8

Collisions and Encounters of Stellar Systems

Our Galaxy and its nearest large neighbor, the spiral galaxy M31, are falling towards one another and will probably collide in about 3 Gyr (see Plate 3 and Box 3.1).

A collision between our Galaxy and M31 would have devastating consequences for the gas in both systems. If a gas cloud from M31 encountered a Galactic cloud, shock waves would be driven into both clouds, heating and compressing the gas. In the denser parts of the clouds, the compressed post-shock gas would cool rapidly and fragment into new stars. The most massive of these would heat and ionize much of the remaining gas and ultimately explode as supernovae, thereby shock-heating the gas still further. Depending on the relative orientation of the velocity vectors of the colliding clouds, the post-collision remnant might lose much of its orbital angular momentum, and then fall towards the bottom of the potential well of the whole system, thereby enhancing the cloud-collision and star-formation rates still further. We do not yet have a good understanding of this complex chain of events, but there is strong observational evidence that collisions between gas-rich galaxies like the Milky Way and M31 cause the extremely high star-formation rates observed in starburst galaxies (§8.5.5).

In contrast to gas clouds, stars emerge unscathed from a galaxy collision.

To see this, consider what would happen to the solar neighborhood in a collision with the disk of M31. According to Table 1.1, the surface density of visible stars in the solar neighborhood is $\simeq 30\,\mathcal{M}_\odot\,\mathrm{pc}^{-2}$. Assuming that most of these are similar to the Sun, the number density of stars is $N \simeq 30\,\mathrm{pc}^{-2}$ and the fraction of the area of the galactic disk that is filled by the disks of these stars is of order $N\pi R_\odot^2 \approx 5 \times 10^{-14}$. Thus even if M31 were to score a direct hit on our Galaxy, the probability that even one of the 10^{11} stars in M31 would collide with any star in our Galaxy is small.[1]

However, the distribution of the stars in the two galaxies would be radically changed by such a collision, because the gravitational field of M31 would deflect the stars of our Galaxy from their original orbits and vice versa for the stars of M31. In this process, which is closely related to violent relaxation (§4.10.2), energy is transferred from ordered motion (the relative motion of the centers of mass of the two galaxies) to random motion. Thus the collision of two galaxies is inelastic, just as the collision of two lead balls is inelastic—in both cases, ordered motion is converted to random motion, of the stars in one case and the molecules in the other (Holmberg 1941; Alladin 1965). Of course, since stars move according to Newton's laws of motion, the total energy of the galactic system is strictly conserved, in contrast to the lead balls where the energy in random motion of the molecules (i.e., heat) is eventually lost as infrared radiation.

A consequence of this inelasticity is that galaxy collisions often lead to **mergers**, in which the final product of the collision is a single merged stellar system. In fact, we believe that both galaxies and larger stellar systems such as clusters of galaxies are created by a hierarchical or "bottom-up" process in which small stellar systems collide and merge, over and over again, to form ever larger systems (§9.2.2).

The most straightforward way to investigate what happens in galaxy encounters is to simulate the process using an N-body code. Figure 8.1 shows an N-body simulation of the collision of the Galaxy and M31. This is an example of a **major merger**, in which the merging galaxies have similar masses, and the violently changing gravitational field leads to a merger remnant that looks quite different from either of its progenitors. In contrast, **minor mergers**, in which one of the merging galaxies is much smaller than the other, leave the larger galaxy relatively unchanged.

Not every close encounter between galaxies leads to a merger. To see this, let v_∞ be the speed at which galaxy A initially approaches galaxy B and consider how the energy that is gained by a star in galaxy A depends on v_∞. As we increase v_∞, the time required for the two galaxies to pass through one another decreases. Hence the velocity impulse $\Delta\mathbf{v} = \int \mathrm{d}t\,\mathbf{g}(t)$ due to the gravitational field $\mathbf{g}(t)$ from galaxy B decreases, and less and less energy is transferred from the relative orbit of the two galaxies to the random motions

[1] We have neglected gravitational focusing, which enhances the collision probability by a factor of about five but does not alter this conclusion (eq. 7.194).

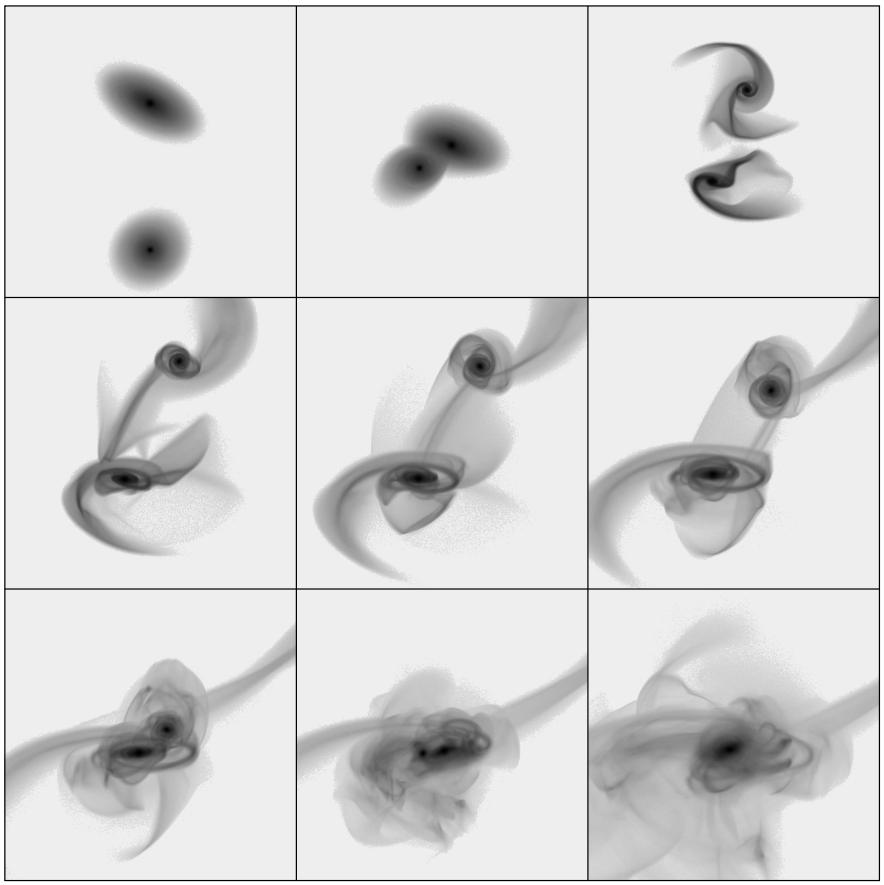

Figure 8.1 An N-body simulation of the collision between the Galaxy (bottom) and M31 (top) which is expected to occur roughly 3 Gyr from now. The simulation follows only the evolution of the stars in the two galaxies, not the gas. Each galaxy is represented by roughly 10^8 stars and dark-matter particles. The viewpoint is from the north Galactic pole. Each panel is 180 kpc across and the interval between frames is 180 Myr. After the initial collision, a open spiral pattern is excited in both disks and long tidal tails are formed. The galaxies move apart by more than 100 kpc and then fall back together for a second collision, quickly forming a remnant surrounded by a complex pattern of shells. The shells then gradually phase mix, eventually leaving a smooth elliptical galaxy. Image provided by J. Dubinski (Dubinski, Mihos, & Hernquist 1996; Dubinski & Farah 2006).

of their stars; in fact, when v_∞ is large, $|\Delta\mathbf{v}| \propto 1/v_\infty$. Thus, when v_∞ exceeds some critical speed v_f, the galaxies complete their interaction with sufficient orbital energy to make good their escape to infinity. If $v_\infty < v_\mathrm{f}$, the galaxies merge, while if $v_\infty \gg v_\mathrm{f}$ the encounter alters both the orbits and

the internal structures of the galaxies only slightly.[2] This simple argument explains why most galaxies in rich clusters have not merged: although the density of galaxies in the clusters is high, so collisions are frequent, the random velocities of cluster galaxies are so high that the loss of orbital energy in a collision is negligible—the galaxies simply pass through one another, like ghosts.

Until the 1970s, most astronomers believed that collisions between galaxies were negligible, except in high-density regions such as clusters. This belief was based on the following argument. The velocities of galaxies are the sum of the Hubble velocity (eq. 1.13) appropriate to that galaxy's position, and a residual, or **peculiar velocity**. Typical peculiar velocities are $v_{\rm p} \approx 100\,{\rm km\,s^{-1}}$ (Willick et al. 1997). The number density of galaxies is described by the Schechter law (eq. 1.18), so the density of luminous galaxies ($L \gtrsim L_\star$) is $n \approx \phi_\star \approx 10^{-2}\,{\rm Mpc^{-3}}$. Most of the stars in a typical luminous galaxy are contained within a radius $R \approx 10\,{\rm kpc}$, so the collision cross-section between two such galaxies is $\Sigma \approx \pi(2R)^2$. If the positions and velocities of the galaxies are uncorrelated, the rate at which an $L_\star$ galaxy suffers collisions with similar galaxies is then expected to be of order $n\Sigma v_{\rm p} \approx 10^{-6}\,{\rm Gyr^{-1}}$, so only about one galaxy in 10^5 would suffer a collision during the age of the universe. Such arguments led astronomers to think of galaxies as island universes that formed and lived in isolation.

This estimate of the collision rate turns out to be far too low, for two reasons. (i) The stars in a galaxy are embedded in a dark halo, which can extend to radii of several hundred kpc. Once two dark halos start to merge, their high-density centers, which contain the stars and other baryonic matter, experience a drag force from dynamical friction (§7.4.4) as they move through the common halo. Dynamical friction causes the baryon-rich central regions to spiral towards the center of the merged halo, where they in turn merge. Thus the appropriate cross-section is proportional to the square of the dark-halo radius rather than the square of the radius of the stellar distribution. (ii) As we describe in §9.1, the departures of the matter distribution in the universe from exact homogeneity arose through gravitational forces, and in particular the peculiar velocities of galaxies relative to the Hubble flow are caused by gravitational forces from nearby galaxies. Consequently, the peculiar velocities of nearby galaxies are correlated—nearby galaxies are falling towards one another, just like our Galaxy and M31—so the collision rate is much higher than it would be if the peculiar velocities were randomly oriented. In §8.5.6 we show that the merger rate for $L_\star$ galaxies is $\sim 0.01\,{\rm Gyr^{-1}}$, 10^4 times larger than our naïve estimate.

When two dark halos of unequal size merge, the smaller halo orbits within the larger one, on a trajectory that steadily decays through dynamical friction. As the orbit decays, the satellite system is subjected to disruptive

[2] Thus, galaxies behave somewhat like the toy putty that is elastic at high impact speeds, but soft and inelastic at low speeds.

processes of growing strength. These include steady tidal forces from the host galaxy, and rapidly varying forces as the smaller halo passes through the pericenter of its orbit. As stars are lost from the satellite, they spread out in long, thin tidal streamers that can provide vivid evidence of ongoing disruption. Eventually the satellite is completely disrupted, and its stars and dark matter phase-mix with those of the host system.

These processes, which we examine in this chapter, are common to a wide variety of astrophysical systems. Dynamical friction (§8.1) drives the orbital evolution not only of satellite galaxies, but also black holes and globular clusters near the centers of galaxies, and bars in barred spiral galaxies. Tidal forces erode satellite galaxies, globular clusters, and galaxies in clusters, and also determine the lifetimes of star clusters and wide binary stars. We shall focus on the effects of tidal forces in two extreme and analytically tractable limits: §8.2 is devoted to impulsive tides, which last for only a short time, while §8.3 examines the effects of static tides. §8.4 describes the dynamics of encounters in galactic disks, and their effect on the kinematics of stars in the solar neighborhood. Finally, in §8.5 we summarize and interpret the observational evidence for ongoing mergers between galaxies, and estimate the merger rate.

8.1 Dynamical friction

A characteristic feature of collisions of stellar systems is the systematic transfer of energy from their relative orbital motion into random motions of their constituent particles. This process is simplest to understand in the limiting case of minor mergers, in which one system is much smaller than the other.

We consider a body of mass M traveling through a population of particles of individual mass $m_a \ll M$. Following §1.2.1 we call M the subject body and the particles of mass m_a field stars. The subject body usually is a small galaxy or other stellar system and thus has non-zero radius, but we shall temporarily assume that it is a point mass. The field stars are members of a much larger host system of mass $\mathcal{M} \gg M$, which we assume to be so large that it can be approximated as infinite and homogeneous. The influence of encounters with the field stars on the subject body can then be characterized using the diffusion coefficients derived in §7.4.4. Because the test body is much more massive than the field stars, the first-order diffusion coefficients $D[\Delta v_i]$ are much larger than the second-order coefficients $D[\Delta v_i \Delta v_j]/v$ (cf. eqs. 7.83 with $m \gg m_a$). Thus the dominant effect of the encounters is to exert dynamical friction (page 583), which decelerates the subject body at a rate

$$\frac{\mathrm{d}\mathbf{v}_M}{\mathrm{d}t} = D[\Delta\mathbf{v}] = -4\pi G^2 M m_a \ln\Lambda \int \mathrm{d}^3\mathbf{v}_a\, f(\mathbf{v}_a) \frac{\mathbf{v}_M - \mathbf{v}_a}{|\mathbf{v}_M - \mathbf{v}_a|^3}, \qquad (8.1a)$$

where

$$\Lambda \approx \frac{b_{\max}}{b_{90}} \approx \frac{b_{\max} v_{\rm typ}^2}{GM} \gg 1 \tag{8.1b}$$

and b_{90} is the 90° deflection radius defined in equation (3.51). Here we have used equations (7.83), assuming $M \gg m_a$ and adjusting the notation appropriately. The field-star DF $f(\mathbf{x}, \mathbf{v}_a)$ is normalized so $\int \mathrm{d}^3\mathbf{v}_a\, f(\mathbf{x}, \mathbf{v}_a) = n(\mathbf{x})$, where n is the number density of field stars in the vicinity of the subject body.

We now estimate the typical value of the factor Λ in the Coulomb logarithm. When a subject body of mass M orbits in a host system of mass $\mathcal{M} \gg M$ and radius $\mathcal{R}$, the typical relative velocity is given by $v_{\rm typ}^2 \approx G\mathcal{M}/\mathcal{R}$. To within a factor of order unity, the maximum impact parameter $b_{\max} \approx R$, where R is the orbital radius of the subject body. Then $\Lambda \approx (\mathcal{M}/M)(R/\mathcal{R})$, which is large whenever $M \ll \mathcal{M}$, unless the subject body is very close to the center of the host.

If the subject body has a non-zero radius, the appropriate value for the Coulomb logarithm is modified to

$$\ln\Lambda = \ln\left(\frac{b_{\max}}{\max(r_{\rm h}, GM/v_{\rm typ}^2)}\right), \tag{8.2}$$

where $r_{\rm h}$ is the half-mass radius of the subject system (see Problem 8.2).

If the field stars have an isotropic velocity distribution,[3] equation (7.88) yields a simpler expression for the dynamical friction,

$$\frac{\mathrm{d}\mathbf{v}_M}{\mathrm{d}t} = -16\pi^2 G^2 M m_a \ln\Lambda \left[\int_0^{v_M} \mathrm{d}v_a\, v_a^2 f(v_a)\right] \frac{\mathbf{v}_M}{v_M^3}; \tag{8.3}$$

thus, only stars moving slower than M contribute to the friction. Like an ordinary frictional drag, the force described by equation (8.3) always opposes the motion ($\mathrm{d}\mathbf{v}_M/\mathrm{d}t$ is anti-parallel to $\mathbf{v}_M$). Equation (8.3) is usually called **Chandrasekhar's dynamical friction formula** (Chandrasekhar 1943a).

If the subject mass is moving slowly, so v_M is sufficiently small, we may replace $f(v_a)$ in the integral of equation (8.3) by $f(0)$ to find

$$\frac{\mathrm{d}\mathbf{v}_M}{\mathrm{d}t} \simeq -\frac{16\pi^2}{3} G^2 M m_a \ln\Lambda f(0)\, \mathbf{v}_M \quad (v_M \text{ small}). \tag{8.4}$$

Thus at low velocity the drag is proportional to v_M—just as in Stokes's law for the drag on a marble falling through honey. On the other hand, for sufficiently large v_M, the integral in equation (8.3) converges to a definite limit equal to the number density n divided by 4π:

$$\frac{\mathrm{d}\mathbf{v}_M}{\mathrm{d}t} = -4\pi G^2 M m_a n \ln\Lambda \frac{\mathbf{v}_M}{v_M^3} \quad (v_M \text{ large}). \tag{8.5}$$

[3] See Problem 8.3 for the case of an ellipsoidal velocity distribution.

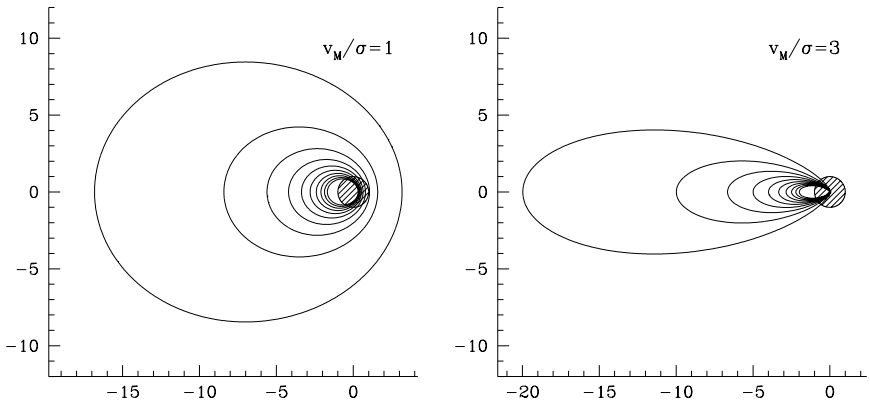

Figure 8.2 A mass M travels from left to right at speed v_M, through a homogeneous Maxwellian distribution of stars with one-dimensional dispersion σ. Deflection of the stars by the mass enhances the stellar density downstream, and the gravitational attraction of this wake on M leads to dynamical friction. The contours show lines of equal stellar density in a plane containing the mass M and the velocity vector $\mathbf{v}_M$; the velocities are $v_M = \sigma$ (left panel) and $v_M = 3\sigma$ (right panel). The fractional overdensities shown are $0.1, 0.2, \ldots, 0.9, 1$. The unit of length is chosen so that $GM/\sigma^2 = 1$. The shaded circle has unit radius and is centered at M. The overdensities are computed using equation (8.148), which is based on linear response theory; for a nonlinear treatment see Mulder (1983).

Thus the frictional force *falls* like ${v_M}^{-2}$—in contrast to the motion of solid bodies through fluids, where the drag force *grows* as the velocity increases.

If $f(\mathbf{v}_a)$ is Maxwellian with dispersion σ, then equation (8.3) becomes (cf. eqs. 7.91–7.93)

$$\frac{\mathrm{d}\mathbf{v}_M}{\mathrm{d}t} = -\frac{4\pi G^2 M n m \ln\Lambda}{v_M^3}\left[\mathrm{erf}(X) - \frac{2X}{\sqrt{\pi}}\mathrm{e}^{-X^2}\right]\mathbf{v}_M, \tag{8.6}$$

where $X \equiv v_M/(\sqrt{2}\sigma)$ and erf is the error function (Appendix C.3). This important formula illustrates two features of dynamical friction:

(i) The frictional drag is proportional to the mass density nm of the stars being scattered, but independent of the mass of each individual star. In particular, if we replace nm in equation (8.6) by the overall background density ρ, we obtain a formula that is equally valid for a host system containing a spectrum of different stellar masses:

$$\frac{\mathrm{d}\mathbf{v}_M}{\mathrm{d}t} = -\frac{4\pi G^2 M \rho \ln\Lambda}{v_M^3}\left[\mathrm{erf}(X) - \frac{2X}{\sqrt{\pi}}\mathrm{e}^{-X^2}\right]\mathbf{v}_M. \tag{8.7}$$

(ii) The frictional acceleration is proportional to M and thus the frictional force must be proportional to M^2. It is instructive to consider why

this is so. Stars are deflected by M in such a way that the density of background stars behind M is greater than in front of it (see Figure 8.2 and Problem 8.4). The amplitude of this density enhancement or wake is proportional to M and the gravitational force that it exerts on M is proportional to M times its amplitude. Hence the force is proportional to M^2.

The validity of Chandrasekhar's formula Although Chandrasekhar's formula (8.3) was derived for a mass moving through an infinite homogeneous background, it can be (and usually is) employed to estimate the drag on a small body traveling through a much larger host system. In such applications we replace $f(v)$ by the value of the DF in the vicinity of the small body, $v_{\rm typ}$ by the local velocity dispersion, and $b_{\rm max}$ by the distance of the subject body from the center of the host. When employed in this way, Chandrasekhar's formula suffers from several internal inconsistencies:

(i) The choices of $b_{\rm max}$ and $v_{\rm typ}$ are rather arbitrary.
(ii) We have neglected the self-gravity of the wake. Thus equation (8.3) takes into account the mutual attraction of M and the background stars, but neglects the attraction of the background stars for each other.
(iii) We obtained equation (8.3) in the approximation that stars move past M on Keplerian hyperbolae. Orbits in the combined gravitational fields of M and the host system would be more complex.

These deficiencies become especially worrisome when M is so large as to be comparable to the mass of the host system that lies interior to M's orbit. Nevertheless, N-body simulations and linearized response calculations show that Chandrasekhar's formula provides a remarkably accurate description of the drag experienced by a body orbiting in a stellar system, usually within a factor of two and often considerably better (Weinberg 1989; Fujii, Funato, & Makino 2006).

The fundamental reasons for this success were discussed in the derivation of the Fokker–Planck equation in §7.4.2, and derive from the large ratio between the maximum and minimum impact parameters that contribute to the Coulomb logarithm $\ln\Lambda = \ln(b_{\rm max}/b_{90})$. Consider, for example, a black hole of mass $M = 10^6\,\mathcal{M}_\odot$, orbiting at radius 1 kpc in a galaxy with velocity dispersion $200\,{\rm km\,s^{-1}}$. Then we may set $b_{\rm max} \approx 1\,{\rm kpc}$ and $V_0 \approx 200\,{\rm km\,s^{-1}}$, so $b_{90} = 0.1\,{\rm pc}$ and $\ln\Lambda = 9.2$. To address the seriousness of problem (i) above, suppose that we have overestimated $b_{\rm max}$ by a factor of two, so the correct value is only half the orbital radius or 0.5 kpc; then $\ln\Lambda = 8.5$, a change of less than 10%. In words, the drag force is insensitive to changes of order unity in $b_{\rm max}$ and $v_{\rm typ}$, because $\ln\Lambda$ is large. To address problems (ii) and (iii) we note the effects of self-gravity are important only on scales comparable to the Jeans length, which in turn is comparable to $b_{\rm max}$. Thus the effects of self-gravity are negligible, and the approximation of a Keplerian hyperbola should be valid, for encounters with impact parameter much less than $b_{\rm max}$. Suppose then that we consider only encounters with $b < 100\,{\rm pc}$

or 10% of the orbital radius. The contribution to the Coulomb logarithm from these encounters is $\ln(100\,\mathrm{pc}/b_{90}) = 5.5$, a difference of only 25% from our original estimate. In words, most of the total contribution to the drag comes from encounters with sufficiently small impact parameters that the neglect of self-gravity and the approximation of Keplerian orbits introduce negligible errors.

A more sophisticated treatment of dynamical friction that avoids the inconsistencies of Chandrasekhar's formula requires the machinery of linear response theory that was developed in §5.3. The subject body is regarded as an external potential $\Phi_{\rm e}(\mathbf{x}, t)$ that excites a response density in the host system, governed by the response function $R(\mathbf{x}, \mathbf{x}', \tau)$. We then solve Poisson's equation to determine the gravitational potential generated by the response density, and the force exerted on the subject body by this response potential is dynamical friction (Weinberg 1986, 1989).

