravitational focusing factor
Ω	orbital frequency
F_D	drag force
A	(surface) area
C_D	drag constant
D_f	fractal index
E	energy
$\mathcal{E}^*$	combined elastic modulus
G	Newton's gravitational constant
J	Jacobi energy
K	rate constant/collision kernel
M	mass (large bodies); mean anomaly (Ch. 2)
$M_\oplus$, $M_\odot$	mass of the Earth, Sun
P	orbital period; also: pressure
Q_T	Toomre stability parameter
R	radius (large bodies)
R_{Bondi}, R_{Hill}	Bondi/Hill radius
Re_p	Reynolds number
T	temperature
Z	dust-to-gas ratio (local or global)

α	turbulence strength parameter; ratio a_2/a_1 (Ch. 2)
γ	surface energy density
ηv_K	headwind velocity gas disk
κ	epicycle frequency
μ	mean molecular weight
ν	viscosity; true anomaly (Ch. 2)
$\rho_\bullet$	internal density
ρ_{gas}	gas density
a	semi-major axis/disk orbital radius
a_c	contact area radius
b	impact parameter
c_s	isothermal sound speed
e	eccentricity
h_{gas}/h_p	disk scaleheight (gas/particles)
i	inclination
k	wave number
k_B	Boltzmann's constant
l_z, $\boldsymbol{l}$	angular momentum (vector)
l_{mfp}	mean free path
ℓ	eddy size
m	mass (small particles)
n	number density; mean motion (Ch. 2)
q_p	mass ratio planet:star
r	radial distance (from origin)
s	radius (small particles, grains)
t	time
t_{stop}	stopping time
t_{settl}	settling time
u	velocity (gas)
v	velocity (particles)
v_K	Keplerian velocity

B
Bibliography

Binney, J. & Tremaine, S. 2008, Galactic Dynamics: Second Edition, ed. Binney, J. & Tremaine, S. (Princeton University Press)
Blum, J., Wurm, G., Poppe, T., & Heim, L.-O. 1998, Earth Moon and Planets, 80, 285
Chambers, J. E., Wetherill, G. W., & Boss, A. P. 1996, Icarus, 119, 261
Chandrasekhar, S. 1961, Hydrodynamic and hydromagnetic stability
Chokshi, A., Tielens, A. G. G. M., & Hollenbach, D. 1993, ApJ, 407, 806
Dominik, C. & Tielens, A. G. G. M. 1997, ApJ, 480, 647
Goldreich, P. & Ward, W. R. 1973, ApJ, 183, 1051
Hayashi, C., Nakazawa, K., & Nakagawa, Y. 1985, in Protostars and Planets II, ed. D. C. Black & M. S. Matthews (Univ. of Arizona Press, Tucson), 1100–1153
Hertz, H. 1882, Verhandlungen des Vereins zur Beforderung des Gewerbefleisses
Iwasaki, K., Tanaka, H., Nakazawa, K., & Hiroyuki, E. 2001, Publications of the Astronomical Society of Japan, 53, 321
Jacquet, E., Balbus, S., & Latter, H. 2011, MNRAS, 415, 3591
Johansen, A., Youdin, A., & Mac Low, M. 2009, ApJ, 704, L75
Johnson, K. L., Kendall, K., & Roberts, A. D. 1971, Proceedings of the Royal Society of London Series A, 324, 301
Kataoka, A., Tanaka, H., Okuzumi, S., & Wada, K. 2013, A&A, 557, L4
Krijt, S., Güttler, C., Heißelmann, D., Dominik, C., & Tielens, A. G. G. M. 2013, Journal of Physics D Applied Physics, 46, 435303
Landau, L. D. & Lifshitz, E. M. 1959, Fluid mechanics
Lin, D. N. C. & Papaloizou, J. 1979, MNRAS, 186, 799
Malyshkin, L. & Goodman, J. 2001, Icarus, 150, 314
Markiewicz, W. J., Mizuno, H., & Voelk, H. J. 1991, A&A, 242, 286
Marois, C., Zuckerman, B., Konopacky, Q. M., Macintosh, B., & Barman, T. 2010, Nature, 468, 1080
Morbidelli, A., Levison, H. F., Tsiganis, K., & Gomes, R. 2005, Nature, 435, 462
Muller, V., Yushchenko, V., & Derjaguin, B. 1980, J. Colloid Interface Sci, 77, 91
Murray, C. D. & Dermott, S. F. 1999, Solar system dynamics
Nakagawa, Y., Sekiya, M., & Hayashi, C. 1986, Icarus, 67, 375
Ormel, C. W. & Cuzzi, J. N. 2007, A&A, 466, 413
Ormel, C. W. & Spaans, M. 2008, ApJ, 684, 1291
Pan, L. & Padoan, P. 2010, Journal of Fluid Mechanics, 661, 73
Raymond, S. N., Quinn, T., & Lunine, J. I. 2006, Icarus, 183, 265
Saffman, P. G. & Turner, J. S. 1956, Journal of Fluid Mechanics, 1, 16
Safronov, V. S. 1969, Evolution of the Protoplanetary Cloud and Formation of Earth and the Planets, ed. V. S. Safronov (Moscow: Nauka. Transl. 1972 NASA Tech. F-677)
Seizinger, A., Krijt, S., & Kley, W. 2013, A&A, 560, A45
Shakura, N. I. & Sunyaev, R. A. 1973, A&A, 24, 337
Smoluchowski, M. V. 1916, Zeitschrift für Physik, 17, 557
Tanaka, H., Himeno, Y., & Ida, S. 2005, ApJ, 625, 414
Völk, H. J., Jones, F. C., Morfill, G. E., & Roeser, S. 1980, A&A, 85, 316
Ward, W. R. 1991, in Lunar and Planetary Institute Science Conference Abstracts, Vol. 22, Lunar and Planetary Institute Science Conference Abstracts, 1463
Weidenschilling, S. J. 1977, Astrophysics and Space Science, 51, 153
Weidenschilling, S. J. 1984, Icarus, 60, 553
Wetherill, G. W. 1990, Icarus, 88, 336
Youdin, A. N. & Goodman, J. 2005, ApJ, 620, 459
Zhou, J.-L., Lin, D. N. C., & Sun, Y.-S. 2007, ApJ, 666, 423

C
Index