Like Landau damping, dynamical friction illustrates the curious fact that irreversible processes can occur in a system with reversible equations of motion. We have seen in §5.5.3 that Landau damping in spherical stellar systems arises from resonances between the oscillations of the system and the orbital frequencies of individual stars. Similarly, dynamical friction can be shown to arise from resonances between the orbital frequencies of the subject body and the stars (Tremaine & Weinberg 1984b). The rich Fourier spectrum of the gravitational potential from an orbiting point mass ensures that many orbital resonances contribute to the drag force, and the cumulative effect of these many weak resonances gives rise to the Coulomb logarithm in Chandrasekhar's formula.

8.1.1 Applications of dynamical friction

(a) Decay of black-hole orbits The centers of galaxies often contain black holes with masses 10^6–$10^9\,\mathcal{M}_\odot$ (§1.1.6). It is natural to ask whether such objects could be also be present at other locations in the galaxy, where they would be even harder to find. To investigate this question, we imagine a black hole of mass M on a circular orbit of radius r, and ask how long is needed for dynamical friction to drag the black hole to the galaxy center.

The flatness of many observed rotation curves suggests that we approximate the density distribution by a singular isothermal sphere (eq. 4.103),

$$\rho(r) = \frac{v_{\rm c}^2}{4\pi G r^2}, \tag{8.8}$$

where $v_{\rm c} = \sqrt{2}\sigma$ is the constant circular speed (eq. 4.104). The DF of the isothermal sphere is Maxwellian, so equation (8.7) gives the frictional force

$\widetilde{F} = M|\mathrm{d}\mathbf{v}_M/\mathrm{d}t|$ on the black hole:

$$\begin{aligned}\widetilde{F} &= \frac{4\pi G^2 M^2 \rho(r)\ln\Lambda}{v_\mathrm{c}^2}\left[\mathrm{erf}(X) - \frac{2X}{\sqrt{\pi}}\mathrm{e}^{-X^2}\right] \\ &= 0.428\ln\Lambda\frac{GM^2}{r^2},\end{aligned} \tag{8.9}$$

where $X = v_\mathrm{c}/(\sqrt{2}\sigma) = 1$.

This force is tangential and directed opposite to the velocity of the black hole, causing it to lose angular momentum $\widetilde{L}$ at a rate

$$\frac{\mathrm{d}\widetilde{L}}{\mathrm{d}t} = -\widetilde{F}r \simeq -0.428\ln\Lambda\,\frac{GM^2}{r}. \tag{8.10}$$

Thus the black hole spirals towards the center of the galaxy, while remaining on a nearly circular orbit. Since the circular-speed curve of the singular isothermal sphere is flat, the black hole continues to orbit at speed v_c as it spirals inward, so its angular momentum at radius r is $\widetilde{L} = Mrv_\mathrm{c}$. Substituting the time derivative of this expression into equation (8.10), we obtain

$$r\frac{\mathrm{d}r}{\mathrm{d}t} = -0.428\ln\Lambda\frac{GM}{v_\mathrm{c}} = -0.302\ln\Lambda\frac{GM}{\sigma}. \tag{8.11}$$

If we neglect the slow variation of $\ln\Lambda$ with radius, we can solve this differential equation subject to the initial condition that the radius is r_i at zero time. We find that the black hole reaches the center after a time[4]

$$t_\mathrm{fric} = \frac{1.65}{\ln\Lambda}\frac{r_\mathrm{i}^2\sigma}{GM} = \frac{19\,\mathrm{Gyr}}{\ln\Lambda}\left(\frac{r_\mathrm{i}}{5\,\mathrm{kpc}}\right)^2\frac{\sigma}{200\,\mathrm{km\,s^{-1}}}\frac{10^8\,\mathcal{M}_\odot}{M}. \tag{8.12}$$

This equation can be cast into a simpler form using the crossing time $t_\mathrm{cross} = r_\mathrm{i}/v_\mathrm{c}$, the time required for the subject body to travel one radian,

$$t_\mathrm{fric} = \frac{1.17}{\ln\Lambda}\frac{\mathcal{M}(r)}{M}t_\mathrm{cross}, \tag{8.13}$$

where $\mathcal{M}(r) = v_\mathrm{c}^2 r/G$ is the mass of the host galaxy contained within radius r. This result is approximately correct even for mass distributions other than the singular isothermal sphere; in words, if the ratio of the mass of the subject body to the interior mass of the host is $\mu \ll 1$, then the subject body spirals to the center of the host in roughly $1/(\mu\ln\Lambda)$ initial crossing times.

For characteristic values $b_\mathrm{max} \approx 5\,\mathrm{kpc}$, $M = 10^8\,\mathcal{M}_\odot$, and $v_\mathrm{typ} \approx \sigma = 200\,\mathrm{km\,s^{-1}}$, we have by equation (8.1b) that $\ln\Lambda \simeq 6$. Thus for the standard

[4] Equation (8.8) overestimates the density inside the galaxy's core, but this leads to a negligible error in the inspiral time, since the decay is rapid at small radii anyway.

parameters in equation (8.12), the inspiral time $t_{\rm fric}$ is only 3 Gyr. Black holes on eccentric orbits have even shorter inspiral times than those on circular orbits with the same mean radius, since the eccentric orbit passes through regions of higher density where the drag force is stronger. We conclude that any $10^8\,\mathcal{M}_\odot$ black hole that is formed within $\sim 10\,$kpc of the center of a typical galaxy will spiral to the center within the age of the universe. Thus massive black holes should normally be found at the center of the galaxy, unless they are far out in the galactic halo.

(b) Galactic cannibalism Most large galaxies are accompanied by several satellite galaxies, small companion galaxies that travel on bound orbits in the gravitational potential of the larger host. The satellites of our own Milky Way galaxy include the Sagittarius dwarf galaxy, the Large and Small Magellanic Clouds (§1.1.3 and Plate 11), and several dozen even smaller galaxies at distances of $\sim$ 100–300 kpc. Two satellite galaxies of the nearby disk galaxy M31 appear in Plate 3.

Satellites orbiting within the extended dark halo of their host experience dynamical friction, leading to orbital decay. As the satellite orbit decays, tidal forces from the host galaxy (§8.3) strip stars from the outer parts of the satellite, until eventually the entire satellite galaxy is disrupted—this process, in which a galaxy consumes its smaller neighbors, is an example of a minor merger, or, more colorfully, **galactic cannibalism**.

The rate of orbital decay for a satellite of fixed mass M is described approximately by equation (8.12). This formula does not, however, allow for mass loss due to tidal stripping as the satellite spirals inward. To account crudely for this process, we shall refer forward to §8.3, in which we show that the outer or tidal radius of a satellite is given approximately by its Jacobi radius $r_{\rm J}$, defined by equation (8.91). Once again we assume that the host galaxy is a singular isothermal sphere, so its mass interior to radius r is $\mathcal{M}(r) = v_{\mathcal{M}}^2 r/G = 2\sigma_{\mathcal{M}}^2 r/G$, where $v_{\mathcal{M}}$ and $\sigma_{\mathcal{M}} = v_{\mathcal{M}}/\sqrt{2}$ are the circular speed and velocity dispersion of the host; in this case equations (8.91) and (8.108) yield

$$r_{\rm J} = \left(\frac{M}{2\mathcal{M}(r)}\right)^{1/3} r = \left(\frac{GMr^2}{4\sigma_{\mathcal{M}}^2}\right)^{1/3}. \tag{8.14}$$

We shall assume that the satellite galaxy is also a singular isothermal sphere, but one that is sharply truncated at $r_{\rm J}$. Thus the total mass of the satellite is $M = 2\sigma_{\rm s}^2 r_{\rm J}/G$, where $\sigma_{\rm s}$ is its velocity dispersion. (A truncated isothermal sphere is not a self-consistent solution of the collisionless Boltzmann and Poisson equations, so this should be regarded as a fitting formula without much dynamical significance.) Equation (8.14) then yields

$$r_{\rm J} = \frac{\sigma_{\rm s}}{\sqrt{2}\sigma_{\mathcal{M}}} r \quad \text{which implies that} \quad M = \frac{\sqrt{2}\sigma_{\rm s}^3 r}{G\sigma_{\mathcal{M}}}. \tag{8.15}$$

Substituting into equation (8.11), we obtain the rate of orbital decay,

$$\frac{\mathrm{d}r}{\mathrm{d}t} = -0.428 \ln\Lambda \frac{\sigma_\mathrm{s}^3}{\sigma_\mathcal{M}^2}, \tag{8.16}$$

and neglecting the slow variation in $\ln\Lambda$ with radius, we find the inspiral time from radius r_i to be

$$\begin{aligned} t_\mathrm{fric} &= \frac{2.34}{\ln\Lambda}\frac{\sigma_\mathcal{M}^2}{\sigma_\mathrm{s}^3} r_\mathrm{i} \\ &= \frac{2.7\,\mathrm{Gyr}}{\ln\Lambda}\frac{r_\mathrm{i}}{30\,\mathrm{kpc}}\left(\frac{\sigma_\mathcal{M}}{200\,\mathrm{km\,s^{-1}}}\right)^2\left(\frac{100\,\mathrm{km\,s^{-1}}}{\sigma_\mathrm{s}}\right)^3. \end{aligned} \tag{8.17}$$

To evaluate the Coulomb logarithm, we use equation (8.2). The half-mass radius r_h of the satellite is half of its Jacobi radius, and the typical velocity may be taken to be $v_\mathrm{typ} = \sigma_\mathcal{M}$. Then the two quantities in the denominator of equation (8.2) are given by equations (8.15),

$$r_\mathrm{h} = \frac{\sigma_\mathrm{s}}{2^{3/2}\sigma_\mathcal{M}} r \quad ; \quad \frac{GM}{v_\mathrm{typ}^2} = \frac{\sqrt{2}\sigma_\mathrm{s}^3}{\sigma_\mathcal{M}^3} r. \tag{8.18}$$

The velocity dispersion of a galaxy is correlated with its mass through the Faber–Jackson law (1.21). Satellite galaxies have smaller luminosities than their hosts, and hence smaller dispersions. If $\sigma_\mathrm{s} \lesssim 0.5\sigma_\mathcal{M}$, the first term in equation (8.18) is larger than the second, so the argument of the Coulomb logarithm is $\Lambda = b_\mathrm{max}/r_\mathrm{h}$; setting $b_\mathrm{max} = r$ we have finally $\Lambda = 2^{3/2}\sigma_\mathcal{M}/\sigma_\mathrm{s}$. Thus, for example, equation (8.17) implies that in a host galaxy with dispersion $200\,\mathrm{km\,s^{-1}}$, a satellite galaxy with dispersion $\sigma \gtrsim 50\,\mathrm{km\,s^{-1}}$ will merge from a circular orbit with radius 30 kpc within 10 Gyr.

(c) Orbital decay of the Magellanic Clouds In general, the orbits of satellites of the Milky Way cannot be determined, because their velocities perpendicular to the line of sight are either unknown or have large observational uncertainties. However, much more information is available for the Large and Small Magellanic Clouds. Not only do we have good estimates for their velocities perpendicular to the line of sight (Kallivayalil et al. 2006), but the correct Cloud orbits must be able to reproduce the dynamics of the Magellanic Stream, a narrow band of neutral hydrogen gas that extends over 120° in the sky and is believed to have been torn off the Small Cloud by the gravitational field of the Galaxy about 1–1.5 Gyr ago. (See BM §8.4.1 and Putman et al. 2003 for a description of the observations.)

Several groups have modeled the dynamics of the Magellanic Stream and the resulting constraints on the Cloud orbits (Murai & Fujimoto 1980; Lin & Lynden–Bell 1982; Gardiner, Sawa, & Fujimoto 1994; Connors et al. 2004). They find that the orbital plane of the Clouds is nearly perpendicular

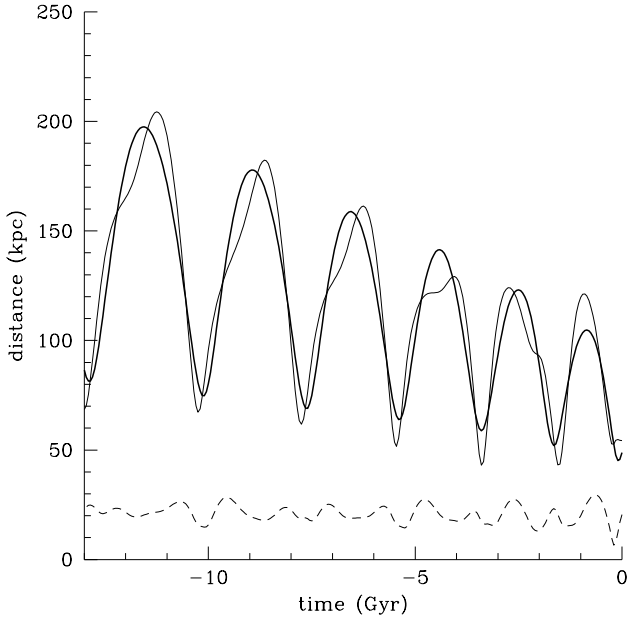

Figure 8.3 The decay of the orbits of the Magellanic Clouds around our Galaxy. The upper curves show the radius of the Clouds from the Galactic center (thick line for the Large Cloud and thin line for the Small Cloud), and the lower, dashed curve shows the distance between the Large and Small Cloud. The Galaxy potential is that of a singular isothermal sphere with circular speed $v_c = 220\,\mathrm{km\,s^{-1}}$, and the drag force is computed using Chandrasekhar's formula (8.7). The initial conditions at $t = 0$ are chosen to reproduce the observed distances and radial velocities of the Clouds and the kinematics of the Magellanic Stream (Gardiner, Sawa, & Fujimoto 1994).

to the Galactic plane; the sense of the orbit is such that the Clouds are approaching the Galactic plane with the Magellanic Stream trailing behind; the orbit is eccentric (the apocenter/pericenter distance is $\gtrsim 2$); and the Clouds are presently near pericenter (Figure 8.3). As seen in the figure, the orbits of the Magellanic Clouds are decaying due to dynamical friction. The ongoing mass loss from the Clouds that generates the Magellanic Stream provides circumstantial evidence that the orbit is continuing to shrink.

In this model the Clouds merge with the Milky Way in about 6 Gyr, although the model is unrealistic beyond about 3 Gyr in the future, when the Galaxy experiences a much more violent merger with M31 (Box 3.1).

(d) Dynamical friction on bars Dynamical friction can be generated by any time-varying large-scale gravitational field. An important example is the interaction between a bar in a disk galaxy and the surrounding dark halo. As a first approximation, let us think of the bar as a rigid body, consisting

of two masses M at the ends of a rod of length $2r$ that revolves around its center at the bar pattern speed $\Omega_{\rm b}$. In a strong bar M would not be much smaller than the mass of the galaxy interior to r. In this circumstance equation (8.13) suggests that the bar should lose its angular momentum in a few crossing times, which is much shorter than the age of the galaxy. Thus we might expect that bars in disk galaxies with massive halos would have zero angular momentum and zero pattern speed.

Improving on this crude model is a challenging analytic task, for several reasons: first, the gravitational potential of a bar is more complicated than the potential from a point mass; second, in contrast to most orbiting bodies, bars are extended objects, so the friction is not dominated by local encounters; third, dynamical friction exerts a torque on the bar but we do not understand the reaction of the bar to that torque: does its pattern speed increase or decrease? does the bar grow stronger or weaker? etc.

Accurate analytic determinations of the frictional torque on a bar from the dark halo can be derived using perturbation theory (Weinberg 1985), but N-body simulations can be more informative because they determine both the torque on the bar and its resulting evolution. Simulations confirm that the halo exerts a strong frictional torque on the bar, and show that in response the bar pattern speed rapidly decays but the bar remains intact (Sellwood 1980; Hernquist & Weinberg 1992; Debattista & Sellwood 1998, 2000).

These theoretical results imply that if massive dark halos are present in the inner parts of barred galaxies, bars should rotate slowly. However, this conclusion is inconsistent with observations: as we saw in §6.5.1a the ratio $\mathcal{R}$ of the corotation radius to the bar semi-major axis (eq. 6.103) generally lies in the range 0.9–1.3, where $\mathcal{R} \simeq 1$ is the maximum allowed rotation rate for a weak bar. This problem can be resolved if spiral galaxies have maximum disks (§6.3.4), for then the halo mass is relatively small in the inner few kpc, where interactions with the bar are strongest.

(e) Formation and evolution of binary black holes Since most galaxies contain black holes at their centers, it is natural to ask what happens to the black holes when a satellite galaxy merges with a larger host.

As the satellite's orbit decays, its stars are stripped by tidal forces that become stronger and stronger as the orbit shrinks (eq. 8.15), until eventually only its central black hole is left. The orbit of the black hole continues to decay from dynamical friction, although at a slower rate since the mass of the black hole is only a small fraction of the mass of the original satellite galaxy. Assuming that the host galaxy also contains a central black hole, we expect that eventually the two black holes will form a bound binary system.

After the black-hole binary is formed, its orbit continues to decay by dynamical friction. Equation (8.3) still describes the drag acting on each black hole, with the maximum impact parameter $b_{\rm max}$ appearing in the Coulomb logarithm now equal to the binary semi-major axis a.

As the binary orbit shrinks, the relative orbital velocity v of the two black holes grows. Eventually the orbital velocity greatly exceeds the velocity dispersion σ of the stars in the galaxy. For a circular orbit, this occurs when the binary semi-major axis a satisfies

$$\frac{G(M_1+M_2)}{a} \gg \sigma^2 \quad \text{or} \quad a \ll 10\,\mathrm{pc}\,\frac{M_1+M_2}{10^8\,\mathcal{M}_\odot}\left(\frac{200\,\mathrm{km\,s^{-1}}}{\sigma}\right)^2, \tag{8.19}$$

where M_1 and M_2 are the masses of the black holes. Following the discussion in §7.5.7, we shall say that the black-hole binary is hard when $v > \sigma$.

For hard binaries Chandrasekhar's dynamical friction formula is no longer valid, but an approximate formula for the rate of orbital decay can be derived by arguments similar to those used to derive the hardening rate for binary stars in equation (7.179). These yield (Quinlan 1996b)

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{a}\right) = -C\frac{G\rho}{\sigma}, \qquad C = 14.3, \tag{8.20}$$

where ρ is the density of stars in the vicinity of the binary. This result is almost independent of the eccentricity of the binary and depends only weakly on the mass ratio M_2/M_1 so long as it is not too far from unity.[5]

Under the assumption that the galaxy has a constant-density core, we can integrate equation (8.20) to obtain $1/a = \mathit{constant} - CG\rho t/\sigma$. Choosing the origin of time so that the constant is zero, we obtain $a = \sigma/(CG\rho t)$. The King radius of the galaxy, r_0, is related to ρ and σ via $4\pi G\rho r_0^2 = 9\sigma^2$ (eq. 4.106). Eliminating ρ with the help of this equation, we have finally

$$a(t) = \frac{4\pi r_0^2}{9C\sigma t} = 0.005\,\mathrm{pc}\,\frac{200\,\mathrm{km\,s^{-1}}}{\sigma}\left(\frac{r_0}{100\,\mathrm{pc}}\right)^2\frac{\mathrm{Gyr}}{t}. \tag{8.21}$$

Thus interactions with stars in the host galaxy can drive the black-hole binary to semi-major axes as small as a few milliparsecs; the corresponding relative speed for a circular orbit is

$$v = \sqrt{\frac{G(M_1+M_2)}{a}} = 2.1\times10^4\,\mathrm{km\,s^{-1}}\left(\frac{M_1+M_2}{10^8\,\mathcal{M}_\odot}\frac{10^{-3}\,\mathrm{pc}}{a}\right)^{1/2}. \tag{8.22}$$

There is an important case in which this analysis fails. Only stars with angular momentum $L \lesssim [G(M_1+M_2)a]^{1/2}$ interact strongly with the binary, and if the binary semi-major axis a is much smaller than the King radius r_0 then this is much smaller than the typical angular momentum $L \sim r_0\sigma$ of

[5] The numerical coefficient differs from the one in equation (7.179) because here the binary components are much more massive than the field stars, while in equation (7.179) the binary components and the field stars all have the same mass.

stars in the core. Thus, only a small fraction of the stars in the core interact strongly with the black holes. The region in phase space with such small angular momentum is called the loss cone, by analogy with the loss cone from which stars are consumed by a single black hole (§7.5.9). The mass of stars in the loss cone shrinks as the binary semi-major axis decreases, and eventually may become smaller than the black-hole mass. In this case the binary can empty the loss cone, and the shrinkage of the semi-major axis will stall. Once the loss cone has been emptied, the rate of continued evolution is much less certain, being determined by the rate at which the loss cone is slowly refilled by two processes: diffusion of angular momentum due to two-body relaxation (Chapter 7), or torques from the host galaxy, if its overall mass distribution is non-spherical (Yu 2002; Makino & Funato 2004).

If the binary semi-major axis shrinks far enough, gravitational radiation takes over as the dominant cause of orbital decay. A binary black hole on a circular orbit with semi-major axis a will coalesce under the influence of gravitational radiation in a time (Peters 1964)

$$\begin{aligned} t_{\rm gr} &= \frac{5c^5a^4}{256G^3M_1M_2(M_1+M_2)} \\ &= 5.81\,{\rm Myr}\left(\frac{a}{0.01\,{\rm pc}}\right)^4\left(\frac{10^8\,\mathcal{M}_\odot}{M_1+M_2}\right)^3\frac{(M_1+M_2)^2}{M_1M_2}. \end{aligned} \tag{8.23}$$

The characteristic decay time due to gravitational radiation therefore scales as a^4. In contrast, the decay time $({\rm d}\ln a/{\rm d}t)^{-1}$ due to dynamical friction varies as $1/a$. Consequently, the actual decay time, which is set by the more efficient of the two processes, has a maximum at the semi-major axis where the two decay times are equal (Begelman, Blandford & Rees 1980). This radius is referred to as the **bottleneck radius**, and lies between 0.003 pc and 3 pc depending on the galaxy density distribution and black-hole masses (Yu 2002). The bottleneck radius is where binary black holes are most likely to be found.

The decay time at the bottleneck is quite uncertain, since it depends on both the extent to which the loss cone is depopulated, and the possible contribution of gas drag. If the bottleneck decay time exceeds 10 Gyr, most galaxies should contain binary black holes at their centers. If the decay time is less than 10 Gyr, most black-hole binaries will eventually coalesce. Coalescing black holes are of great interest because they generate strong bursts of gravitational radiation that should ultimately be detectable, even at cosmological distances, and thus provide a unique probe of both galaxy evolution and general relativity.

(f) Globular clusters These systems may experience significant orbital decay from dynamical friction. The rate of decay and inspiral time can be described approximately by equations (8.11) and (8.12). For a typical cluster

mass $M = 2 \times 10^5\,\mathcal{M}_\odot$ (Table 1.3) the inspiral time from radius r_i is

$$t_\mathrm{fric} = 64\,\mathrm{Gyr}\frac{\sigma}{200\,\mathrm{km\,s^{-1}}}\left(\frac{r_\mathrm{i}}{1\,\mathrm{kpc}}\right)^2, \tag{8.24}$$

where σ is the velocity dispersion of the host galaxy and we have assumed $\ln\Lambda = 5.8$, from equation (8.2) with $b_\mathrm{max} = 1\,\mathrm{kpc}$ and $r_\mathrm{h} = 3\,\mathrm{pc}$ (Table 1.3). Orbital decay is most important for low-luminosity host galaxies, which have small radii and low velocity dispersions. Many dwarf elliptical galaxies exhibit a deficit of clusters near their centers and compact stellar nuclei, which may arise from the inspiral and merger of these clusters (Lotz et al. 2001). A puzzling exception is the Fornax dwarf spheroidal galaxy, a satellite of the Milky Way, which contains five globular clusters despite an estimated inspiral time of only $t_\mathrm{fric} \sim 1\,\mathrm{Gyr}$ (Tremaine 1976a). Why these clusters have not merged at the center of Fornax remains an unsolved problem.

8.2 High-speed encounters

One of the most important classes of interaction between stellar systems is high-speed encounters. By "high-speed" we mean that the duration of the encounter—the interval during which the mutual gravitational forces are significant—is short compared to the crossing time within each system. A typical example is the collision of two galaxies in a rich cluster of galaxies (§1.1.5). The duration of the encounter is roughly the time it takes the two galaxies to pass through one another; given a galaxy size $r \sim 10\,\mathrm{kpc}$ and the typical encounter speed in a rich cluster, $V \approx 2000\,\mathrm{km\,s^{-1}}$, the duration is $r/V \approx 5\,\mathrm{Myr}$. For comparison the internal dispersion of a large galaxy is $\sigma \approx 200\,\mathrm{km\,s^{-1}}$ so the crossing time is $r/\sigma \approx 50\,\mathrm{Myr}$, a factor of ten larger.

As we saw at the beginning of this chapter, the effect of an encounter on the internal structure of a stellar system *decreases* as the encounter speed increases. Hence high-speed encounters can be treated as small perturbations of otherwise steady-state systems.

We consider an encounter between a stellar system of mass M_s, the subject system, and a passing perturber—a galaxy, gas cloud, dark halo, black hole, etc.—of mass M_p. At the instant of closest approach, the centers of the subject system and the perturber are separated by distance b and have relative speed V. If the relative speed is high enough, then:

(i) The kinetic energy of relative motion of the two systems is much larger than their mutual potential energy, so the centers travel at nearly uniform velocity throughout the encounter.

(ii) In the course of the encounter, the majority of stars will barely move from their initial locations with respect to the system center. Thus the

gravitational force from the perturber can be approximated as an impulse of very short duration, which changes the velocity but not the position of each star. A variety of analytic arguments (see page 658) and numerical experiments (Aguilar & White 1985) suggest that this **impulse approximation** yields remarkably accurate results, even when the duration of the encounter is almost as long as the crossing time.[6]

We now ask how the structure of the subject system is changed by the passage of the perturber. We work in a frame that is centered on the center of mass of the subject system before the encounter. Let m_α be the mass of the αth star of the subject system, and let $\dot{\mathbf{v}}'_\alpha$ be the rate of change in its velocity due to the force from the perturber. We break $\dot{\mathbf{v}}'_\alpha$ into two components. The component that reflects the rate of change of the center-of-mass velocity of the subject system is

$$\dot{\mathbf{v}}_{\rm cm} \equiv \frac{1}{M_{\rm s}} \sum_\beta m_\beta \dot{\mathbf{v}}'_\beta, \quad \text{where} \quad M_{\rm s} \equiv \sum_\beta m_\beta \tag{8.25a}$$

is the mass of the subject system. The component

$$\dot{\mathbf{v}}_\alpha \equiv \dot{\mathbf{v}}'_\alpha - \dot{\mathbf{v}}_{\rm cm} \tag{8.25b}$$

gives the acceleration of the αth star with respect to the center of mass.

Let $\Phi(\mathbf{x}, t)$ be the gravitational potential due to the perturber. Then

$$\dot{\mathbf{v}}'_\alpha = -\boldsymbol{\nabla}\Phi(\mathbf{x}_\alpha, t), \tag{8.26}$$

and equation (8.25b) can be written

$$\dot{\mathbf{v}}_\alpha = -\boldsymbol{\nabla}\Phi(\mathbf{x}_\alpha, t) + \frac{1}{M_{\rm s}} \sum_\beta m_\beta \boldsymbol{\nabla}\Phi(\mathbf{x}_\beta, t). \tag{8.27}$$

In the impulse approximation, $\mathbf{x}_\alpha$ is constant during an impulsive encounter, so

$$\Delta\mathbf{v}_\alpha = \int_{-\infty}^{\infty} {\rm d}t\, \dot{\mathbf{v}}_\alpha = \int_{-\infty}^{\infty} {\rm d}t \left[-\boldsymbol{\nabla}\Phi(\mathbf{x}_\alpha, t) + \frac{1}{M_{\rm s}} \sum_\beta m_\beta \boldsymbol{\nabla}\Phi(\mathbf{x}_\beta, t) \right]. \tag{8.28}$$

The potential energy of the subject system does not change during the encounter, so in the center-of-mass frame the change in the energy, $\widetilde{E}$, is simply

[6] Condition (ii) almost always implies condition (i), but condition (i) need not imply condition (ii): for example, a star passing by the Sun at a relative velocity $v \simeq 50\,{\rm km\,s^{-1}}$ and an impact parameter $b \simeq 0.01\,{\rm pc}$ will hardly be deflected at all and hence satisfies condition (i). However, the encounter time $b/v \simeq 200\,{\rm yr}$ is much larger than the orbital period of most of the planets so the encounter is adiabatic, rather than impulsive.

the change in the internal kinetic energy, $\widetilde{K}$. Here tildes on the symbols are a reminder that these quantities have units of mass $\times$(velocity)^2, in contrast to the usual practice in this book where E and K denote energy per unit mass. We have

$$\begin{aligned}\Delta\widetilde{E} = \Delta\widetilde{K} &= \tfrac{1}{2}\sum_\alpha m_\alpha \left[(\mathbf{v}_\alpha + \Delta\mathbf{v}_\alpha)^2 - \mathbf{v}_\alpha^2\right] \\ &= \tfrac{1}{2}\sum_\alpha m_\alpha \left[|\Delta\mathbf{v}_\alpha|^2 + 2\mathbf{v}_\alpha\cdot\Delta\mathbf{v}_\alpha\right].\end{aligned} \tag{8.29}$$

In any static axisymmetric system, $\sum_\alpha m_\alpha \mathbf{v}_\alpha \cdot \Delta\mathbf{v}_\alpha = 0$ by symmetry (see Problem 8.5). Thus the energy changes that are first-order in the small quantity $\Delta\mathbf{v}$ average to zero, and the change of internal energy is given by the second-order quantity

$$\Delta\widetilde{E} = \Delta\widetilde{K} = \tfrac{1}{2}\sum_\alpha m_\alpha |\Delta\mathbf{v}_\alpha|^2. \tag{8.30}$$

This simple derivation masks several subtleties:

(a) Mass loss Equation (8.29) shows that the encounter redistributes a portion of the system's original energy stock: stars in which $\mathbf{v}_\alpha\cdot\Delta\mathbf{v}_\alpha > 0$ gain energy, while those with $\mathbf{v}_\alpha\cdot\Delta\mathbf{v}_\alpha < 0$ may lose energy. The energy gained by some stars may be so large that they escape from the system, and then the overall change in energy of the stars that remain bound can be negative. Thus the energy per unit mass of the bound remnant system may decrease (become more negative) as the result of the encounter, even though the encounter always adds energy to the original system.

(b) Return to equilibrium After the increments (8.28) have been added to the velocities of all the stars of the subject system, it no longer satisfies the virial theorem (4.250). Hence the encounter initiates a period of readjustment, lasting a few crossing times, during which the subject system settles to a new equilibrium configuration.

If the perturbation is weak enough that no stars escape, some properties of this new equilibrium can be deduced using the virial theorem. Let the initial internal kinetic and total energies be $\widetilde{K}_0$ and $\widetilde{E}_0$, respectively. Then the virial theorem implies that

$$\widetilde{K}_0 = -\widetilde{E}_0. \tag{8.31}$$

Since the impulsive encounter increases the kinetic energy by $\Delta\widetilde{K}$ and leaves the potential energy unchanged, the final energy is

$$\widetilde{E}_1 = \widetilde{E}_0 + \Delta\widetilde{K}. \tag{8.32}$$

Once the subject system has settled to a new equilibrium state, the final kinetic energy is given by the virial theorem,

$$\widetilde{K}_1 = -\widetilde{E}_1 = -(\widetilde{E}_0 + \Delta\widetilde{K}) = \widetilde{K}_0 - \Delta\widetilde{K}. \tag{8.33}$$

Thus if the impulsive encounter *increases* the kinetic energy by $\Delta\widetilde{K}$, the subsequent relaxation back to dynamical equilibrium *decreases* the kinetic energy by $2\Delta\widetilde{K}$!

(c) Adiabatic invariance The impulse approximation is valid only if the encounter time is short compared to the crossing time. In most stellar systems the crossing time is a strong function of energy or mean orbital radius, so the impulse approximation is unlikely to hold for stars near the center. Indeed, sufficiently close to the center, the crossing times of most stars may be so short that their orbits deform adiabatically as the perturber approaches (§3.6.2c). In this case, changes that occur in the structure of the orbits as the perturber approaches will be reversed as it departs, and the encounter will leave most orbits in the central region unchanged.

If we approximate the potential near the center of the stellar system as that of a harmonic oscillator with frequency Ω, then the energy change imparted to the stars in an encounter of duration τ is proportional to $\exp(-\alpha\Omega\tau)$ for $\Omega\tau \gg 1$, where α is a constant of order unity (see §3.6.2a). However, this strong exponential dependence does not generally hold in realistic stellar systems. The reason is that some of the stars are in resonance with the slowly varying perturbing force, in the sense that $\mathbf{m}\cdot\boldsymbol{\Omega} \simeq 0$ where the components of $\boldsymbol{\Omega}(\mathbf{J})$ are the fundamental frequencies of the orbit (eq. 3.190), and $\mathbf{m}$ is an integer triple. At such a resonance, the response to a slow external perturbation is large—in the language of Chapter 5, the polarization matrix diverges (eq. 5.95). A careful calculation of the contribution of both resonant and non-resonant stars shows that the total energy change in an encounter generally declines only as $(\Omega\tau)^{-1}$ for $\Omega\tau \gg 1$, rather than exponentially (Weinberg 1994a).

8.2.1 The distant-tide approximation

The calculation of the effects of an encounter simplifies considerably when the size of the subject system is much less than the impact parameter.

Let $\Phi(\mathbf{x}, t)$ be the gravitational potential of the perturber, in a frame in which the center of mass of the subject system is at the origin. When the distance to the perturber is much larger than the size of the subject system, the perturbing potential will vary smoothly across it, and we may therefore expand the field $-\boldsymbol{\nabla}\Phi(\mathbf{x}, t)$ in a Taylor series about the origin:

$$-\frac{\partial\Phi}{\partial x_j}(\mathbf{x}, t) = -\Phi_j(t) - \sum_k \Phi_{jk}(t)x_k + \mathrm{O}(|\mathbf{x}|^2), \tag{8.34a}$$

where $\mathbf{x} = (x_1, x_2, x_3)$ and

$$\Phi_j \equiv \left.\frac{\partial \Phi}{\partial x_j}\right|_{\mathbf{x}=0} \quad ; \quad \Phi_{jk} \equiv \left.\frac{\partial^2 \Phi}{\partial x_j \partial x_k}\right|_{\mathbf{x}=0}. \tag{8.34b}$$

Dropping the terms $\mathrm{O}(|\mathbf{x}|^2)$ constitutes the **distant-tide approximation**. Encounters for which the both the distant-tide and impulse approximations are valid are often called **tidal shocks**.

Substituting into equations (8.25a) and (8.26), we find that Φ_{jk} does not contribute to the center-of-mass acceleration $\dot{\mathbf{v}}_{\rm cm}$, because the center of mass is at the origin so $\sum_\beta m_\beta \mathbf{x}_\beta = 0$. Similarly, substituting (8.34a) into (8.27), we find that Φ_j does not contribute to $\dot{\mathbf{v}}_\alpha$ because $\sum_\beta m_\beta = M_{\rm s}$. Thus

$$\dot{\mathbf{v}}_\alpha = -\sum_{j,k=1}^{3} \hat{\mathbf{e}}_j \Phi_{jk} x_{\alpha k}. \tag{8.35}$$

If the perturber is spherical and centered at $\mathbf{X}(t)$, we may write $\Phi(\mathbf{x}, t) = \Phi\left(|\mathbf{x} - \mathbf{X}(t)|\right)$ and (cf. Box 2.3)

$$\Phi_j = -\Phi' \frac{X_j}{X} \quad ; \quad \Phi_{jk} = \left(\Phi'' - \frac{\Phi'}{X}\right)\frac{X_j X_k}{X^2} + \frac{\Phi'}{X}\delta_{jk}, \tag{8.36}$$

where $X = |\mathbf{X}|$ and all derivatives of Φ are evaluated at X.

An important special case occurs when the impact parameter is large enough that we may approximate the perturber as a point mass $M_{\rm p}$. Then $\Phi(X) = -GM_{\rm p}/X$ and we have

$$\Phi_j = -\frac{GM_{\rm p}}{X^3} X_j \quad ; \quad \Phi_{jk} = \frac{GM_{\rm p}}{X^3}\delta_{jk} - \frac{3GM_{\rm p}}{X^5} X_j X_k. \tag{8.37}$$

Thus the equation of motion (8.35) becomes

$$\dot{\mathbf{v}}_\alpha = -\frac{GM_{\rm p}}{X^3}\mathbf{x}_\alpha + \frac{3GM_{\rm p}}{X^5}(\mathbf{X}\cdot\mathbf{x}_\alpha)\mathbf{X}. \tag{8.38}$$

We argued at the beginning of this section that in the impulse approximation, the orbit of the perturber can be assumed to have constant relative velocity $\mathbf{V}$. We align our coordinate axes so that $\mathbf{V}$ lies along the z axis, and the perturber's orbit lies in the yz plane, and choose the origin of time to coincide with the point of closest approach. Then $\mathbf{X}(t) = (0, b, Vt)$, where b is the impact parameter, and we have

$$\dot{\mathbf{v}}_\alpha = -\frac{GM_{\rm p}\mathbf{x}_\alpha}{[b^2 + (Vt)^2]^{3/2}} + \frac{3GM_{\rm p}(y_\alpha b + z_\alpha Vt)}{[b^2 + (Vt)^2]^{5/2}}(b\hat{\mathbf{e}}_y + Vt\hat{\mathbf{e}}_z). \tag{8.39}$$

In the impulse approximation, $\mathbf{x}_\alpha$ is constant during the encounter, so

$$\begin{aligned}\Delta\mathbf{v}_\alpha &= \int_{-\infty}^{\infty} \mathrm{d}t\, \dot{\mathbf{v}}_\alpha \\ &= GM_\mathrm{p} \int_{-\infty}^{\infty} \mathrm{d}t \left\{ -\frac{(x_\alpha, y_\alpha, z_\alpha)}{[b^2+(Vt)^2]^{3/2}} + 3(0, b, Vt)\frac{y_\alpha b + z_\alpha Vt}{[b^2+(Vt)^2]^{5/2}} \right\} \\ &= \frac{GM_\mathrm{p}}{b^2 V}\Bigg(- x_\alpha \int_{-\infty}^{\infty} \frac{\mathrm{d}u}{(1+u^2)^{3/2}},\ y_\alpha \int_{-\infty}^{\infty} \mathrm{d}u\, \frac{2-u^2}{(1+u^2)^{5/2}}, \\ &\qquad\qquad z_\alpha \int_{-\infty}^{\infty} \mathrm{d}u\, \frac{2u^2-1}{(1+u^2)^{5/2}}\Bigg),\end{aligned} \tag{8.40}$$

where we have made the substitution $u = Vt/b$. Evaluating the integrals in equation (8.40), we obtain finally

$$\Delta\mathbf{v}_\alpha = \frac{2GM_\mathrm{p}}{b^2 V}(-x_\alpha, y_\alpha, 0). \tag{8.41}$$

The error introduced in this formula by the distant-tide approximation is of order $|\mathbf{x}|/b \ll 1$. The velocity increments tend to deform a sphere of stars into an ellipsoid whose long axis lies in the direction of the perturber's point of closest approach. This distortion is reminiscent of the way in which the Moon raises tides on the surface of the oceans.

By equations (8.30) and (8.41) the change in the energy per unit mass in the distant-tide approximation is (Spitzer 1958)

$$\Delta\widetilde{E} = \frac{2G^2 M_\mathrm{p}^2}{V^2 b^4} \sum_\alpha m_\alpha (x_\alpha^2 + y_\alpha^2). \tag{8.42}$$

If the subject system is spherical, then $\sum m_\alpha x_\alpha^2 = \sum m_\alpha y_\alpha^2 = \frac{1}{3} M_\mathrm{s} \langle r^2 \rangle$, where $\langle r^2 \rangle$ is the mass-weighted mean-square radius of the stars in the subject system. In this case equation (8.42) simplifies to

$$\Delta\widetilde{E} = \frac{4G^2 M_\mathrm{p}^2 M_\mathrm{s}}{3V^2 b^4} \langle r^2 \rangle. \tag{8.43}$$

Equation (8.43) shows that for large impact parameter b the energy input in tidal shocks varies as b^{-4}. Thus the encounters that have the strongest effect on a stellar system are those with the smallest impact parameter b, which unfortunately are also those for which the approximation of a point-mass perturber is invalid. Fortunately, it is a straightforward numerical task to generalize these calculations to a spherical perturber with an arbitrary mass distribution, using equations (8.30), (8.35), and (8.36) (Aguilar & White 1985; Gnedin, Hernquist, & Ostriker 1999). Let $U(b/r_\mathrm{h})$ be the ratio

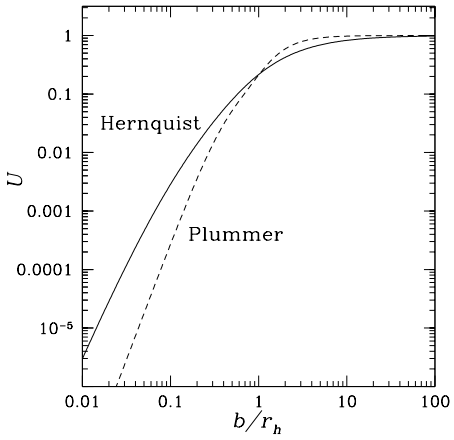

Figure 8.4 Energy input in a tidal shock due to a perturber with a Plummer or Hernquist mass distribution (eqs. 2.44b and 2.67). Here b is the impact parameter, $r_{\rm h}$ is the half-mass radius of the Plummer or Hernquist model, and U is the ratio of the energy input to that caused by a point mass perturber (eq. 8.44). The integral $W = \int {\rm d}x\, U(x)/x^3 = 0.5675$ for the Plummer model and 1.239 for the Hernquist model (eq. 8.52).

of the impulsive energy change caused by a perturber of half-mass radius $r_{\rm h}$ to the input from a point of the same total mass, which is given by (8.43). Then we have

$$\Delta\widetilde{E} = \frac{4G^2M_{\rm p}^2M_{\rm s}}{3V^2b^4}\,U(b/r_{\rm h})\langle r^2\rangle. \tag{8.44}$$

Figure 8.4 shows $U(x)$ for the Plummer and Hernquist mass distributions.

8.2.2 Disruption of stellar systems by high-speed encounters

In many cases we are interested in the cumulative effect of encounters on a stellar system that is traveling through a sea of perturbers. Let us assume that the perturbers have mass $M_{\rm p}$ and number density $n_{\rm p}$, and a Maxwellian DF with velocity dispersion $\sigma_{\rm p}$ in one dimension. Similarly, we assume that the subject system is a randomly chosen member of a population having a Maxwellian velocity distribution with dispersion $\sigma_{\rm s}$.

Consider the rate at which the subject system encounters perturbers at relative speeds in the range $(V, V+{\rm d}V)$ and impact parameters in the range $(b, b+{\rm d}b)$. With our assumption of Maxwellian velocity distributions, the distribution of relative velocities of encounters is also Maxwellian, with dispersion (Problem 8.8)

$$\sigma_{\rm rel} = (\sigma_{\rm s}^2 + \sigma_{\rm p}^2)^{1/2}. \tag{8.45}$$

Thus the probability that the subject system and perturber have relative speed in the given range is

$${\rm d}P = \frac{4\pi V^2{\rm d}V}{(2\pi\sigma_{\rm rel}^2)^{3/2}}\exp\left(-\frac{V^2}{2\sigma_{\rm rel}^2}\right), \tag{8.46}$$

and the average rate at which a subject system encounters perturbers with speed V and impact parameter b is

$$\dot{C} = n_{\rm p}\, V\, 2\pi b\, {\rm d}b\, {\rm d}P = \frac{2\sqrt{2\pi} n_{\rm p} b\, {\rm d}b}{\sigma_{\rm rel}^3} \exp\left(-\frac{V^2}{2\sigma_{\rm rel}^2}\right) V^3 {\rm d}V. \tag{8.47}$$

In the distant-tide approximation, the energy input to a star is proportional to the square of its radius (eq. 8.42). So we focus on stars in the outer parts of the subject system, for which we can assume that the gravitational potential of the subject system is Keplerian. In a Keplerian potential, the time averaged mean-square radius of an orbit with semi-major axis a and eccentricity e is $(1 + \frac{3}{2}e^2)a^2$ (Problem 3.9). If the DF of the subject system is isotropic in velocity space, the average of e^2 over all the stars with a given semi-major axis or energy is $\frac{1}{2}$ (Problem 4.8). Thus, if we average over stars with different orbital phases and eccentricities but the same energy, $\langle r^2\rangle = \frac{7}{4}a^2$, and equation (8.44) yields an average change in energy per unit mass of

$$\langle \Delta E\rangle = \frac{\langle \Delta \widetilde{E}\rangle}{M_{\rm s}} = \frac{7G^2 M_{\rm p}^2 a^2}{3V^2 b^4}\, U(b/r_{\rm h}). \tag{8.48}$$

From Figure 8.4 we see that for $b \gg r_{\rm h}$, $U \simeq 1$ so $\langle \Delta E\rangle \propto b^{-4}$ is a steeply declining function of impact parameter. The frequency of encounters with impact parameters in the range $(b, b+{\rm d}b)$ is proportional to $b\,{\rm d}b$ so the rate at which energy is injected by encounters in this range decreases with increasing b as ${\rm d}b/b^3$. On the other hand, U rapidly decreases with decreasing b once $b \lesssim r_{\rm h}$, with the result that encounters with impact parameters $b \sim r_{\rm h}$ inflict the most damage.

If the damage from a single encounter with $b \sim r_{\rm h}$ is not fatal for the system, we say that we are in the **diffusive regime** because the effects from a whole sequence of encounters will accumulate, as in the diffusive relaxation processes that we discussed in Chapter 7. If, by contrast, a single encounter at impact parameter $b \sim r_{\rm h}$ will shatter the system, the damage sustained by the system will be small until it is disrupted by a single, closest encounter, and we say that we are in the **catastrophic regime**.

(a) The catastrophic regime We first determine the largest impact parameter parameter b_1 at which a single encounter can disrupt the system. Since we are in the catastrophic regime, we may assume that $b_1 \gtrsim r_{\rm h}$ so the energy per unit mass injected by an encounter at impact parameter b_1 is given by equation (8.48) with $U(b/r_{\rm h}) \simeq 1$. Equating this to the absolute value of the energy of an individual star $E = -GM_{\rm s}/2a$ (eq. 3.32), we obtain

$$1 = \frac{\langle \Delta E\rangle}{|E|} = \frac{14 G M_{\rm p}^2 a^3}{3 M_{\rm s} V^2 b_1^4} \quad \text{so} \quad b_1(V) = 1.5\left(\frac{G M_{\rm p}^2 a^3}{M_{\rm s} V^2}\right)^{1/4}. \tag{8.49}$$

The rate at which disruptive encounters occur is then given by equation (8.47):

$$\begin{aligned}\mathcal{R} &\equiv \frac{2\sqrt{2\pi}n_{\rm p}}{\sigma_{\rm rel}^3}\int_0^\infty {\rm d}V\, V^3 \exp\left(-\frac{V^2}{2\sigma_{\rm rel}^2}\right)\int_0^{b_1(V)} b\,db \\ &= \sqrt{\frac{14}{3}}\pi\frac{G^{1/2}n_{\rm p}M_{\rm p}a^{3/2}}{M_{\rm s}^{1/2}}.\end{aligned} \tag{8.50}$$

The disruption time of a subject system with semi-major axis a is

$$t_{\rm d} \simeq \mathcal{R}^{-1} \simeq k_{\rm cat}\frac{1}{G\rho_{\rm p}}\left(\frac{GM_{\rm s}}{a^3}\right)^{1/2}, \tag{8.51}$$

where $\rho_{\rm p} \equiv M_{\rm p}n_{\rm p}$ is the mass density of perturbers and $k_{\rm cat}$ is of order unity. Our analytic treatment yields $k_{\rm cat} = 0.15$ but Monte-Carlo simulations of catastrophic disruption suggest that $k_{\rm cat} \simeq 0.07$ (Bahcall, Hut, & Tremaine 1985). It is remarkable that the disruption time in the catastrophic regime is independent of both the velocity dispersion $\sigma_{\rm rel}$ and the mass of individual perturbers, so long as their overall mass density $\rho_{\rm p}$ is fixed.

(b) The diffusive regime In this regime, each encounter imparts a velocity impulse $\Delta\mathbf{v}$ (eq. 8.28) that satisfies $|\Delta\mathbf{v}| \ll |\mathbf{v}|$. The corresponding change in the energy per unit mass is $\Delta E = \mathbf{v}\cdot\Delta\mathbf{v} + \frac{1}{2}|\Delta\mathbf{v}|^2$. Thus the **diffusion term** $\mathbf{v}\cdot\Delta\mathbf{v}$ is much larger than the **heating term** $\frac{1}{2}|\Delta\mathbf{v}|^2$. On the other hand the direction of the velocity impulse, which depends on the relative orientation of the star and the perturber, is usually uncorrelated with the direction of the velocity $\mathbf{v}$ of the star relative to the center of mass of the subject system, which depends on the orbital phase of the star. Thus the average of the diffusion term over many encounters is zero, while the heating term systematically increases the energy.[7] An accurate description of the evolution of the energy under the influence of many high-speed encounters requires the inclusion of both terms, using the Fokker–Planck equation that we described in §7.4.2. Nevertheless, for the sake of simplicity, and since our estimates will be crude anyway, we focus our attention exclusively on the heating term.

Combining equations (8.47) and (8.48), we find that the rate of energy increase for stars with semi-major axis a is

$$\begin{aligned}\dot{E} &= \dot{C}\,\langle\Delta E\rangle \\ &= \frac{14}{3}\sqrt{2\pi}\frac{G^2M_{\rm p}^2n_{\rm p}a^2}{\sigma_{\rm rel}^3}\int_0^\infty {\rm d}V\, V\exp\left(-\frac{V^2}{2\sigma_{\rm rel}^2}\right)\int\frac{{\rm d}b}{b^3}U(b/r_{\rm h}) \\ &= \frac{14}{3}\sqrt{2\pi}\frac{G^2M_{\rm p}^2n_{\rm p}a^2}{\sigma_{\rm rel}r_{\rm h}^2}W, \quad \text{where} \quad W \equiv \int\frac{{\rm d}x}{x^3}U(x).\end{aligned} \tag{8.52}$$

[7] This argument is similar to, but distinct from, the argument leading from equation (8.29) to equation (8.30), which involved an average over the effects of a single collision on many stars rather than an average over the effect of many collisions on a single star.

In general, W must be evaluated numerically for a given mass model. For a Plummer model, $W = 0.5675$ and for a Hernquist model $W = 1.239$ (Figure 8.4).

For point-mass perturbers, $U(x) = 1$, and the heating rate is

$$\dot{E} = \frac{14}{3}\sqrt{2\pi}\frac{G^2 M_{\rm p}^2 n_{\rm p} a^2}{\sigma_{\rm rel}} \int \frac{{\rm d}b}{b^3}. \tag{8.53}$$

This integral over impact parameter diverges at small b. In practice, the distant-tide approximation fails when the impact parameter is comparable to the size of the subject system, so the integration should be cut off at this point.

Comparing the heating rate (8.52) to the energy of an individual star $E = -\frac{1}{2}GM_{\rm s}/a$, we obtain the time required for the star to escape:

$$t_{\rm d} \simeq \frac{|E|}{\dot{E}} \simeq \frac{0.043}{W}\frac{\sigma_{\rm rel} M_{\rm s} r_{\rm h}^2}{G M_{\rm p}^2 n_{\rm p} a^3}. \tag{8.54}$$

For point-mass perturbers, we use equation (8.53), with the integration over impact parameter cut off at $b_{\rm min}$:

$$t_{\rm d} \simeq \frac{|E|}{\dot{E}} \simeq 0.085\frac{\sigma_{\rm rel} M_{\rm s} b_{\rm min}^2}{G M_{\rm p}^2 n_{\rm p} a^3}. \tag{8.55}$$

These are only approximate estimates. A more accurate treatment would employ the Fokker–Planck equation (7.123); in this equation the diffusion coefficient $D[\Delta E]$ is the quantity here called $\dot{E}$, and the diffusion coefficient $D[(\Delta E)^2]$ would be computed similarly as the rate of change of the mean-square energy. Generally this treatment gives a half-life for a star with a given semi-major axis that is a few times shorter than the estimate (8.54).

(c) Disruption of open clusters The masses of open clusters lie in the range $10^2\,\mathcal{M}_\odot \lesssim M_{\rm c} \lesssim 10^4\,\mathcal{M}_\odot$, and their half-mass radii and internal velocity dispersions are $r_{\rm h,c} \approx 2\,{\rm pc}$ and $\sigma_{\rm c} \approx 0.3\,{\rm km\,s^{-1}}$ (Table 1.3). The crossing time at the half-mass radius is $r_{\rm h,c}/\sigma_{\rm c} \approx 10\,{\rm Myr}$. Much of the interstellar gas in our Galaxy is concentrated into a few thousand **giant molecular clouds** of mass $M_{\rm GMC} \gtrsim 10^5\,\mathcal{M}_\odot$ and radius $r_{\rm h,GMC} \approx 10\,{\rm pc}$. Both open clusters and molecular clouds travel on nearly circular orbits through the Galactic disk, with random velocities of order $7\,{\rm km\,s^{-1}}$; thus the dispersion in relative velocity is $\sigma_{\rm rel} \simeq \sqrt{2}\times 7\,{\rm km\,s^{-1}} \simeq 10\,{\rm km\,s^{-1}}$ (eq. 8.45). The duration of a cluster-cloud encounter with impact parameter $b > r_{\rm GMC}$ is then $b/\sigma_{\rm rel} \simeq (b/10\,{\rm pc})\,{\rm Myr}$, which is shorter than the cluster crossing time for $b \lesssim 100\,{\rm pc}$. Thus we may use the impulse approximation to study the effect of close encounters with molecular clouds on open clusters.

The impact parameter at which a typical encounter with a point-mass perturber would disrupt the cluster is given by equation (8.49); identifying

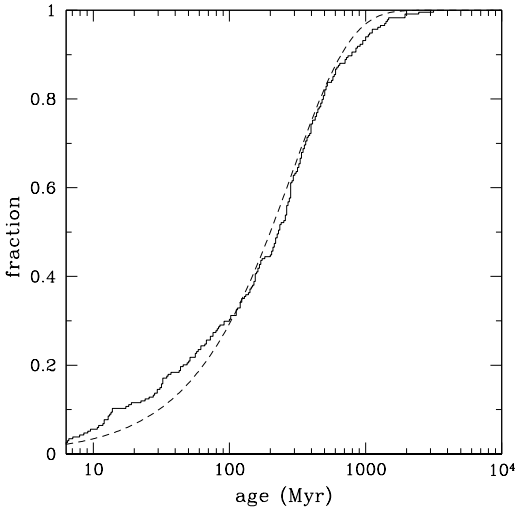

Figure 8.5 The fraction of nearby open clusters younger than a given age. The cluster sample is from Piskunov et al. (2007). The curve is derived from a simple theoretical model in which clusters are born at a constant rate and the probability that a cluster survives for time t is $\exp(-t/\tau)$ with $\tau \simeq 300\,\mathrm{Myr}$. A Kolmogorov–Smirnov test (Press et al. 1986) shows that the two distributions are statistically indistinguishable.

the open cluster with the subject system and the molecular cloud with the perturber we obtain

$$b_1(\sigma_{\rm rel}) = 15\,{\rm pc}\left(\frac{M_{\rm GMC}}{10^5\,\mathcal{M}_\odot}\right)^{1/2}\left(\frac{300\,\mathcal{M}_\odot}{M_{\rm c}}\right)^{1/4} \times \left(\frac{a}{2\,{\rm pc}}\right)^{3/4}\left(\frac{10\,{\rm km\,s^{-1}}}{\sigma_{\rm rel}}\right)^{1/2}. \tag{8.56}$$

Since this distance is larger than the cloud size $r_{\rm h,GMC} \approx 10\,{\rm pc}$, even when the semi-major axis is as small as the typical cluster half-mass radius of 2 pc, the encounters are in the catastrophic regime. Hence the disruption time is given by equation (8.51):

$$t_{\rm d} \simeq 250\,{\rm Myr}\frac{k_{\rm cat}}{0.07}\frac{0.025\,\mathcal{M}_\odot\,{\rm pc^{-3}}}{\rho_{\rm GMC}}\left(\frac{M_{\rm c}}{300\,\mathcal{M}_\odot}\right)^{1/2}\left(\frac{2\,{\rm pc}}{a}\right)^{3/2}, \tag{8.57}$$

where the mean density of gas in molecular clouds is taken to be about half of the total gas density in the solar neighborhood (see Table 1.1).

This result is quite uncertain, not only because the derivation of equation (8.57) is highly idealized, but also because of uncertainties in the molecular cloud parameters and the large dispersion in open-cluster parameters. Nevertheless, the available data suggest that the median lifetime of open clusters is remarkably close to this simple estimate (Figure 8.5). In fact, it was the observation that there are few open clusters with ages $\gtrsim 500\,{\rm Myr}$ that prompted Spitzer (1958) to argue that clusters might be dissolved by the very clouds that bring them into the world.

(d) Disruption of binary stars Binary stars can be thought of as clusters with just two members and, like clusters, they can be disrupted by

encounters with passing perturbers. Obviously the vulnerability of a binary to disruption is an increasing function of the semi-major axis a of its components. Binary semi-major axes are usually measured in terms of the **astronomical unit**, $1\,\mathrm{AU} = 1.496 \times 10^{11}\,\mathrm{m} = 4.848 \times 10^{-6}\,\mathrm{pc}$ (approximately the mean Earth-Sun distance; see Appendix A).

First we consider disruption of binaries in the solar neighborhood by passing stars. We focus on stars—both the binary components and their perturbers—that have ages comparable to the age of the Galaxy and masses comparable to that of the Sun, since these contain most of the stellar mass in the solar neighborhood. The velocity distribution of such stars is triaxial, but we may approximate this distribution by an isotropic Maxwellian with a one-dimensional dispersion $\sigma_\star \simeq 30\,\mathrm{km\,s^{-1}}$ ($1/\sqrt{3}$ of the RMS velocity, from Table 1.2). The velocity distribution is the same for single and binary stars, so the relative dispersion is $\sigma_{\rm rel} = \sqrt{2} \times 30\,\mathrm{km\,s^{-1}} \simeq 40\,\mathrm{km\,s^{-1}}$ (eq. 8.45).

According to equation (8.49), the maximum impact parameter for a catastrophic encounter between a binary star of total mass $M_{\rm b}$ and a passing star of mass $M_{\rm p}$ is

$$\begin{aligned} b_1(\sigma_{\rm rel}) &\simeq 1.5a \left(\frac{GM_{\rm p}^2}{M_{\rm b}\sigma_{\rm rel}^2 a}\right)^{1/4} \\ &\simeq 0.11a \left(\frac{2\,\mathcal{M}_\odot}{M_{\rm b}}\frac{10^4\,\mathrm{AU}}{a}\right)^{1/4} \left(\frac{M_{\rm p}}{1\,\mathcal{M}_\odot}\frac{40\,\mathrm{km\,s^{-1}}}{\sigma_{\rm rel}}\right)^{1/2}. \end{aligned} \tag{8.58}$$

Unless the semi-major axis is so small that the probability of a close encounter is negligible, this result shows that $b_1 \lesssim a$ for solar-type stars in the solar neighborhood, and thus that the encounters are in the diffusive regime. The disruption time is given by equation (8.55), setting $b_{\rm min} \sim a$, where the distant-tide approximation fails; thus (Öpik 1932; Heggie 1975)

$$t_{\rm d} \simeq k_{\rm diff}\frac{\sigma_{\rm rel} M_{\rm b}}{GM_{\rm p}^2 n_{\rm p} a}, \tag{8.59}$$

where $k_{\rm diff} \equiv 0.085(b_{\rm min}/a)^2$. We can refine this estimate by recalling the discussion of the disruption of soft binaries in §7.5.7a; equation (7.173) in that section describes the disruption time in the diffusive regime for the case in which the component stars of the binary have the same mass as the perturbing stars, so $M_{\rm b} = 2M_{\rm p}$, and the velocity dispersion σ of the perturbers and the binaries is the same, so $\sigma_{\rm rel} = \sqrt{2}\sigma$. Equating the two expressions, we find $k_{\rm diff} \simeq 0.022/\ln\Lambda$, where $\Lambda \approx \sigma_{\rm rel}^2 a/(GM_{\rm p})$. For binaries with $a \sim 10^4\,\mathrm{AU}$ in the solar neighborhood, this formula yields $k_{\rm diff} \simeq 0.002$, and Monte Carlo simulations yield a similar value (Bahcall, Hut, & Tremaine 1985). The rather small value of $k_{\rm diff}$ arises in part because $\dot{E}$ grows as a^2, so the heating rate accelerates as the binary gains energy, and in part

because close encounters with either member of the binary contribute to the disruption rate, an effect not accounted for in equation (8.53).

In the solar neighborhood, equation (8.59) yields

$$t_{\rm d} \simeq 15\,{\rm Gyr}\,\frac{k_{\rm diff}}{0.002}\frac{\sigma_{\rm rel}}{40\,{\rm km\,s^{-1}}}\frac{M_{\rm b}}{2\,\mathcal{M}_\odot}\left(\frac{1\,\mathcal{M}_\odot}{M_{\rm p}}\right)^2\frac{0.05\,{\rm pc^{-3}}}{n_{\rm p}}\frac{10^4\,{\rm AU}}{a}. \quad (8.60)$$

Thus the upper limit to the semi-major axes of old binary stars in the solar neighborhood is $a \simeq 2\times10^4\,{\rm AU}$.

Now consider the effects of molecular clouds. Replacing the perturber mass $M_{\rm p}$ in equation (8.58) by the typical cloud mass $M_{\rm GMC}\approx10^5\,\mathcal{M}_\odot$, we find that the maximum impact parameter for impulsive disruption is

$$b_1(\sigma_{\rm rel}) \simeq 1.9\,{\rm pc}\left(\frac{2\,\mathcal{M}_\odot}{M_{\rm b}}\right)^{1/4}\left(\frac{a}{10^4\,{\rm AU}}\right)^{3/4}\left(\frac{M_{\rm GMC}}{10^5\,\mathcal{M}_\odot}\frac{30\,{\rm km\,s^{-1}}}{\sigma_{\rm rel}}\right)^{1/2}. \quad (8.61)$$

We have used a fiducial value $\sigma_{\rm rel}=30\,{\rm km\,s^{-1}}$, which is the sum in quadrature of the dispersions of the stars, $\sigma_\star\simeq30\,{\rm km\,s^{-1}}$, and the clouds, $\sigma_{\rm GMC}\simeq7\,{\rm km\,s^{-1}}$. Since b_1 is smaller than the cloud radius $r_{\rm h,GMC}\approx10\,{\rm pc}$, the encounters are in the diffusive regime. The disruption time can be estimated from equation (8.54), using the value $W=0.5675$ appropriate for a Plummer model of the cloud's density distribution:

$$t_{\rm d} \simeq 0.075\frac{\sigma_{\rm rel}M_{\rm b}r_{\rm h,GMC}^2}{GM_{\rm GMC}^2n_{\rm GMC}a^3}. \quad (8.62)$$

The cloud parameters $M_{\rm GMC}$, $n_{\rm GMC}$, and $r_{\rm h,GMC}$ are all poorly known. Fortunately, they enter this equation in terms of the observationally accessible combinations $\Sigma_{\rm GMC}\equiv(M/\pi r_{\rm h}^2)_{\rm GMC}$, the mean surface density of a cloud, and $\rho_{\rm GMC}=(Mn)_{\rm GMC}$, the mean density of molecular gas. We adopt $\Sigma_{\rm GMC}\simeq300\,\mathcal{M}_\odot\,{\rm pc^{-2}}$ and $\rho_{\rm GMC}\simeq0.025\,\mathcal{M}_\odot\,{\rm pc^{-3}}$ (Hut & Tremaine 1985). Thus

$$t_{\rm d} \simeq 380\,{\rm Gyr}\,\frac{M_{\rm b}}{2\,\mathcal{M}_\odot}\left(\frac{10^4\,{\rm AU}}{a}\right)^3\frac{\sigma_{\rm rel}}{30\,{\rm km\,s^{-1}}}. \quad (8.63)$$

Although this result is subject to substantial uncertainties, together with equation (8.60) it implies that binaries with semi-major axes $\gtrsim2\times10^4\,{\rm AU}\simeq0.1\,{\rm pc}$ cannot survive in the solar neighborhood for its lifetime of $\sim10\,{\rm Gyr}$, due to the combined effects of high-speed encounters with molecular clouds and other stars.

The widest known binary stars in the disk do indeed have separations of about 0.1 pc (Chanamé & Gould 2004); however, there is little evidence for or against the cutoff in the binary distribution that we have predicted at this separation (Wasserman & Weinberg 1987). Binary stars in the stellar halo appear to exist with even larger separations; such binaries can survive

because their velocity $\sigma_{\rm rel}$ relative to the disk is much higher, and because they spend only a fraction of their orbit in the disk, so the disruptive effects from disk stars and molecular clouds are much weaker (Yoo, Chanamé, & Gould 2004).

(e) Dynamical constraints on MACHOs One possible constituent of the dark halo is MACHOs, compact objects such as black holes or nonluminous stars (§1.1.2). Suppose that MACHOs contribute a fraction $f_{\rm h} \gtrsim 0.5$ of the radial force in the solar neighborhood; this is close to the maximum allowed since the disk contributes a fraction $f_{\rm d} = 1 - f_{\rm h} \gtrsim 0.4$ (§6.3.3). Then limits on the optical depth of the dark halo to gravitational lensing (Alcock et al. 2001; Tisserand et al. 2007) imply that the MACHO mass

$$m \lesssim 10^{-7}\,\mathcal{M}_\odot \quad \text{or} \quad m \gtrsim 30\,\mathcal{M}_\odot. \tag{8.64}$$

In §7.4.4 we showed that encounters between MACHOs and disk stars add kinetic energy to the disk stars and thereby increase both the velocity dispersion and the disk thickness; even if this is the only mechanism that heats the disk—and we shall see in §8.4 that it is not—the observed dispersion requires that $m \lesssim 5$–$10 \times 10^6\,\mathcal{M}_\odot$ (eq. 7.104). We now investigate what additional constraints can be placed on the MACHO mass by the effect of high-speed encounters of MACHOs on binary stars.

We write the number density of MACHOs as $n = \rho/m$. If the MACHO mass is small enough, disruption is in the diffusive regime, and we can use equation (8.59) to estimate the disruption time:

$$t_{\rm d,diff} \simeq k_{\rm diff}\frac{\sigma_{\rm rel} M_{\rm b}}{Gm\rho a} \qquad (k_{\rm diff} \approx 0.002), \tag{8.65a}$$

where $M_{\rm b}$ is the mass of the binary. In the catastrophic regime, the disruption time is given by equation (8.51):

$$t_{\rm d,cat} \simeq k_{\rm cat}\frac{M_{\rm b}^{1/2}}{G^{1/2}\rho a^{3/2}} \qquad (k_{\rm cat} \approx 0.07). \tag{8.65b}$$

The transition between these two regimes occurs when the critical impact parameter $b_1(\sigma_{\rm rel})$ (eq. 8.49) is of order the binary semi-major axis a; however, a more accurate way to determine the transition is to set the actual disruption time to

$$t_{\rm d} = \min\left(t_{\rm d,diff}, t_{\rm d,cat}\right) \tag{8.66}$$

and identify the transition with the MACHO mass $m_{\rm crit}$ at which $t_{\rm d,diff} = t_{\rm d,cat}$. Thus we find

$$m_{\rm crit} = \frac{k_{\rm diff}}{k_{\rm cat}}\left(\frac{\sigma_{\rm rel}^2 M_{\rm b} a}{G}\right)^{1/2} \qquad (k_{\rm diff}k_{\rm cat} \approx 0.03). \tag{8.67}$$

Notice that for $m > m_{\rm crit}$, the disruption time $t_{\rm d,cat}$ depends on the overall density contributed by the MACHOs but not their individual masses. Thus the survival of a given type of binary system either rules out *all* MACHO masses above $m_{\rm crit}$ and *some* masses below $m_{\rm crit}$ (if the system's age exceeds $t_{\rm d,cat}$) or does not rule out *any* masses (if its age is less than $t_{\rm d,cat}$).

To plug in numbers for the solar neighborhood, we use the simple model for the DF of MACHOs in the dark halo that we described on page 584. In this model the local density of MACHOs is given by equation (7.94), and the relative dispersion between MACHOs is $\sigma_{\rm rel} = \sqrt{2}\sigma = v_{\rm c}$, where $v_{\rm c}$ is the circular speed (eq. 8.45)—this is also roughly the dispersion between the MACHOs and stars, whether they belong to the disk or the stellar halo. Then

$$m_{\rm crit} \simeq 30\,\mathcal{M}_\odot\,\frac{k_{\rm diff}/k_{\rm cat}}{0.03}\left(\frac{M_{\rm b}}{2\,\mathcal{M}_\odot}\frac{a}{10^4\,{\rm AU}}\right)^{1/2}\frac{v_{\rm c}}{220\,{\rm km\,s^{-1}}}. \tag{8.68}$$

To evaluate the disruption time in the catastrophic regime, $m > m_{\rm crit}$, we use equation (8.65b), and take the local MACHO density from equation (7.94). Assuming the solar radius $R_0 = 8\,{\rm kpc}$ and the solar circular speed $v_{\rm c} = v_0 = 220\,{\rm km\,s^{-1}}$, we have

$$t_{\rm d,cat} \simeq 20\,{\rm Gyr}\,\frac{0.5}{f_{\rm h}}\frac{k_{\rm cat}}{0.07}\left(\frac{10^4\,{\rm AU}}{a}\right)^{3/2}. \tag{8.69}$$

For dark-halo fractions $f_{\rm h} \simeq 0.5$, the disruption time $t_{\rm d,cat}$ is larger than 10 Gyr for semi-major axes $a \lesssim 1.6\times10^4\,{\rm AU}$. In the diffusive regime, the disruption time is even longer. Disk binaries with semi-major axes larger than this limit are likely to be disrupted by encounters with other disk stars (eq. 8.59) and so we cannot probe the MACHO mass with disk binaries. Halo binaries are much less susceptible to other stars and molecular clouds, because they spend only a small fraction of their time in the disk, and therefore might be present with semi-major axes large enough to provide useful constraints on the MACHO population. Thus, if a population of halo binaries with $a \gtrsim 2\times10^4\,{\rm AU}$ were discovered, we could rule out a substantial contribution to the local gravitational field for all MACHO masses exceeding $30\,\mathcal{M}_\odot$ (eq. 8.68). Yoo, Chanamé, & Gould (2004) offer evidence that halo binaries exist with semi-major axes as large as $a \sim 10^5\,{\rm AU}$. Together with the microlensing constraint (8.64) this conclusion, if verified by larger samples, would virtually rule out MACHOs as a significant constituent of the dark halo in the solar neighborhood.

(f) Disk and bulge shocks Globular clusters in disk galaxies pass through the disk plane twice per orbit. As they cross the plane, the gravitational field of the disk exerts a compressive gravitational force which is superposed on the cluster's own gravitational field, pinching the cluster briefly along the normal to the disk plane. Repeated pinching at successive passages

through the disk can eventually disrupt the cluster. This process is known as **disk shocking** (Ostriker, Spitzer, & Chevalier 1972).

Let $Z \equiv Z_{\rm cm} + z$ be the height above the disk midplane of a cluster star, with $Z_{\rm cm}(t)$ the height of the cluster's center of mass. Then so long as the cluster is small compared to the disk thickness, we may use the distant-tide approximation, and equation (8.35) yields

$$\dot{v}_z = -\left(\frac{\partial^2\Phi_{\rm d}}{\partial Z^2}\right)_{\rm cm} z, \tag{8.70}$$

where $v_z = \dot{z}$ is the z-velocity of the star relative to the cluster center.

The gravitational potential arising from a thin disk of density $\rho_{\rm d}(R,z)$ is $\Phi_{\rm d}(R,Z)$, where (eq. 2.74)

$$\frac{{\rm d}^2\Phi_{\rm d}}{{\rm d}Z^2} = 4\pi G\rho_{\rm d}. \tag{8.71}$$

Thus

$$\dot{v}_z = -4\pi G\rho_{\rm d}(R, Z_{\rm cm})z, \tag{8.72}$$

where R is the radius at which the cluster crosses the disk.

If the passage of the cluster through the disk is sufficiently fast for the impulse approximation to hold, z is constant during this passage, and the velocity impulse is

$$\Delta v_z = \int {\rm d}t\, \dot{v}_z = -4\pi G z \int {\rm d}t\, \rho_{\rm d}[R, Z_{\rm cm}(t)]. \tag{8.73}$$

To a good approximation we can assume that the velocity of the center of mass of the cluster is constant as it flies through the disk, so $Z_{\rm cm}(t) = V_z t + \textit{constant}$, where V_z is the Z-velocity of the cluster; eliminating the dummy variable t in favor of $Z_{\rm cm}$ we have

$$\Delta v_z = -\frac{4\pi G z}{|V_z|}\int {\rm d}Z_{\rm cm}\, \rho_{\rm d}(R, Z_{\rm cm}) = -\frac{4\pi G\Sigma_{\rm d}(R) z}{|V_z|}, \tag{8.74}$$

where $\Sigma_{\rm d}(R) \equiv \int {\rm d}Z \rho_{\rm d}(R,Z)$ is the surface density of the disk.

From equation (8.30), the energy per unit mass gained by the cluster in a single disk passage is

$$\Delta E = \tfrac{1}{2}\langle(\Delta v_z)^2\rangle = \frac{8\pi^2 G^2\Sigma_{\rm d}^2}{V_z^2}\langle z^2\rangle. \tag{8.75}$$

If the cluster is spherically symmetric, the average value of z^2 for stars at a given radius r is $\frac{1}{3}r^2$. As shown on page 662, if the cluster has an ergodic DF

the average value of r^2 for stars with a given semi-major axis is $\frac{7}{4}a^2$. Thus the energy gain is

$$\Delta E = \frac{14\pi^2 G^2 \Sigma_{\rm d}^2 a^2}{3V_z^2}. \tag{8.76}$$

The cluster passes through the disk twice in each orbital period T_ψ, so the disruption time is

$$t_{\rm d} \simeq \tfrac{1}{2} T_\psi \frac{|E|}{\Delta E} = 0.005 \frac{M_{\rm gc} V_z^2 T_\psi}{G\Sigma_{\rm d}^2 a^3}, \tag{8.77}$$

where we have set $E = \frac{1}{2}GM_{\rm gc}/a$, since the potential is Keplerian in the outer parts of the cluster, where the effect of disk shocking is strongest.[8]

In the solar neighborhood the Galactic disk has a midplane volume density $\rho \simeq 0.10\,\mathcal{M}_\odot\,{\rm pc}^{-3}$, and surface density $\Sigma_{\rm d} \simeq 50\,\mathcal{M}_\odot\,{\rm pc}^{-2}$ (Table 1.1). The effective thickness of the disk is $h \equiv \Sigma_{\rm d}/\rho \simeq 500\,{\rm pc}$. If we approximate the potential of the Milky Way as spherically symmetric, with circular speed $v_{\rm c}$ at all radii, then the mean-square speed of a collection of test particles such as clusters is $\langle V^2\rangle = v_{\rm c}^2$ (Problem 4.35), so if the cluster distribution is spherical, we expect that $\langle V_z^2\rangle = \frac{1}{3}v_{\rm c}^2$; thus $\langle V_z^2\rangle^{1/2} \simeq 130\,{\rm km\,s^{-1}}$ for $v_{\rm c} \simeq 220\,{\rm km\,s^{-1}}$. Equation (8.77) can be rewritten

$$\begin{aligned} t_{\rm d} \simeq 340\,{\rm Gyr}\,&\frac{M_{\rm gc}}{2\times10^5\,\mathcal{M}_\odot}\,\frac{T_\psi}{200\,{\rm Myr}} \\ &\times \left(\frac{V_z}{130\,{\rm km\,s^{-1}}}\right)^2 \left(\frac{50\,\mathcal{M}_\odot\,{\rm pc}^{-2}}{\Sigma_{\rm d}}\right)^2 \left(\frac{10\,{\rm pc}}{a}\right)^3. \end{aligned} \tag{8.78}$$

This result is based on the impulse approximation, whose validity we must check. The duration of the encounter of the cluster with the disk is $\tau \approx h/V_z \simeq 4\,{\rm Myr}$ for $V_z \simeq 130\,{\rm km\,s^{-1}}$ and $h \simeq 500\,{\rm pc}$. The crossing time in the outer parts of the cluster is roughly the inverse of the orbital frequency, $(a^3/GM_{\rm gc})^{1/2} \simeq 1\,{\rm Myr}\,(a/10\,{\rm pc})^{3/2}$ for a cluster mass of $2\times10^5\,\mathcal{M}_\odot$. Thus the impulse approximation is valid only in the outer parts of the cluster, $a \gtrsim 30\,{\rm pc}$, and at these semi-major axes the disruption time is $\lesssim 10\,{\rm Gyr}$. We conclude that disk shocks can lead to substantial erosion of the outermost stars in a typical globular cluster orbiting at the solar radius. For clusters orbiting at smaller radii, disk shocks are even more important, since the orbital time is shorter and the disk surface density is larger.

Bulge shocking is a closely related process. Here the rapidly changing external gravitational field arises as a cluster on a highly eccentric orbit

[8] A subtle but important assumption in deriving this result is that the orbital phase and thus the value of the height z is uncorrelated between successive disk passages. This assumption is plausible because the interval between successive disk passages is likely to vary considerably for an eccentric, inclined cluster orbit in a realistic disk galaxy potential.

plunges through the bulge of a disk galaxy or the dense center of an elliptical galaxy. In this case the use of the term "shock" is less apt, since the duration of the encounter is not very short compared to the crossing time, even in the outer parts of the cluster, and the impulse approximation is not strictly valid. Nevertheless, the results are similar: the encounters systematically pump energy into the stars in the outer cluster, leading to the escape of stars and the eventual dissolution of the cluster.

The evolution of the globular-cluster population under the influence of disk and bulge shocks is described in §7.5.6.

(g) High-speed interactions in clusters of galaxies The study of galaxies in clusters provides unique insights into galaxy formation and evolution, not only because many dynamical processes are stronger and more obvious in the high-density cluster environment, but also because clusters can be detected at high redshift, enabling the evolution of the galaxy population to be studied directly.

The relative velocities between galaxies in a rich cluster, $\sim 2000\,\mathrm{km\,s^{-1}}$, are so large that collisions of galaxies last only a few Myr, far less than the crossing time in the galaxy, so they can be treated by the impulse approximation. We model the cluster as a singular isothermal sphere, with density $\rho(r) = \sigma^2/(2\pi G r^2)$ (eq. 4.103). In clusters the galaxies and dark matter have a similar distribution, so it is reasonable to assume that the ratio $M_\star \equiv \rho/n$ of the mass density to the galaxy number density—in other words the mass per galaxy—is constant. (Note that this is not necessarily the mass *of* the galaxy, since most of the mass is probably spread uniformly through the cluster and is not associated with any individual galaxies.) We focus on galaxies with the characteristic luminosity $L_\star = 2.9 \times 10^{10}\,L_\odot$ in the R band (eq. 1.18). The mass-to-light ratio in rich clusters is $\Upsilon_R \approx 200\Upsilon_\odot$ (eq. 1.25) so the mass associated with each $L_\star$ galaxy is

$$M_\star = \Upsilon_R L_\star \approx 6 \times 10^{12}\,\mathcal{M}_\odot. \tag{8.79}$$

Consequently, the number density of $L_\star$ galaxies is

$$n(r) = \frac{\rho(r)}{M_\star} = \frac{\sigma^2}{2\pi G M_\star r^2}. \tag{8.80}$$

Now let us estimate the rate at which a given galaxy encounters other galaxies. The velocity dispersion in clusters is so high that gravitational focusing is negligible, so if galaxies are deemed to collide when their centers come within a collision radius $r_{\rm coll}$, then from equation (7.194) the collision time is given by

$$\frac{1}{t_{\rm coll}} = 4\sqrt{\pi} n \sigma r_{\rm coll}^2 \simeq \frac{2}{\sqrt{\pi}} \frac{\sigma^3 r_{\rm coll}^2}{G M_\star r^2}. \tag{8.81}$$

After replacing $r_{\rm coll}$ with $2r_{\rm h}$, where $r_{\rm h}$ is the galaxy radius, equation (8.81) yields

$$t_{\rm coll} \simeq 0.2\frac{GM_\star r^2}{\sigma^3 r_{\rm h}^2} \simeq 6\,{\rm Gyr}\left(\frac{800\,{\rm km\,s^{-1}}}{\sigma}\right)^3\left(\frac{20\,{\rm kpc}}{r_{\rm h}}\frac{r}{0.5\,{\rm Mpc}}\right)^2\frac{M_\star}{5\times10^{12}\,\mathcal{M}_\odot}. \tag{8.82}$$

Thus the stellar component of a galaxy in the central 0.5 Mpc of a rich cluster is likely to have suffered at least one close encounter with another galaxy. What are the consequences of such encounters?

As we have seen, high-speed collisions between galaxies have only a small effect on the distribution of stars, but if both galaxies contain gas disks the gas will suffer a violent collision and be lost from the galaxies. Spitzer & Baade (1951) suggested that collisions might transform spiral galaxies into gas-free lenticular galaxies, thereby explaining the observation that spirals are replaced by lenticulars in high-density environments such as clusters (§1.1.3). An alternative and more likely explanation (Gunn & Gott 1972) is that ram pressure, heating, and other interactions with hot intergalactic gas in the cluster have gradually eroded the gas disks of spiral galaxies (see van Gorkom 2004 for a review).

Isolated galaxies have dark halos that extend to several hundred kpc. In clusters, encounters strip off the outer parts of these halos, so we expect that cluster galaxies will have much smaller and less massive halos than galaxies in low-density environments. We can make a crude estimate of this effect using equation (8.54). For this purpose, we assume that the subject system and the perturber are identical. Thus $M_{\rm s} = M_{\rm p} = M$ and $\sigma_{\rm rel} = \sqrt{2}\sigma$ where σ is the velocity dispersion in the cluster. We set the dimensionless parameter $W \approx 1$, and set $r_{\rm h} = a$ to estimate the disruption time for stars at the half-mass radius. Thus

$$t_{\rm d} \simeq 0.06\frac{\sigma}{GMnr_{\rm h}}. \tag{8.83}$$

We eliminate the number density n using equation (8.80), and write the mass M of the galaxy in terms of $r_{\rm h}$ and its velocity dispersion $\sigma_{\rm s}^2 = \frac{1}{3}\langle v^2\rangle$, using the virial theorem in the form (4.249b); the dispersion for an $L_\star$ galaxy is $\sigma_{\rm s} \simeq 200\,{\rm km\,s^{-1}}$ (eq. 1.21). Thus

$$\begin{aligned} t_{\rm d} &\simeq 0.056\frac{GM_\star r^2}{\sigma\sigma_{\rm s}^2 r_{\rm h}^2} \\ &\simeq 3.8\,{\rm Gyr}\frac{M_\star}{5\times10^{12}\,\mathcal{M}_\odot}\frac{800\,{\rm km\,s^{-1}}}{\sigma}\left(\frac{200\,{\rm km\,s^{-1}}}{\sigma_{\rm s}}\right)^2\left(\frac{r}{0.5\,{\rm Mpc}}\frac{50\,{\rm kpc}}{r_{\rm h}}\right)^2. \end{aligned} \tag{8.84}$$

Note that the disruption time is related to the collision time (eq. 8.82) by the simple formula $t_{\rm d}/t_{\rm coll} \simeq 0.25(\sigma/\sigma_{\rm s})^2$; as the velocity dispersion σ of the

cluster increases relative to the dispersion σ_s of the galaxies, gravitational interactions become less and less important relative to physical collisions.

This result shows that encounters with other galaxies will erode the dark halos of galaxies residing in the inner 0.5 Mpc of a rich cluster to a radius $r_h \lesssim 50$ kpc. Most of the dark-halo mass has therefore been stripped from the individual galaxies, and is now smoothly distributed throughout the cluster (Richstone 1976). Since the disruption time is proportional to M_s/a^3 (eq. 8.54) and thus to the mean density of the subject system, low-density galaxies are more severely affected by encounters, and can be completely disrupted near the cluster center. The large-scale static tidal field of the cluster also strips the outer halo of cluster galaxies, a process that we shall investigate in the next section.

Clusters form hierarchically from smaller systems that resemble groups of galaxies (§§1.1.5 and 9.2.2). Groups have velocity dispersions of only $\sim 300\,\mathrm{km\,s^{-1}}$, so encounters in groups occur at lower speeds and have stronger effects—they frequently lead to mergers—and dynamical friction is more powerful. Most of the galaxy evolution that we see in the centers of rich clusters may thus be due to "pre-processing" in the groups that later merged to form the cluster. For example, the exceptionally luminous brightest cluster galaxies or cD galaxies (§1.1.3) that are found at the centers of clusters probably arise from the merger of galaxies in precursor groups (Dubinski 1998).

8.3 Tides

In the last section we examined how tidal shocks from high-speed encounters heat stellar systems and erode their outer parts. We now consider the opposite limiting case of a static tidal field. The simplest example of a static tide occurs when a satellite travels on a circular orbit in the gravitational field of a much larger spherical host system. In this case, the satellite experiences no shocks—in fact, in the frame rotating with the satellite the external tidal field is stationary—so in the absence of other relaxation effects, a sufficiently small system could survive indefinitely. However, a static tidal field prunes distant stars from the satellite system, thereby enforcing an upper limit on its size. Observationally, globular clusters and other satellite systems often show a fairly sharp outer boundary, which is called the tidal radius on the assumption that it is caused by this process (see §4.3.3c, BM §6.1.10, and King 1962).

In the following subsections, we analyze the effect of a tidal field on a satellite in a circular orbit using two complementary approaches.

8.3.1 The restricted three-body problem

Let us suppose that the host and satellite systems are point masses M and m, traveling at separation R_0 in a circular orbit around their mutual center of mass. The **restricted three-body problem** is to find the trajectory of a massless test particle that orbits in the combined gravitational field of these two masses (Szebehely 1967; Hénon 1997; Valtonen & Karttunen 2006). Solutions of this problem provide a good approximation to the motion of stars in the outer parts of a satellite stellar system that is on a circular orbit near or beyond the outer edge of a spherical host system.

The two masses orbit their common center of mass with angular speed

$$\Omega = \sqrt{\frac{G(M+m)}{R_0^3}}, \tag{8.85}$$

so the gravitational field is stationary when referred to a coordinate system centered on the center of mass that rotates at speed Ω. We orient this coordinate system so that the centers of the satellite and host systems are at $\mathbf{x}_m = [MR_0/(M+m), 0, 0]$ and $\mathbf{x}_M = [-mR_0/(M+m), 0, 0]$, and the angular speed is $\mathbf{\Omega} = (0, 0, \Omega)$. In §3.3.2 we showed that on any orbit in such a system, the Jacobi integral

$$\begin{aligned} E_{\rm J} &= \tfrac{1}{2}v^2 + \Phi(\mathbf{x}) - \tfrac{1}{2}|\mathbf{\Omega}\times\mathbf{x}|^2 \\ &= \tfrac{1}{2}v^2 + \Phi_{\rm eff}(\mathbf{x}) \end{aligned} \tag{8.86}$$

is conserved (eq. 3.113). Since $v^2 \geq 0$, a star with Jacobi integral $E_{\rm J}$ can never trespass into a region where $\Phi_{\rm eff}(\mathbf{x}) > E_{\rm J}$. Consequently, the surface $\Phi_{\rm eff}(\mathbf{x}) = E_{\rm J}$, the zero-velocity surface for stars of Jacobi integral $E_{\rm J}$, forms an impenetrable wall for such stars. Figure 8.6 shows contours of constant $\Phi_{\rm eff}$ in the equatorial plane of two orbiting point masses; the Lagrange points are the extrema (maxima and saddle points) of this surface. It is instructive to compare these contours to those in a bar-like potential, shown in Figure 3.14.[9] The stability of orbits near the Lagrange points in the restricted three-body problem is discussed in Problem 3.25.

From the figure we see that the zero-velocity surfaces near each body are centered on it, but farther out the zero-velocity surfaces surround both bodies. Hence, at the critical value of $\Phi_{\rm eff}$ corresponding to the last zero-velocity surface to enclose only one body, there is a discontinuous change in the region confined by the Jacobi integral. The last closed zero-velocity surface surrounding a single body is called its **tidal** or **Roche surface**; since this surface touches the Lagrange point L_3 that lies between the two masses

[9] Note that different authors use different conventions for the numbering of the Lagrange points L_1, L_2, L_3.

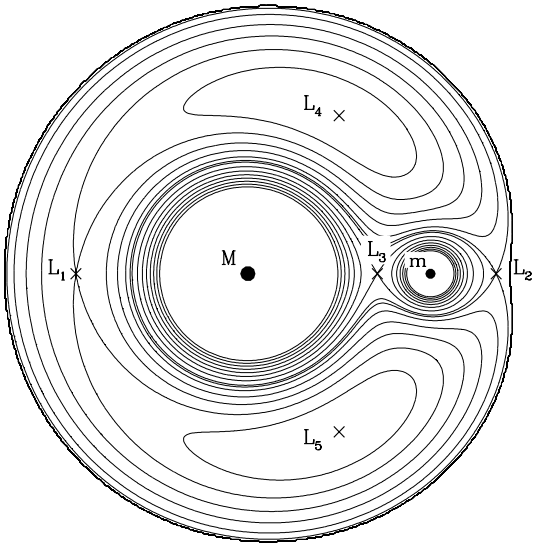

Figure 8.6 Contours of equal effective potential $\Phi_{\rm eff}$ defined by equation (8.88) for two point masses in a circular orbit. The mass ratio $m/M = \frac{1}{9}$. The points $L_1, \ldots, L_5$ are the Lagrange points. The L_4 and L_5 points form an equilateral triangle with the two masses (Problem 3.25).

on the line connecting them, it is natural to identify the tidal radius of m as the distance $r_{\rm J}$ between m and L_3.

We may evaluate $r_{\rm J}$ by noticing that at $(x_m - r_{\rm J}, 0, 0)$ the effective potential has a saddle point, so

$$\left(\frac{\partial \Phi_{\rm eff}}{\partial x}\right)_{(x_m - r_{\rm J}, 0, 0)} = 0. \tag{8.87}$$

For two point masses a distance R_0 apart, equations (8.85) and (8.86) imply

$$\Phi_{\rm eff}(\mathbf{x}) = -G\left[\frac{M}{|\mathbf{x} - \mathbf{x}_M|} + \frac{m}{|\mathbf{x} - \mathbf{x}_m|} + \frac{M+m}{2R_0^3}(x^2 + y^2)\right]. \tag{8.88}$$

At a point between the two masses, (8.87) is satisfied if

$$0 = \frac{1}{G}\left(\frac{\partial \Phi_{\rm eff}}{\partial x}\right)_{(x_m - r_{\rm J}, 0, 0)} = \frac{M}{(R_0 - r_{\rm J})^2} - \frac{m}{r_{\rm J}^2} - \frac{M+m}{R_0^3}\left(\frac{MR_0}{M+m} - r_{\rm J}\right). \tag{8.89}$$

This equation leads to a fifth-order polynomial whose roots give $r_{\rm J}$. In general these roots must be found numerically. However, if the satellite is small, $m \ll M$, then $r_{\rm J} \ll R_0$, and we can expand $(R_0 - r_{\rm J})^{-2}$ in powers of $r_{\rm J}/R_0$ to find

$$0 = \frac{M}{R_0^2}\left(1 + \frac{2r_{\rm J}}{R_0} + \cdots\right) - \frac{m}{r_{\rm J}^2} - \frac{M}{R_0^2} + \frac{M+m}{R_0^3} r_{\rm J} \simeq \frac{3Mr_{\rm J}}{R_0^3} - \frac{m}{r_{\rm J}^2}. \quad (8.90)$$

Truncating the series in this way is none other than the distant-tide approximation. Then to first order in $r_{\rm J}/R_0$,

$$r_{\rm J} = \left(\frac{m}{3M}\right)^{1/3} R_0. \quad (8.91)$$

We call the radius $r_{\rm J}$ the **Jacobi radius** of the mass m; alternative names are the **Roche** or **Hill radius**. The Jacobi radius of an orbiting stellar system is expected to correspond to the observational tidal radius, the maximum extent of the satellite system. However this correspondence is only approximate, for several reasons:

(i) The tidal surface is not spherical (see Problem 8.12), so it cannot be fully characterized by a single radius.

(ii) All we have established is that a test particle can never cross the tidal surface if it lies inside the tidal surface and has a velocity (in the rotating frame) small enough that $E_{\rm J} < \Phi_{\rm eff}(L_3)$. Stars with larger velocities may or may not escape from the satellite; conversely, stars that lie outside the tidal surface can, in some cases, remain close to the satellite for all future times (see Problem 8.14 and Hénon 1970). The approximate correspondence between the Jacobi radius and the observational tidal radius arises because the fraction of velocity space occupied by orbits that remain close to the satellite diminishes rapidly beyond $r_{\rm J}$.

(iii) In most applications, the satellite system is not on a circular orbit. When m is on an eccentric orbit, there is no reference frame in which the potential experienced by a test particle is stationary, and no analog of the Jacobi integral exists.[10] Thus no direct generalization of our derivation of the Jacobi radius to the case of non-circular satellite orbits is possible. King (1962) and others have argued that if the satellite is on a non-circular orbit, the tidal radius is still given by equation (8.91), but with R_0 replaced by the pericenter distance (we used an analogous argument to describe the tidal disruption of stars orbiting a massive black hole; see eq. 7.200). A more accurate approach is to recognize that the effect of tidal fields on satellites in non-circular orbits is intermediate between tidal radii—a concept that applies to circular orbits—and tidal shocks—which apply to high-velocity or plunging orbits. The tidal radius limits the satellite at a fixed size, no matter how many orbits it

[10] Although analogs of the Lagrange points can exist (Szebehely 1967).

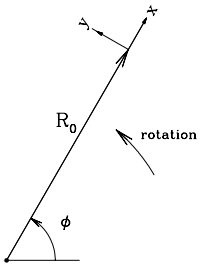

Figure 8.7 The rotating (x, y) coordinate system used in Hill's approximation.

travels, while tidal shocks prune the satellite more and more at every pericenter passage.

(iv) Stars are usually lost from the satellite as a result of weak perturbations, such as two-body relaxation, that drive $E_{\rm J}$ slightly higher than $\Phi_{\rm eff}(L_3)$. Such stars drift slowly away from the satellite and thus can remain close to the satellite for many orbital periods, thereby contributing to the star counts even though they are no longer bound to the satellite (Fukushige & Heggie 2000).

(v) In many cases, the satellite orbits within the body of the host system, so the point-mass approximation used in deriving equation (8.91) is not accurate. This defect, at least, is easy to remedy—see equation (8.106) below.

Tidal forces can be thought of as imposing a limit on the density of a satellite. Let $\overline{\rho} \equiv m/(\frac{4}{3}\pi r_{\rm J}^3)$ be the mean density of the satellite within a distance $r_{\rm J}$, and $\overline{\rho}_{\rm h} \equiv M/(\frac{4}{3}\pi R_0^3)$ be the mean density of the host inside the orbital radius R_0. Then equation (8.91) states that

$$\overline{\rho} = 3\overline{\rho}_{\rm h}; \tag{8.92}$$

to within a factor of order unity, a satellite is pruned by tidal forces until its mean density equals the mean density of its host interior to its orbital radius.

8.3.2 The sheared-sheet or Hill's approximation

When the satellite is much smaller than the distance to the center of the host system, we can use the distant-tide approximation for the host's gravitational field (§8.2.1). We consider a spherically symmetric host system with potential $\Phi(R)$ at a distance R from its center; here we do *not* assume that the host is a point mass, so $\Phi(R)$ is not necessarily the Keplerian potential $-GM/R$. We assume that the satellite travels on a circular orbit at distance R_0 from the center of the host. We work in a frame with origin at the center of mass

of the satellite, in which the x-y plane coincides with the orbital plane of the satellite, $\hat{\mathbf{e}}_x$ points directly away from the center of the host system, and $\hat{\mathbf{e}}_y$ points in the direction of the orbital motion of the satellite (see Figure 8.7). This frame rotates with the circular frequency $\mathbf{\Omega}_0 \equiv \Omega_0 \hat{\mathbf{e}}_z$, so the acceleration of a particle in the satellite is given by equation (3.116),

$$\frac{\mathrm{d}^2\mathbf{x}}{\mathrm{d}t^2} = -\boldsymbol{\nabla}\Phi - 2\mathbf{\Omega}_0 \times \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} - \mathbf{\Omega}_0 \times (\mathbf{\Omega}_0 \times \mathbf{x}), \tag{8.93}$$

where

$$\boldsymbol{\nabla}\Phi = \boldsymbol{\nabla}\Phi_{\mathrm{s}} + \sum_{k=1}^{3} \Phi_{jk} x_k. \tag{8.94}$$

Here $\Phi_{\mathrm{s}}(\mathbf{x})$ is the gravitational potential from the satellite, and the second term arises from the distant-tide approximation (8.35). In our coordinate system, the center of the host is located at $\mathbf{X} = (-R_0, 0, 0)$ and from equation (8.36):

$$\Phi_{xx} = \Phi''(R_0) \quad ; \quad \Phi_{yy} = \Phi_{zz} = \frac{\Phi'(R_0)}{R_0} \quad ; \quad \Phi_{xy} = \Phi_{xz} = \Phi_{yz} = 0. \tag{8.95}$$

The equations of motion (8.93) read

$$\begin{aligned} \ddot{x} &= 2\Omega_0 \dot{y} + \left[\Omega_0^2 - \Phi''(R_0)\right] x - \frac{\partial \Phi_{\mathrm{s}}}{\partial x}; \\ \ddot{y} &= -2\Omega_0 \dot{x} + \left[\Omega_0^2 - \frac{\Phi'(R_0)}{R_0}\right] y - \frac{\partial \Phi_{\mathrm{s}}}{\partial y}; \\ \ddot{z} &= -\frac{\Phi'(R_0)}{R_0} z - \frac{\partial \Phi_{\mathrm{s}}}{\partial z}. \end{aligned} \tag{8.96}$$

Using the relation $\Phi'(R_0) = R_0\Omega_0^2$ we see that the term in square brackets in the second line vanishes. Moreover we can rewrite $\Omega_0^2 - \Phi''(R_0)$ as $-2R_0\Omega_0\Omega'(R_0)$ and this in turn can be rewritten as $4\Omega_0 A_0$ where $A_0 = A(R_0)$ is given by equation (3.83). Thus

$$\ddot{x} - 2\Omega_0\dot{y} - 4\Omega_0 A_0 x = -\frac{\partial \Phi_{\mathrm{s}}}{\partial x} \; ; \; \ddot{y} + 2\Omega_0 \dot{x} = -\frac{\partial \Phi_{\mathrm{s}}}{\partial y} \; ; \; \ddot{z} + \Omega_0^2 z = -\frac{\partial \Phi_{\mathrm{s}}}{\partial z}. \tag{8.97}$$

These are the equations of motion in the **sheared sheet** or **Hill's approximation**, named after the mathematician G. W. Hill, who used this approach to study the motion of the Moon in the nineteenth century (Murray & Dermott 1999).

(a) The epicycle approximation and Hill's approximation We first consider the trajectories of test particles in the absence of a satellite, $\Phi_{\mathrm{s}}(\mathbf{x}) = 0$. The simplest solutions of equations (8.97) have the form

$$x(t) = x_{\mathrm{g}} = \text{constant} \quad ; \quad y(t) = -2A_0 x_{\mathrm{g}} t + \text{constant} \quad ; \quad z(t) = 0. \tag{8.98}$$

These are the analogs of circular orbits in the host system. The general solution is

$$\begin{aligned} x(t) &= x_{\rm g} + X\cos(\kappa_0 t + \alpha), \\ y(t) &= y_{\rm g}(t) - Y\sin(\kappa_0 t + \alpha), \quad \text{where} \quad y_{\rm g}(t) = y_{\rm g0} - 2A_0 x_{\rm g} t, \\ z(t) &= Z\cos(\Omega_0 t + \alpha_z), \end{aligned} \tag{8.99}$$

where $x_{\rm g}$, $y_{\rm g0}$, X, Z, α and α_z are arbitrary constants and

$$\kappa_0^2 = 4\Omega_0(\Omega_0 - A_0) = -4\Omega_0 B_0 \quad ; \quad \frac{X}{Y} = \frac{\kappa_0}{2\Omega_0}. \tag{8.100}$$

Here $B_0 = A_0 - \Omega_0$ (eq. 3.84). Thus we have re-derived the epicycle approximation of §3.2.3, in particular the relation between the epicycle frequency κ_0 and Oort's constants (eq. 3.84) and the ratio of the axes of the epicycle (eq. 3.95). The difference between the two derivations is that §3.2.3 described an *approximate* solution of the *exact* equations of motion for a particle on a nearly circular orbit, while here we have found an *exact* solution of Hill's *approximate* equations of motion. Note that in Hill's approximation all particles have the same epicycle frequency.

It is straightforward to verify that when $\Phi_{\rm s} = 0$, the following expressions are integrals of the motion:

$$E_\parallel \equiv \tfrac{1}{2}(\dot{x}^2 + \dot{y}^2 - 4\Omega_0 A_0 x^2) \quad ; \quad E_\perp \equiv \tfrac{1}{2}(\dot{z}^2 + \Omega_0^2 z^2) \quad ; \quad L \equiv \dot{y} + 2\Omega_0 x; \tag{8.101}$$

$E_\parallel + E_\perp$ and $R_0 L$ differ from the Jacobi integral and the angular momentum by constant terms and terms of order $\mathrm{O}(x^3, y^3)$. These expressions are related to the constants in the orbit solutions (8.99) by

$$E_\parallel = 2A_0 B_0 x_{\rm g}^2 + \tfrac{1}{2}\kappa_0^2 X^2 \quad ; \quad E_\perp = \tfrac{1}{2}\Omega_0^2 Z^2 \quad ; \quad L = -2B_0 x_{\rm g}. \tag{8.102}$$

A circular orbit has $E_\parallel = \frac{1}{2}A_0 L^2/B_0$; hence it is natural to define the **epicycle energy** E_x as the difference

$$\begin{aligned} E_x &\equiv E_\parallel - \frac{A_0 L^2}{2B_0}, \\ &= \tfrac{1}{2}[\dot{x}^2 + \kappa_0^2 (x - x_{\rm g})^2], \\ &= \tfrac{1}{2}\kappa_0^2 X^2, \\ &= \tfrac{1}{2}\dot{x}^2 + \frac{2\Omega_0^2}{\kappa_0^2}(\dot{y} + 2A_0 x)^2. \end{aligned} \tag{8.103}$$

Some of these results, derived in other ways and with slightly different notation, have already appeared as equations (3.86) and (3.102).

(b) The tidal radius in Hill's approximation If a satellite is present, with potential $\Phi_{\rm s}(\mathbf{x})$, the integrals in equation (8.101) are no longer conserved; the only remaining classical integral is (Problem 8.15)

$$E \equiv \tfrac{1}{2}(\dot{x}^2 + \dot{y}^2 + \dot{z}^2 - 4\Omega_0 A_0 x^2 + \Omega_0^2 z^2) + \Phi_{\rm s}(\mathbf{x}). \tag{8.104}$$

This integral is the analog of the Jacobi integral (8.86).

Now let us imagine that the satellite potential $\Phi_{\rm s}$ arises from a mass m that is located at $\mathbf{x} = \mathbf{0}$. The equations of motion (8.97) become

$$\ddot{x} = 2\Omega_0\dot{y} + 4\Omega_0 A_0 x - \frac{Gmx}{r^3} \ ; \ \ddot{y} = -2\Omega_0\dot{x} - \frac{Gmy}{r^3} \ ; \ \ddot{z} = -\Omega_0^2 z - \frac{Gmz}{r^3}, \tag{8.105}$$

where $r^2 = x^2 + y^2 + z^2$. The test particle remains stationary ($\ddot{x} = \dot{x} = \ddot{y} = \dot{y} = \ddot{z} = 0$) if and only if $y = z = 0$ and $4\Omega_0 A_0 = Gm/|x|^3$. These conditions are satisfied for the points on the x axis with

$$x = \pm r_{\rm J}, \quad \text{where} \quad r_{\rm J} \equiv \left(\frac{Gm}{4\Omega_0 A_0}\right)^{1/3}. \tag{8.106}$$

These stationary points are analogs to the Lagrange points L_2 and L_3 in the restricted three-body problem (Figure 8.6). If the host is a point mass $M \gg m$, then $\Omega(R) = (GM/R^3)^{1/2}$ so $A_0 = \frac{3}{4}\Omega_0$ and

$$r_{\rm J} = \left(\frac{m}{3M}\right)^{1/3} R_0. \tag{8.107}$$

Thus we recover expression (8.91) for the Jacobi radius. For a spherical host with mass $M(R)$ interior to radius R, it is straightforward to show that this expression is modified by replacing M by $M(R_0)$ and multiplying the Jacobi radius by a factor

$$f = \left(1 - \tfrac{1}{3}\frac{{\rm d}\ln M}{{\rm d}\ln R}\right)^{-1/3}. \tag{8.108}$$

The factor f is unity for a point mass and 1.145 for a singular isothermal sphere ($M \propto R$). For a homogeneous sphere ($M \propto R^3$) f diverges, so there is no Jacobi radius: in this case the host potential $\Phi(R) = \frac{1}{2}\Omega_0^2 R^2$ and Oort's constant $A_0 = 0$, so the tidal field $4\Omega_0 A_0 x$ in the equations of motion (8.97) is absent. Physically, there is no tidal radius because all stars in this potential have the same orbital period: thus, even if the satellite mass were zero, stars in nearly circular orbits with similar radii and azimuths will continue to have similar radii and azimuths at all future times.

8.3.3 Tidal tails and streamers

We now investigate what happens to stars after they are stripped from a satellite by tidal forces, with the help of the angle-action variables described in §3.5 (Helmi & White 1999; Tremaine 1999). Consider a satellite of mass m orbiting a host that has mass $M \gg m$ interior to the satellite orbit. At its pericenter, a distance R from the center of the host, the satellite is pruned by tidal forces to a radius $r \approx R(m/M)^{1/3}$. Its velocity at pericenter is $V \approx$

$(GM/R)^{1/2}$. To a first approximation, we may assume that stars lost from the satellite no longer feel its gravitational field, and follow orbits determined solely by the field of the host. On such orbits, the actions $\mathbf{J}$ are constant and the angles $\boldsymbol{\theta}$ increase linearly with time, at a rate $\dot{\boldsymbol{\theta}} = \boldsymbol{\Omega} = \partial H/\partial \mathbf{J}$ where H is the Hamiltonian corresponding to the host potential. The stripped stars have a range of actions and angles, which we write as $\mathbf{J}_0 \pm \Delta\mathbf{J}$, $\boldsymbol{\theta}_0 \pm \Delta\boldsymbol{\theta}$. The mean actions $\mathbf{J}_0$ and the mean angles $\boldsymbol{\theta}_0$ at the time the satellite passes through pericenter are very nearly the actions and angles of the satellite at that time, since the tidal forces are symmetric about its center of mass. The spread in actions and angles in the stripped stars arises from two effects: (i) the stars are lost from both the inner and outer edge of the satellite (near the Lagrange points L_3 and L_2), and (ii) the stars have a range of velocities, roughly equal to the velocity dispersion σ of the satellite. These effects lead to a fractional spread $r/R \sim (m/M)^{1/3}$ in position and σ/V in velocity. Since $\sigma \approx (Gm/r)^{1/2} \sim V(m/M)^{1/2}(R/r)^{1/2} \sim V(m/M)^{1/3}$ the two effects yield approximately the same fractional spread. Thus, the stripped stars are initially distributed through ranges in action and angle given by

$$\frac{\Delta J_i}{J_i}, \Delta\theta_i \sim \left(\frac{m}{M}\right)^{1/3}. \tag{8.109}$$

The spread in actions leads to a spread in orbital frequencies

$$\Delta\Omega_i \sim \sum_{j=1}^{3} H_{ij}\Delta J_j, \quad \text{where} \quad D_{ij} \equiv \frac{\partial^2 H}{\partial J_i \partial J_j} \tag{8.110}$$

is the Hessian of the Hamiltonian. The spread in angles grows linearly with time, such that

$$\Delta\boldsymbol{\theta}(t) = \Delta\boldsymbol{\theta}(0) + \Delta\boldsymbol{\Omega}t, \tag{8.111}$$

where $t = 0$ is the time at which the stars were stripped. At large times the second term dominates, so we have

$$\Delta\boldsymbol{\theta}(t) \simeq t\,\mathbf{D}\cdot\Delta\mathbf{J}. \tag{8.112}$$

Since the matrix $\mathbf{D}$ is symmetric, it is diagonalizable, that is, there exists an orthogonal matrix $\mathbf{A}$ such that

$$\mathbf{A}\mathbf{D}\mathbf{A}^{\mathrm{T}} = \widetilde{\mathbf{D}}, \tag{8.113}$$

where $\mathbf{A}^{\mathrm{T}} = \mathbf{A}^{-1}$ is the transpose of $\mathbf{A}$ ($A^{\mathrm{T}}_{jk} = A_{kj}$), and $\widetilde{\mathbf{D}}$ is the diagonal matrix formed by the eigenvalues λ_i of $\mathbf{D}$. We now make a canonical transformation to new angle-action variables $(\boldsymbol{\theta}', \mathbf{J}')$ using the generating function $S(\boldsymbol{\theta}, \mathbf{J}') = \mathbf{J}'\cdot\mathbf{A}\cdot\boldsymbol{\theta}$ (eq. D.93); thus

$$\boldsymbol{\theta}' = \frac{\partial S}{\partial \mathbf{J}'} = \mathbf{A}\cdot\boldsymbol{\theta} \quad ; \quad \mathbf{J} = \frac{\partial S}{\partial \boldsymbol{\theta}} = \mathbf{A}^{\mathrm{T}}\cdot\mathbf{J}'. \tag{8.114}$$

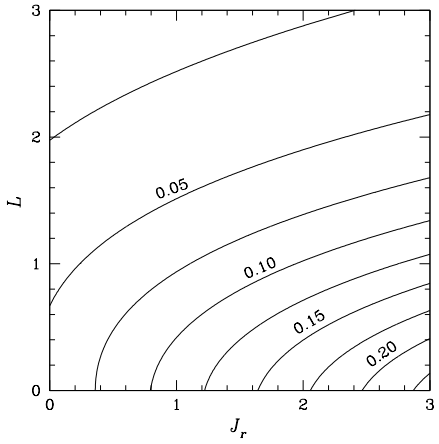

Figure 8.8 The ratio of the two largest eigenvalues of $\mathbf{D}$, the Hessian of the Hamiltonian, for the isochrone potential (see §3.1c and Problem 3.41). The axes are the radial action J_r and the angular momentum L. When this ratio is small compared to unity, tidally stripped stars form a one-dimensional filament or tidal streamer.

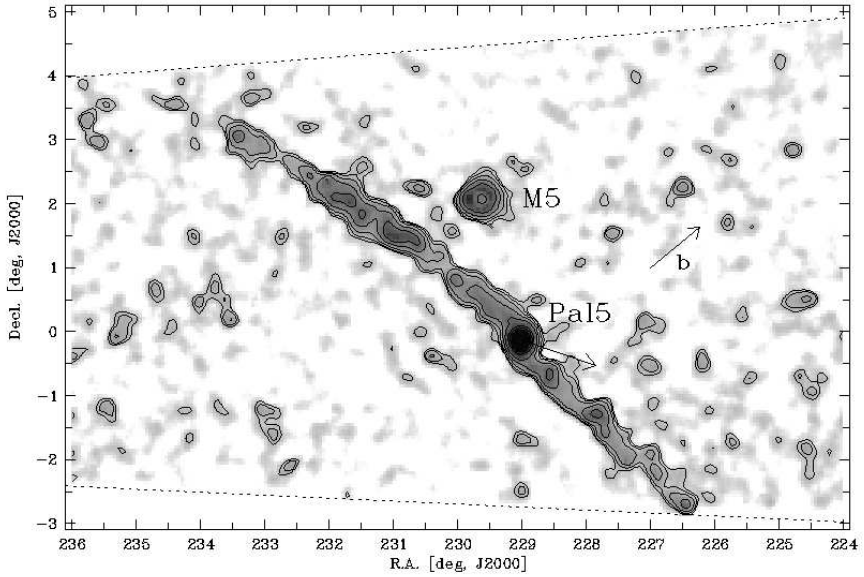

Figure 8.9 Tidal streamers emerging from the globular cluster Pal 5. The plot shows the surface density of stars whose distances are consistent with the cluster distance. The clump labeled "M5" is a residual feature from the unrelated cluster M5. The arrow labeled "b" shows the direction of increasing Galactic latitude. The dotted lines mark the borders of the field. See Grillmair & Dionatos (2006) for maps of the streamers at even larger distances from the cluster. From Odenkirchen et al. (2003), by permission of the AAS.

In terms of the new variables, equation (8.112) becomes

$$\Delta\boldsymbol{\theta}'(t) \simeq t\,\widetilde{\mathbf{D}}\cdot\Delta\mathbf{J}' \quad \text{or} \quad \Delta\theta_i'(t) \simeq t\lambda_i\Delta J_i' \quad \text{(no summation over } i\text{)}. \tag{8.115}$$

This result shows that the cloud of escaped stars spreads into three of the six phase-space dimensions, at rates determined by the initial spread in actions and the eigenvalues λ_i of the matrix $\mathbf{D}$. Small satellites have a smaller spread in actions and so disperse more slowly. If one of the three eigenvalues is zero, or at least much smaller than the other two, the cloud will expand in two dimensions in phase space, creating a sheet; this is the situation for tidal streamers in a spherical host galaxy. If two of the three eigenvalues are zero, the cloud will expand in one dimension to produce a filament; this is the situation in a Keplerian potential. Even when two or more of the eigenvalues of $\mathbf{D}$ are non-zero, usually one is large enough compared to the others that the disrupted stars form a relatively thin tail, which is called a **tidal streamer** or **tail** (see Figure 8.8)—usually the term "tail" is reserved for the long, prominent, massive streamers formed in major mergers of two disk galaxies.

Known tidal streamers are associated with the Magellanic Clouds (the Magellanic Stream, already described in §8.1.1c), the globular cluster Pal 5 (Figure 8.9), and the Sagittarius galaxy (Figure 8.10).

Unlike comet tails, tidal streamers are symmetrical structures that both lead and lag the satellite along its orbit. For example, in Figure 8.9 the upper streamer is made up of stars that have longer orbital periods than the cluster, and hence trail behind it; conversely, the streamer at lower right contains stars that are on shorter-period orbits, and race ahead of it.

In Chapter 9 we shall argue that galaxies form by hierarchical merging of smaller subunits. In the merging process, these subunits are disrupted by tidal forces, and the debris—both stars and dark matter—forms a vast web of tidal streamers. The number of streamers per unit volume and the corresponding degree of irregularity in the mass distribution of the halo depend on the distance from the center of the galaxy: at small radii, the galaxy is hundreds or thousands of crossing times old and the tidal streamers are thoroughly phase-mixed, while at large radii subunits are falling in for the first time and the substructure will be much more prominent (Helmi, White, & Springel 2003). At any given radius, the substructure is likely to be stronger in the baryons (stars and gas) than in the dark matter, since the baryons are concentrated in the dense centers of the dark-matter halos and thus are less susceptible to tidal forces. Efforts to detect and disentangle this web are still in their infancy.

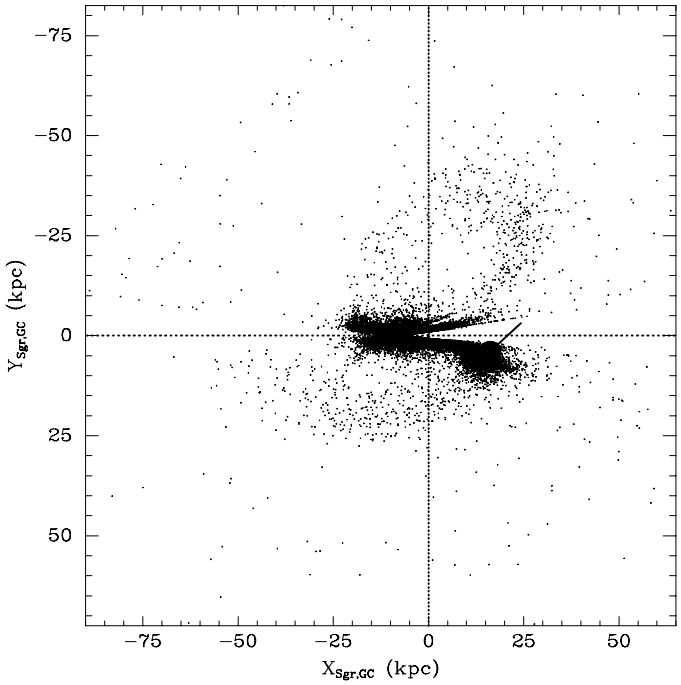

Figure 8.10 The distribution of M-giant stars lying within 7 kpc of the orbital plane of the Sagittarius dwarf galaxy. The figure is a projection onto this orbital plane, which is tipped by 77° from the Galactic plane. The Galactic disk lies along $Y = 0$, the Galactic center is at the origin, and the Sun is at $X \simeq -8\,\mathrm{kpc}$, $Y \simeq 0$ (by coincidence, the Sun lies nearly in the Sagittarius orbital plane). Stars that are highly reddened have been removed, which creates the wedge-shaped gap stretching right from the Sun. The Sagittarius galaxy is located at $X \simeq 15\,\mathrm{kpc}$, $Y \simeq 5\,\mathrm{kpc}$, and the line extending from it indicates the direction of its velocity vector. Tidal debris from the galaxy is evident as the prominent arc passing through $(X, Y) \simeq (25\,\mathrm{kpc}, -30\,\mathrm{kpc})$ above the Galactic plane, and through $(X, Y) \simeq (-15\,\mathrm{kpc}, 15\,\mathrm{kpc})$ below the plane. Most of the width of the arcs is probably due to errors in the stellar distances. From Majewski et al. (2003).

8.4 Encounters in stellar disks

The velocity distribution of stars in the solar neighborhood is approximately described by the Schwarzschild distribution introduced in §4.4.3 (see also Problem 8.17). In this DF, the number of stars with velocity $\mathbf{v}$ in a small range $\mathrm{d}^3\mathbf{v}$ is

$$f(\mathbf{v})\mathrm{d}^3\mathbf{v} = \frac{n_0\,\mathrm{d}^3\mathbf{v}}{(2\pi)^{3/2}\sigma_R\sigma_\phi\sigma_z}\exp\left[-\left(\frac{v_R^2}{2\sigma_R^2}+\frac{v_\phi^2}{2\sigma_\phi^2}+\frac{v_z^2}{2\sigma_z^2}\right)\right]. \qquad (8.116)$$

Here n_0 is the number of stars per unit volume, σ_R, σ_ϕ, and σ_z are the velocity dispersions along the three axes of a cylindrical coordinate system

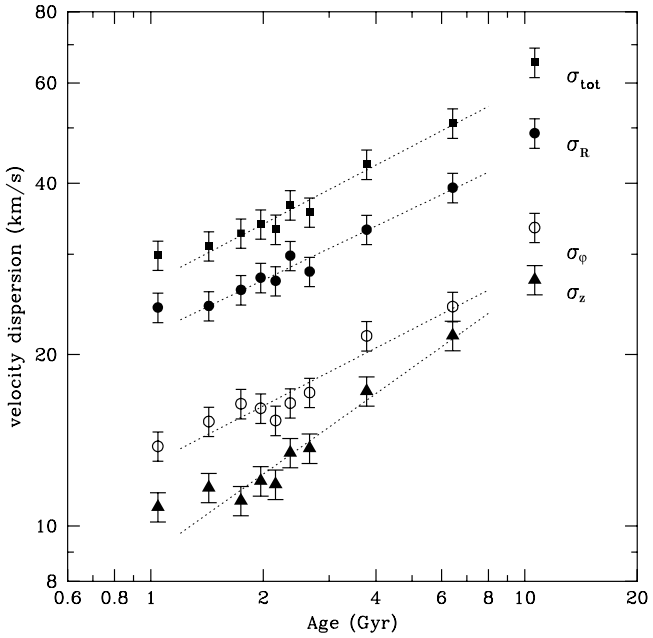

Figure 8.11 The velocity dispersion of stars in the solar neighborhood as a function of age, from Nordström et al. (2004). From bottom to top, the plots show the vertical dispersion σ_z, the azimuthal dispersion σ_ϕ, the radial dispersion σ_R, and the RMS velocity $(\sigma_R^2+\sigma_\phi^2+\sigma_z^2)^{1/2}$. The lines show fits of the form $\sigma_i \propto t^\alpha$ where t is the age; from bottom to top the best-fit exponents α are 0.47, 0.34, 0.31, and 0.34.

centered on the center of the Galaxy, and $\mathbf{v}$ is the velocity relative to the velocity of a circular orbit passing through the solar neighborhood (the Local Standard of Rest; see §1.1.2). The radial and azimuthal dispersions are approximately related by

$$\frac{\sigma_\phi}{\sigma_R} = \frac{\kappa}{2\Omega}, \tag{8.117}$$

where κ and Ω are the epicycle frequency and the azimuthal frequency for nearly circular orbits in the solar neighborhood (see eq. 3.100 and Problem 4.43). Equation (8.116) states that the density of stars in velocity space is constant on ellipsoids with principal axes σ_R, σ_ϕ and σ_z, called velocity ellipsoids in §4.1.2.

Although the shape of the velocity ellipsoid is approximately the same for different types of stars, its size is not: the dispersions σ_i $(i = R, \phi, z)$ of cool, red stars are almost three times as large as those of hot, blue stars (BM Figure 10.12 and Tables 10.2 and 10.3). Since blue stars are young, while red stars are a mixture of mostly old and a few young stars, this trend suggests that stars are born on nearly circular or "cold" orbits, and as a

stellar population ages it "heats up" in the sense that the dispersions σ_i increase. This hypothesis can be confirmed by measuring the age-velocity dispersion relation for nearby stars (Figure 8.11). These observations show that $\sigma_i \propto t^\alpha$, where $\alpha \simeq 0.5$ for σ_z and $\alpha \simeq 0.3$ for both σ_R and σ_ϕ—the exponent is necessarily the same for these two dispersions, because they are related by equation (8.117).

We refer to the steady increase of these dispersions with time as **disk heating**, and in this section we investigate the dynamics of this process. A natural first step in this investigation is to wonder whether disk heating can be due to the accumulation of small velocity kicks from passing stars. This process was described briefly in §1.2.1 and more thoroughly in §7.4. In particular, in the discussion following equation (7.106) we saw that encounters between stars in the solar neighborhood have a negligible effect on their velocities over the age of the Galaxy. Thus we must seek other explanations.

The simplest mechanism for disk heating is encounters with hypothetical massive objects in the dark halo, or MACHOs. This process was investigated in §7.4.4, where we found that the predicted time dependence of the velocity dispersion σ_R is incorrect. Moreover, the required MACHO mass appears to be incompatible with observations of wide binary stars in the halo (§8.2.2e). We therefore examine other possibilities.

8.4.1 Scattering of disk stars by molecular clouds

Long before molecular clouds were detected, Spitzer & Schwarzschild (1951, 1953) suggested that encounters between disk stars and massive gas clouds might be responsible for the random velocities of old disk stars. In Figure 8.12 we illustrate how a molecular cloud or other mass m on a circular orbit in a disk affects the orbits of nearby stars. Since the cloud mass is $\lesssim 10^{-5}$ times the mass of the Galaxy, we may use Hill's approximation (§8.3.2), in which the cloud is at rest at the origin of a rotating Cartesian coordinate system, with the x axis pointing radially outward and the y axis in the direction of rotation. For simplicity we neglect motion perpendicular to the x–y plane. The stellar trajectories are given by the equations of motion (8.97), where the cloud potential $\Phi_{\rm s} = -Gm/(x^2+y^2)^{1/2}$. In the figure, we write the distances in terms of the Jacobi radius of the cloud (eq. 8.106),

$$r_{\rm J} = \left(\frac{Gm}{4\Omega_0 A_0}\right)^{1/3} = 52\,{\rm pc}\left(\frac{m}{10^5\,\mathcal{M}_\odot}\frac{\Omega_0}{A_0}\right)^{1/3}\left(\frac{220\,{\rm km\,s^{-1}}}{v_{\rm c}}\frac{R_0}{8\,{\rm kpc}}\right)^{2/3}. \tag{8.118}$$

In these units, the equations of motion are independent of m, so Figure 8.12 applies to clouds of any mass.

The figure shows only stars on initially circular orbits that are larger than the cloud's orbit. The behavior of orbits that are smaller than the cloud's can be deduced by reflecting the orbits shown through the origin of

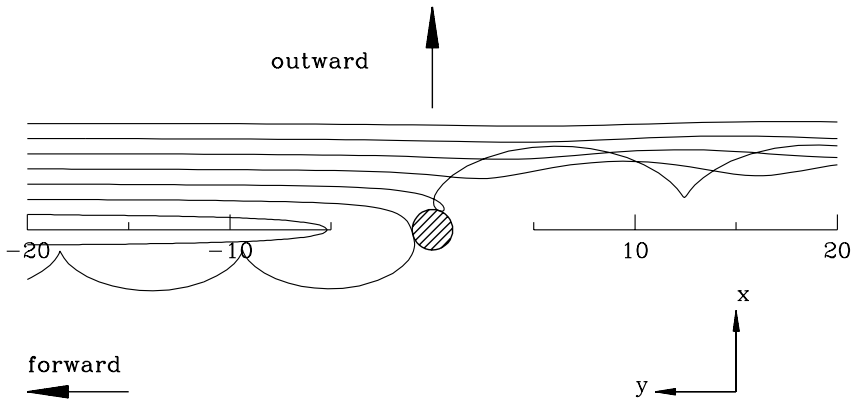

Figure 8.12 The trajectories of stars near a mass point in a disk. The orbit of the mass point is circular, as are the initial orbits of the stars. The coordinate frame co-rotates with the mass point, which is therefore fixed at the origin; in this expanded view of the area near the mass point, upwards is radially outward and the mass point is traveling to the left in an inertial frame. All orbits are restricted to the $z = 0$ plane. Circular orbits appear as straight horizontal lines. A sequence of seven orbits is shown, all initially circular with radius slightly larger than that of the point mass, so it overtakes them. The behavior of orbits inside the point mass is given by reflecting these orbits through the origin. The trajectories of the stars are described by equations (8.97). The disk is assumed to have a flat circular-speed curve, so Oort's constant $A_0 = \Omega_0/2$ and the epicycle frequency $\kappa_0 = \sqrt{2}\Omega_0$. Distances are measured in units of the Jacobi radius $r_{\rm J}$ of the mass point (eq. 8.106), which is also the radius of the circle representing its location.

the figure. The initial orbits shown have angular speeds that are smaller than the cloud's, so the cloud overtakes them (i.e., they move to the right in the cloud frame of reference that is used in the figure). If the initial difference in orbital radii $\Delta r \lesssim r_{\rm J}$, the encounter simply reverses the direction of the orbit relative to the cloud, without imparting any significant epicycle motion.[11] If $\Delta r \sim r_{\rm J}$, the encounter imparts significant epicycle motion—the epicycle amplitude is comparable to Δr and the encounter may or may not reverse the overall direction of motion of the orbit relative to the cloud. For $\Delta r \gtrsim r_{\rm J}$, the star passes the cloud and acquires a small epicyclic motion. It is this excitation of epicycle motion by encounters with clouds that warms the disk.

We may estimate the efficiency of this process by using the impulse approximation to find the radial velocity acquired by a star that is initially on a circular orbit of radius R. If $R_{\rm c}$ is the radius of the cloud orbit, then in the sheared sheet the star's initial orbit is $x = R - R_{\rm c} \equiv b = \textit{constant}$,

[11] This is an example of the donkey effect, described in Box 3.3. No epicycle motion is excited because the star approaches the cloud slowly, so its eccentricity is an adiabatic invariant. For a comprehensive discussion of the trajectories in this problem, see Petit & Hénon (1986).

$y(t) = constant - 2A_0bt$ (eq. 8.99). Integrating the gravitational attraction of the potential $\Phi_s(\mathbf{x})$ along this trajectory yields

$$\dot{x} = -\int_{-\infty}^{\infty} \mathrm{d}t \frac{Gmb}{[b^2 + y^2(t)]^{3/2}} = -\frac{Gm}{A_0 b^2}. \tag{8.119}$$

In the impulse approximation, immediately after the encounter we have

$$x = b \quad ; \quad \dot{x} = -\frac{Gm}{A_0 b^2} \quad ; \quad \dot{y} = -2A_0 b, \tag{8.120}$$

and the corresponding epicycle energy and amplitude are given by equations (8.103),

$$\Delta E_x = \tfrac{1}{2}\dot{x}^2 = \frac{f^2}{2}\left(\frac{Gm}{A_0 b^2}\right)^2 \quad ; \quad X = f\frac{Gm}{A_0 \kappa_0 b^2}, \tag{8.121a}$$

where $f = 1$. We have introduced the correction factor f because the impulse approximation is not accurate: since the radial velocity oscillates with the epicycle frequency κ_0, the impulse approximation requires that the duration of the encounter is much less than the epicycle period. In fact, the duration is $\approx b/\dot{y} = (2A)^{-1}$, which is comparable to the epicycle period $2\pi/\kappa_0$. Hence the impulse approximation makes an error of order unity in the epicycle energy. The correct derivation (Julian & Toomre 1966 and Problem 8.20) yields

$$f = \frac{\Omega_0}{A_0} K_0\Big(\frac{\kappa_0}{2A_0}\Big) + \frac{\kappa_0}{2A_0} K_1\Big(\frac{\kappa_0}{2A_0}\Big), \tag{8.121b}$$

where K_ν is a modified Bessel function (Appendix C.7). For a flat circular-speed curve, $A = \frac{1}{2}\Omega$, $\kappa = \sqrt{2}\Omega$, and $f = 0.923$; for a Keplerian curve $f = 1.680$.

This result is based on linear perturbation theory, and hence is valid only when the epicycle amplitude X induced by the encounter is much smaller than the impact parameter b. Requiring that $X \lesssim b$ implies that

$$\frac{b}{r_J} \gtrsim \left(\frac{4f\Omega_0}{\kappa_0}\right)^{1/3} \approx 1, \tag{8.122}$$

where we have used equation (8.118) for the Jacobi radius. Thus, equations (8.121) are valid for encounters with impact parameters that are large compared to the Jacobi radius, and they show that the epicycle energy excited in an encounter falls off as b^{-4}. On the other hand, Figure 8.12 shows that encounters at small impact parameters, $b \lesssim r_J$, simply switch the star from one circular orbit to another, with no sensible increase in the star's random velocity. These considerations imply that the strongest encounters

have $b \sim r_{\rm J}$, and hence justify our treatment of the molecular cloud as a point mass: since the typical cloud radius $R \sim 10\,{\rm pc}$ is much smaller than its Jacobi radius (8.118), the non-zero cloud size has little influence on the rate of disk heating.

The speeds with which the stars in Figure 8.12 approach the scattering cloud are due entirely to the differential rotation of the galactic disk, $\dot{y} = -2A_0 b$. Once the stars have acquired non-zero epicycle energy, we have to consider two types of encounter. For large impact parameters b or small epicycle amplitudes X, the approach speed is still dominated by the contribution from differential rotation (**shear-dominated** encounters), but at impact parameters $b \lesssim \kappa_0 X/A_0$, the encounter geometry will be determined mainly by the star's epicyclic motion (**dispersion-dominated** encounters).

In contrast to shear-dominated encounters, dispersion-dominated encounters usually *can* be treated using the impulse approximation. In this approximation, the magnitude of the velocity change $\Delta\mathbf{v}$ in a single encounter is proportional to v^{-1} where v is the encounter velocity. Thus the average change in epicycle energy E_x is proportional to v^{-2}, and since the number of encounters per unit time is proportional to v and $v^2 \sim E_x$ we expect (cf. Problem 8.21)

$$\frac{{\rm d}E_x}{{\rm d}t} \propto v^{-1} \propto \frac{1}{\sqrt{E_x}}. \tag{8.123}$$

Integrating this result we find that $E_x \propto t^{2/3}$ so the velocity dispersion $v \propto t^\alpha$ with $\alpha = \frac{1}{3}$. This simple calculation somewhat overestimates the rate of growth of the dispersion, since the thickness of the stellar disk is larger than the thickness of the cloud layer, so stars spend a smaller and smaller fraction of their time in the cloud layer as the vertical dispersion, and the resulting thickness of the stellar disk, continue to grow. The numerical calculations described below are consistent with this argument, suggesting that $\alpha \simeq 0.2$–0.25 for heating by molecular clouds. This value is too low to match the observations shown in Figure 8.11—just the opposite problem from MACHO-dominated heating, which gives an exponent that is too large (eq. 7.102).

We have shown that encounters with clouds "heat" the disk, in the sense that the mean epicycle energy increases with time. It is instructive to ask where this energy comes from, since the total energy or Jacobi integral of the star (eq. 8.104) is conserved during each encounter. The first of equations (8.103) shows that in a razor-thin disk the difference in epicycle energy E_x before and after the encounter is equal and opposite to the difference in $\frac{1}{2}A_0 L^2/B_0$; in most galactic potentials $A_0/B_0 < 0$ so we conclude that an increase in E_x is accompanied by an increase in $|L|$ or in $|x_{\rm g}|$ (eq. 8.102), where $x_{\rm g}$ is the difference in radius between the guiding center of the stellar orbit and the orbital radius of the molecular cloud. In words, the gravitational interaction with the cloud *repels* the stars, in the sense that their mean

orbital radius is shifted away from the cloud. Thus the energy to heat the disk comes from a redistribution of the surface mass density of the stars in the disk; the cloud acts as a catalyst to expedite this redistribution of energy but does not contribute any of its own energy to the disk heating.

Numerous authors have estimated the rate at which star-cloud encounters heat disks (Spitzer & Schwarzschild 1951, 1953; Jenkins 1992; Hänninen & Flynn 2002). The best estimate of the number density and masses of molecular clouds in the solar neighborhood leads to a rate of velocity-dispersion growth that is too small by a factor of two or more; but the heating rate is likely to be enhanced by the swing-amplified response or spiral wake induced in the stellar disk by the gravitational field from the molecular cloud, which can be several times more massive than the cloud itself (Julian & Toomre 1966; Julian 1967).

These studies also show that the predicted ratio σ_z/σ_R of the vertical and radial dispersions is $\simeq 0.6$ (Ida, Kokubo, & Makino 1993), not far from the observed ratio of 0.5. However, the predicted age-velocity dispersion relation is approximately a power law, $\sigma_i \propto t^\alpha$, with exponent $\alpha \simeq 0.2$–0.25. This is significantly lower than the observed exponent, which is 0.3 for σ_R and σ_ϕ and even larger for σ_z (Figure 8.11). This result suggests that molecular clouds are unlikely to be the primary cause of disk heating.

8.4.2 Scattering of disk stars by spiral arms

The disks of spiral galaxies are far from smooth. Gas, dust, and young stars are always concentrated into spiral arms. Spiral features are also found in the old stars that make up most of the mass of galactic disks (§6.1.2), so it is natural to ask whether the gravitational fields of spiral features, like the fields from molecular clouds, are able to heat galactic disks (Barbanis & Woltjer 1967).

Consider a weak spiral potential with pattern speed $\Omega_{\rm p}$,

$$\Phi_{\rm s}(R,\phi,t) = \epsilon F(R)\cos[f(R) + m\phi - \Omega_{\rm p}t], \tag{8.124}$$

where $\epsilon \ll 1$. To illustrate the effect of this potential on a stellar orbit, we shall make two assumptions that simplify the algebra but still retain most of the important dynamics: (i) we work in the sheared-sheet approximation (§8.3.2); (ii) we consider only tightly wound spirals, for which the wavenumber $k \equiv {\rm d}f/{\rm d}R$ is large compared to $1/R$ (eq. 6.4).

The sheared-sheet approximation is valid in a neighborhood of the disk centered at a point $[R_0, \phi_0(t)]$ that rotates at the circular angular speed $\dot\phi_0 = \Omega_0 = \Omega(R_0)$. We expand the spiral potential in a Taylor series around this point, using the coordinates $x = R\cos(\phi-\phi_0) - R_0 \simeq R - R_0$ and $y = R\sin(\phi-\phi_0) \simeq R_0(\phi-\phi_0)$. In this neighborhood, we can approximate

the shape function as $f(R) \simeq f(R_0) + kx$. Since the amplitude $F(R)$ varies slowly, it can be replaced by a constant, $F_0 \equiv F(R_0)$. Thus we have

$$\begin{aligned}\Phi_{\rm s}(x,y,t) &= \epsilon F_0 \cos[f(R_0) + kx + my/R_0 + m\phi_0 - \Omega_{\rm p} t] \\ &= \epsilon F_0 \cos[kx + my/R_0 + m(\Omega_0 - \Omega_{\rm p})t + \text{constant}].\end{aligned} \tag{8.125}$$

We now substitute this potential into the equations of motion (8.97) of the sheared sheet, neglecting motion in the z-direction perpendicular to the disk plane:

$$\begin{aligned}\ddot{x} - 2\Omega_0 \dot{y} - 4\Omega_0 A_0 x &= \epsilon k F_0 \sin[kx + my/R_0 + m(\Omega_0 - \Omega_{\rm p})t + \text{constant}]; \\ \ddot{y} + 2\Omega_0 \dot{x} &= \frac{\epsilon m}{R_0} F_0 \sin[kx + my/R_0 + m(\Omega_0 - \Omega_{\rm p})t + \text{constant}].\end{aligned}$$

Since the wave is assumed to be tightly wound, its pitch angle is small so $|k| \gg m/R_0$ (eq. 6.7). Thus the right side of the second equation is much smaller than the corresponding term in the first, and can be neglected. The second equation can then be integrated to yield $\dot{y} + 2\Omega_0 x = \textit{constant}$, and this can be substituted into the first equation to give

$$\ddot{x} + \kappa_0^2 x + \text{constant} = \epsilon k F_0 \sin[kx + my/R_0 + m(\Omega_0 - \Omega_{\rm p})t + \text{constant}]; \tag{8.126}$$

here κ_0 is the epicycle frequency (8.100). The constant on the left side can be dropped, since it can be absorbed by a shift in the origin of the x-coordinate.

In the absence of a spiral ($F_0 = 0$) the solution to this equation is given by equations (8.99); we shall assume that the unperturbed motion is circular, so the trajectory is $\mathbf{x}_0(t) = (x_{\rm g}, y_{\rm g0} - 2A_0 x_{\rm g} t)$. Now consider how this motion is modified by the weak spiral potential on the right side of equation (8.126). We write the trajectory as $\mathbf{x}(t) = \mathbf{x}_0(t) + \epsilon \mathbf{x}_1(t)$, where $\epsilon \mathbf{x}_1(t)$ is the perturbation induced by the spiral. Then the terms of order ϵ in equation (8.126) yield

$$\begin{aligned}\ddot{x}_1 + \kappa_0^2 x_1 &= k F_0 \sin[kx_{\rm g} + m(y_{\rm g0} - 2A_0 x_{\rm g} t)/R_0 + m(\Omega_0 - \Omega_{\rm p})t + \text{constant}] \\ &= k F_0 \sin(kx_{\rm g} + \omega t + c).\end{aligned} \tag{8.127}$$

In the last expression we have absorbed $y_{\rm g0}$ in the constant c, and set $\omega = m(\Omega_0 - 2A_0 x_{\rm g}/R_0 - \Omega_{\rm p})$; this is the frequency at which the unperturbed orbit encounters successive crests of the spiral potential.

This equation can be solved to yield

$$x_1(t) = \frac{k F_0}{\kappa_0^2 - \omega^2} \sin(kx_{\rm g} + \omega t + c). \tag{8.128}$$

The solution diverges when $\omega = \pm\kappa_0$. These points can be thought of as the Lindblad resonances of the sheared sheet: at these locations, like the Lindblad resonances in a disk, the frequency of excitation by the spiral potential

coincides with the frequency κ_0 of the particle's natural radial oscillation. This result is a close analog of equation (3.148), which was derived in the context of weak bars.

Equation (8.128) shows that the spiral potential imposes a forced radial oscillation on the star but does not lead to any steady growth in the radial oscillation $x_1(t)$. In other words, *a spiral potential with a fixed pattern speed cannot heat the disk*, except perhaps at the Lindblad resonances where our simple derivation fails.

This result implies that disk heating requires *transitory* rather than steady spiral patterns. To illustrate this, let us multiply the potential (8.124) or (8.125) by a Gaussian function of time, $p(t) = (2\pi s^2)^{-1/2} \exp(-\frac{1}{2}t^2/s^2)$. Equation (8.127) is thereby modified to read

$$\ddot{x}_1 + \kappa_0^2 x_1 = kF_0 p(t) \sin(kx_{\rm g} + \omega t + c), \tag{8.129}$$

which has the solution

$$x_1(t) = X_1 \cos(\kappa_0 t + \alpha_1) + \frac{kF_0}{\kappa_0} \int_{-\infty}^{t} {\rm d}t'\, p(t') \sin(kx_{\rm g} + \omega t' + c) \sin[\kappa_0 (t - t')], \tag{8.130}$$

where X_1 and α_1 are arbitrary constants. Inserting the chosen form for $p(t)$ and setting the amplitude X_1 of the free oscillation to zero, we obtain

$$\begin{aligned} x_1(t \to \infty) = \frac{kF_0}{2\kappa_0} \Big[& \cos(\kappa_0 t - kx_{\rm g} - c) {\rm e}^{-s^2(\omega+\kappa_0)^2/2} \\ & - \cos(\kappa_0 t + kx_{\rm g} + c) {\rm e}^{-s^2(\omega-\kappa_0)^2/2} \Big]. \end{aligned} \tag{8.131}$$

Thus the transitory spiral pattern has induced a permanent epicyclic oscillation. When the characteristic duration of the transient, s, is much greater than the orbital period, the induced epicycle amplitude is strongly peaked near the Lindblad resonances $\omega = \pm\kappa_0$. On the other hand, when the duration of the transient is short, the arguments of the exponential are small and epicycle motion is induced over a wide range of radii in the disk.

This example shows that the ability of spiral structure to heat the disk is strongly dependent on its temporal structure. According to the Lin–Shu hypothesis (§6.1), in which spiral structure is a stationary wave with a single, well-defined pattern speed, disk heating is negligible except at the Lindblad resonances. In such models the disk can be heated over a wide range of radii only if the pattern speed evolves with time, so the Lindblad resonances slowly sweep across most of the disk. On the other hand, if the spiral structure is transient, the whole disk can be heated—this situation is likely to occur in flocculent spirals, intermediate-scale spirals, or grand-design spirals excited by recent encounters.

Let us suppose that a given star is subjected to N independent transient perturbations. Each transient induces an epicyclic motion whose radial

component can be written in the form $x_1(t) = a_i \cos(\kappa_0 t + \alpha_i), i = 1, \ldots, N$, where a_i and α_i are given by equations similar to (8.131). After N transients,

$$\begin{aligned} x_1(t) &= \sum_{i=1}^{N} a_i \cos(\kappa_0 t + \alpha_i) \\ &= \Big(\sum a_i \cos\alpha_i\Big) \cos\kappa_0 t - \Big(\sum a_i \sin\alpha_i\Big) \sin\kappa_0 t \\ &\equiv a_{\rm f} \cos(\kappa_0 t + \alpha_{\rm f}) \end{aligned} \tag{8.132}$$

where

$$a_{\rm f}^2 = \Big(\sum a_i \cos\alpha_i\Big)^2 + \Big(\sum a_i \sin\alpha_i\Big)^2 = \sum_{i,j=1}^{N} a_i a_j \cos(\alpha_i - \alpha_j). \tag{8.133}$$

Since the transients are uncorrelated, the phases of the epicyclic oscillations that they induce are also uncorrelated. Hence on average $\cos(\alpha_i - \alpha_j)$ will be zero when $i \neq j$, and the only terms in the sum that contribute to the final amplitude $a_{\rm f}$ will be those with $i = j$. Thus $a_{\rm f}^2 \simeq N\langle a^2\rangle$, where $\langle a^2\rangle$ is the mean-square amplitude induced by a single transient. If the rate of occurrence and the strength of new transients are independent of time, we conclude that $a_{\rm f}^2$, and hence the squared velocity dispersion v^2, should grow linearly with time. In other words, $v \propto t^\alpha$, where $\alpha = 0.5$. This behavior holds only so long as $a_{\rm f}$ is not too large: once the epicycle size becomes comparable to the radial wavelength of the spiral arms, the effects of the spiral tend to average out over the epicycle period, so the heating is weaker—this is the same effect that leads to the reduction factor in the WKB dispersion relation for spiral waves, as described in §6.2.2d. Estimates of the heating rate at larger amplitudes can be obtained using the Fokker–Planck equation (Jenkins & Binney 1990; Jenkins 1992) or numerical simulations (De Simone, Wu, & Tremaine 2004); these calculations show that α can vary between 0.25 and 0.5 depending on the properties of the spiral transients (duration, strength, pitch angle, etc.). This range of α is nicely consistent with the observed exponent for the growth of the radial dispersion, $\alpha \simeq 0.3$, and provides a substantially better fit than the values predicted for heating by molecular clouds.

The radial and azimuthal velocity dispersions are related by equation (8.117), so the exponent in the age-velocity dispersion relation must be the same for these two axes of the velocity ellipsoid. However, spiral structure cannot excite velocities in the z-direction effectively, since its spatial and temporal scales are much larger than the amplitude or period of oscillations perpendicular to the disk plane. Thus, scattering by spiral arms cannot explain the relation between age and the z-velocity dispersion σ_z. Probably gravitational scattering by molecular clouds redistributes the radial and azimuthal velocities into the direction perpendicular to the plane (Carlberg

1987; Jenkins & Binney 1990). Thus molecular clouds are responsible for the shape, but not the size, of the velocity ellipsoid.

Transient spiral arms have other interesting consequences for the distribution of disk stars. Strong transients can produce long-lived clumps of stars in velocity space, sometimes called star streams or moving groups (see page 327 and De Simone, Wu, & Tremaine 2004). Spiral waves also redistribute the angular momenta of disk stars, leading to substantial inward and outward migration of individual stars over the lifetime of the Galaxy (Sellwood & Binney 2002).

8.4.3 Summary

There is little doubt that irregularities in the Galaxy's gravitational field heat the disk and thereby determine the velocity distribution of disk stars. It is less clear *which* irregularities dominate this process. We have discussed the influence of hypothetical massive objects in the dark halo (MACHOs), molecular clouds, and transient spiral arms. Other possibilities include merging satellite galaxies (Walker, Mihos, & Hernquist 1996; Velázquez & White 1999), substructure in the dark halo (Benson et al. 2004), or the Galactic bar (Kalnajs 1991; Dehnen 2000a). The simplest explanation that appears to be consistent with most of the observations is the combined effects of spiral transients and molecular clouds.

8.5 Mergers

So far we have investigated galaxy mergers and encounters through limiting cases that are analytically tractable. For example, minor mergers occur through dynamical friction (§8.1), which leads to gradual orbital decay, and as the orbit shrinks tidal forces and tidal shocks (§§8.2 and 8.3) become stronger and stronger, until either the small galaxy is completely disrupted or its core comes to rest at the center of the larger galaxy.

In major mergers the physical processes are qualitatively similar, but harder to quantify. The relative velocity of the centers of mass of the two galaxies is converted into randomly directed velocities of their individual stars—the same process as dynamical friction—but the conversion is so rapid that the galaxies merge into a single steady-state system within a few crossing times. Thus, numerical simulations such as the one shown in Figure 8.1, rather than analytic arguments, are the primary tool for understanding major mergers.

In this section we shall focus on features of major mergers that have direct observational consequences; these are important because they provide the "smoking gun" that enables us to identify galaxies that are participating in ongoing mergers, and thus to explore the physics of mergers. Reviews

of interacting and merging galaxies are given by Barnes & Hernquist (1992) and Kennicutt, Schweizer, & Barnes (1998).

8.5.1 Peculiar galaxies

A small fraction of galaxies are found in a highly disturbed state. The importance of these puzzling systems was emphasized by Arp (1966), who compiled an *Atlas of Peculiar Galaxies* containing over 300 such objects (see also Arp & Madore 1987). Arp argued that "if we could analyze a galaxy in a laboratory, we would deform it, shock it, probe it, in order to discover its properties" and that the peculiarities of the galaxies in his atlas offered a range of experiments on galaxies furnished to us by nature, which we should learn from. At one time it was widely believed that unusual systems of this kind were exploding galaxies or galaxies with very strong magnetic fields, but by the early 1970s it became clear that most are actually colliding systems, and that many of these collisions will result in mergers.

Figure 8.13 shows the pair of galaxies NGC 4038/4039 from the Arp atlas. This system consists of overlapping blobs of light from which two curved tails of much lower surface brightness emerge, giving rise to its common name "the Antennae". From end to end, the tails span over 100 kpc. Can this striking morphology be the signature of a merger? In a classic paper, Toomre & Toomre (1972) showed that this is indeed the case. The Toomres studied encounters between disks of massless particles orbiting around point masses: even with this grossly oversimplified model of a galaxy—the disk has no self-gravity, there is no massive halo, and the disk circular-speed curve is Keplerian rather than flat—they were able to show that for a suitable choice of initial conditions, it is possible to find a pair of colliding stellar systems that is remarkably similar to Figure 8.13. We show their model in Figure 8.14. More accurate models that include the self-gravity of the disk and a massive halo largely confirm the Toomres' conclusions (Barnes 1988; Dubinski, Mihos, & Hernquist 1999).

The Toomres' model predicted the line-of-sight velocity at each point in the system. The observed velocities were found to be in complete agreement with the model, and show that the point of closest approach of the two galaxies, when the tails were launched, occurred 0.5 Gyr ago (Hibbard et al. 2001).

The tails seen in the Antennae differ in one important respect from the tidal streamers discussed in §8.3.3. The streamers discussed in that section are composed of stars stripped from small satellites of much larger stellar systems; the streamers are narrow because the satellite is small. In contrast, the two merging systems in the Antennae have comparable size; the tails are narrow because the stars come from cold stellar systems—the disks of the two merging spiral galaxies—so all the stars near a given location have nearly the same initial velocity. Mergers of hot stellar systems of comparable size do not generate narrow tidal tails.

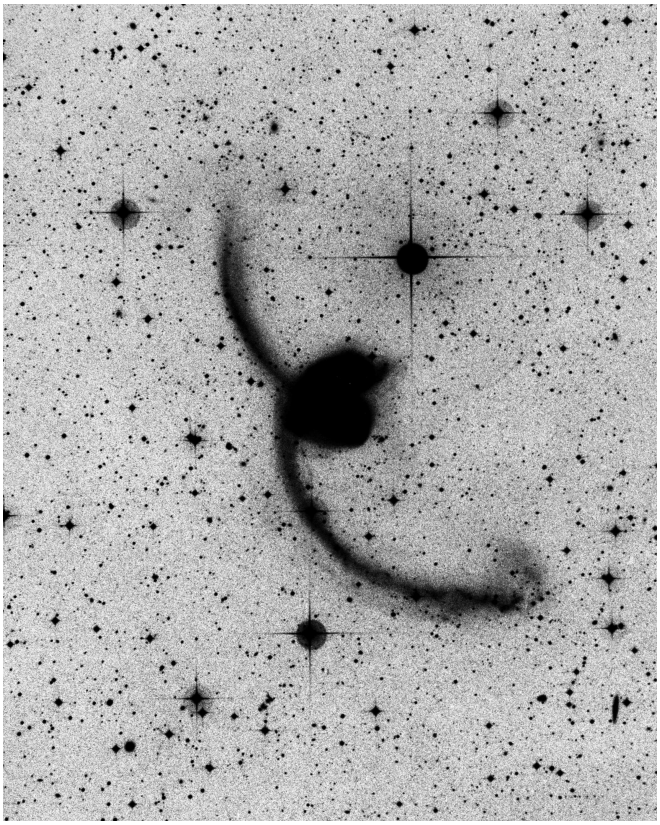

Figure 8.13 The interacting galaxies NGC 4038 and NGC 4039, the "Antennae". This is an overexposed image to emphasize the low surface-brightness tidal tails. The distance from the overlapping blobs at the center to the bright star above and to the right of them is 40 kpc. Courtesy of D. F. Malin and the Anglo–Australian Telescope Board.

Another route to the same conclusion is through the collisionless Boltzmann equation, which shows that the density of stars in phase space is conserved (eq. 4.10). A long-lived tidal tail or streamer must have high phase-space density, since the spatial density must be high if the tail is to be visible against the background galaxy, and the velocity dispersion must be low if it is not to disperse quickly. Thus the progenitor of the tidal tail or streamer must have high phase-space density, a condition that is satisfied by both satellite stellar systems (because their spatial densities are high and their velocity dispersions are low compared to the larger host galaxy) and disks (because the velocity dispersion is low).

Another galaxy with prominent tidal tails that is almost certainly an ongoing merger is NGC 4676 ("the Mice"), shown in Figure 8.15.

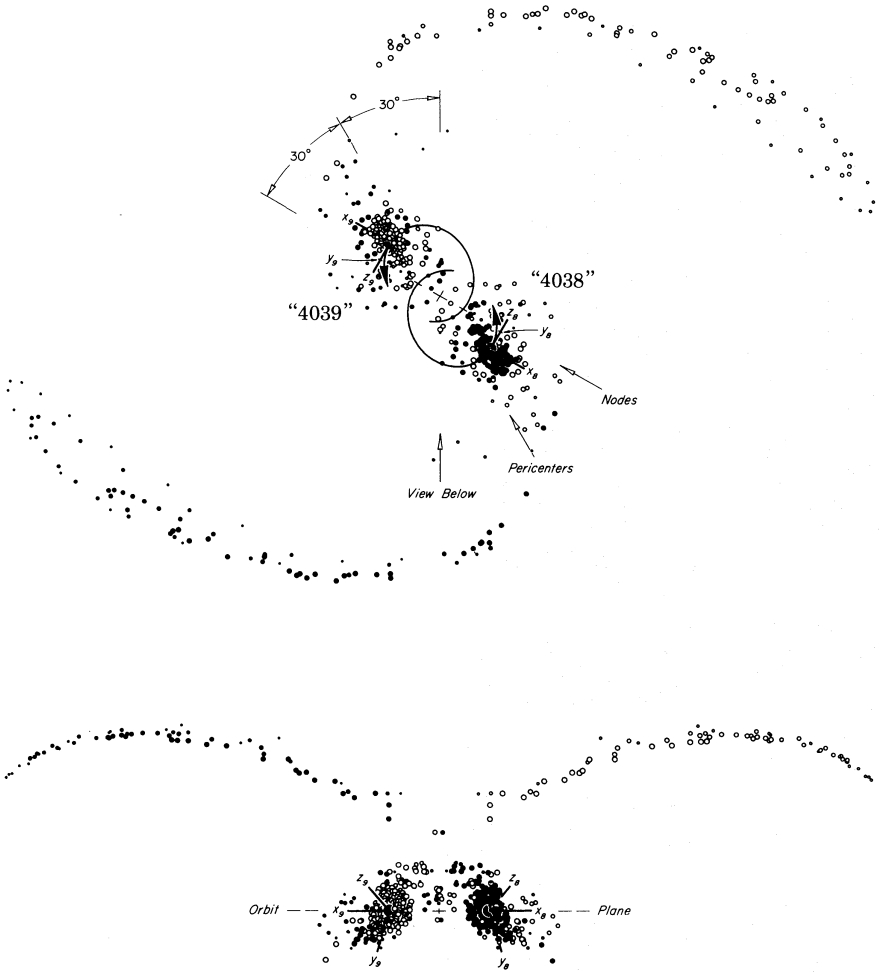

Figure 8.14 A model of the NGC 4038/4039 pair by Toomre & Toomre (1972). Reproduced by permission of *The Astrophysical Journal*.

8.5.2 Grand-design spirals

We have seen in Chapter 6 that grand-design spirals such as M51 (Plate 1) or M81 (Plate 8) often have companion galaxies nearby, and that the gravitational forces from an encounter with a companion can excite a strong but transitory spiral pattern (Figure 6.26). In most cases the orbit of the companion galaxy that excited the spiral will decay by dynamical friction, so the two galaxies are likely to merge in the future. Thus many of the most beautiful and striking spiral galaxies in the sky are likely to be the product of major mergers.

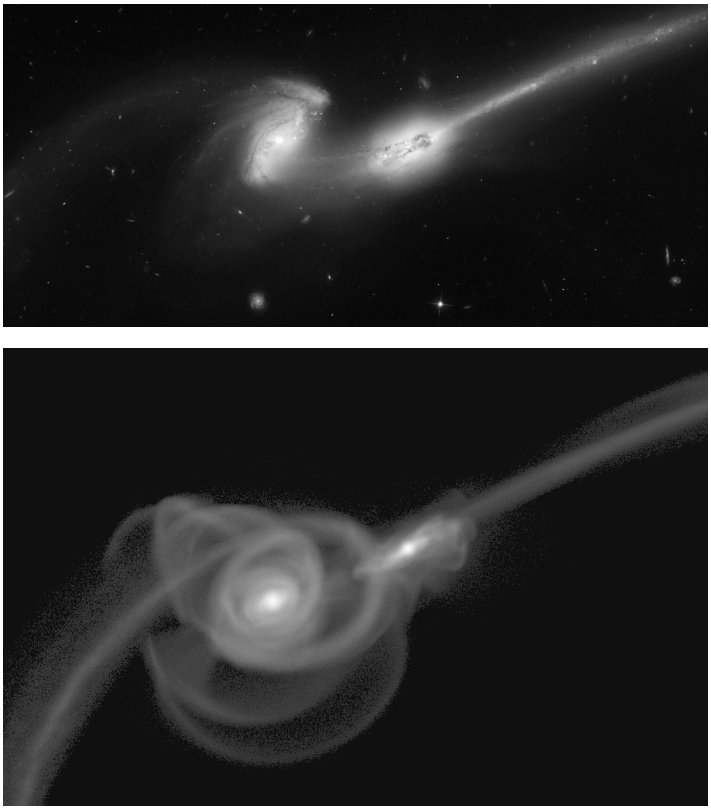

Figure 8.15 The Mice, NGC 4676, a pair of interacting galaxies at a distance of 95 Mpc. Top: optical image from the Hubble Space Telescope. Bottom: an N-body model. Credit for HST image: NASA, H. Ford (JHU), G. Illingworth (UCSC/LO), M. Clampin (STScI), G. Hartig (STScI), the ACS Science Team, and ESA. Credit for N-body model: J. Dubinski (Dubinski & Farah 2006).

8.5.3 Ring galaxies

A handful of galaxies exhibit a distinctive morphology consisting of a luminous ring of young stars that is both rotating and expanding, usually with one or more compact companion galaxies nearby. Figure 8.16 shows one example, the "Cartwheel Galaxy". These remarkable systems are known as **ring galaxies** or sometimes **collisional ring galaxies** (Appleton & Struck–Marcell 1996).[12]

Ring galaxies form when a disk galaxy collides head-on with another system (Lynds & Toomre 1976). The collision excites a radially expanding

[12] These are distinct from the prominent rings that are seen in some barred galaxies, which are thought to arise from rapid star formation in gas that is in resonance with the bar (see §6.5.2d and Buta 1995).

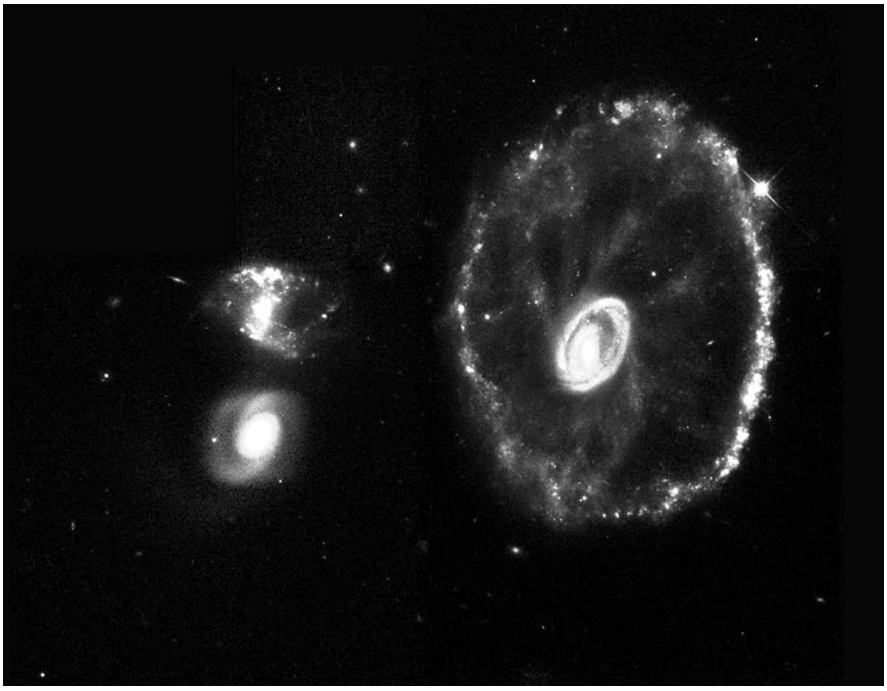

Figure 8.16 The Cartwheel galaxy, the prototypical ring galaxy, at a distance of 125 Mpc. The ring diameter is about 45 kpc. The lower of the two compact stellar systems to the left of the ring—both members of the same group of galaxies as the Cartwheel—is probably responsible for the ring structure. Credit: K. Borne (George Mason University) and NASA.

density wave that triggers star formation in the disk as it passes. The compact systems are the surviving central cores of the colliding galaxies. Ring galaxies are rare—about one in 10^4 galaxies—because they are short-lived, and because they are produced only in collisions with near-zero impact parameter.

We can use the impulse approximation to develop an instructive model of this process, even though this approximation may not hold for all ring galaxies. Consider a singular isothermal sphere that contains a rotating, cold, disk of test particles in the plane $z = 0$, and suppose that it collides with a second singular isothermal sphere having the same circular speed v_c, traveling along the z axis with relative speed $V \gg v_c$. The gravitational potential of each sphere is $\Phi(r) = v_c^2 \ln r$, and in Problem 8.7 it is shown that the change Δv_R in the velocity of a disk star at initial radius R is then

$$\Delta v_R = -2R\frac{v_c^2}{V}\int_R^\infty \frac{\mathrm{d}r}{r\sqrt{r^2 - R^2}} = -\frac{\pi v_c^2}{V}. \tag{8.134}$$

If V/v_c is sufficiently large, the velocity impulse $\Delta v_R/v_c$ will be small, so

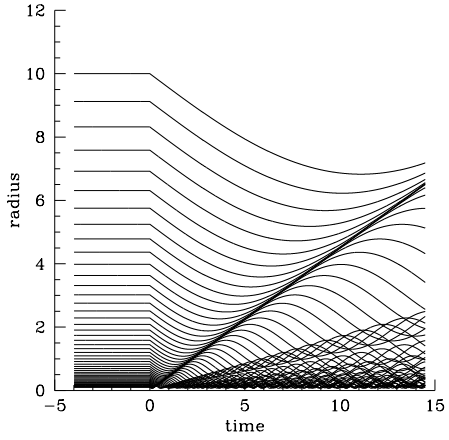

Figure 8.17 The evolution of the radii of particles in a disk after a head-on encounter, as described by equation (8.135), following Toomre (1978). The ratio $V/v_c = 7$.

(i) we may neglect the changes in the target's potential that are generated by the collision; (ii) we can describe the subsequent motion of the disk stars using the epicycle approximation. If the collision is assumed to occur at time $t = 0$, the radius of a star that is initially at R_0 is given by the solution of equation (3.78a) that satisfies the initial conditions $x(0) = R - R_0 = 0$, $\dot{x}(0) = \Delta v_R$:

$$R(R_0, t) = R_0 + \frac{\Delta v_R}{\kappa_0} \sin(\kappa_0 t) \quad (t > 0). \tag{8.135}$$

Here κ_0 is the epicycle frequency at R_0, which is given by $\kappa_0^2 = 2v_c^2/R_0^2$ (eq. 3.79a).

The evolution of the radii of particles in the disk is shown in Figure 8.17. The crowding of particle orbits gives rise to strong axisymmetric density waves that propagate out through the disk. The point of maximum compression of the particle orbits is likely to be a region of enhanced star formation, which we identify with the ring of luminous young stars. The outward propagation of the density waves implies that the region inside the ring should contain older, redder stars that were formed when the ring was smaller, and such radial color gradients are indeed observed in several ring galaxies.

In practice, the collision of two galaxies is never precisely along the z axis of the disk, as assumed in this simple model. However, numerical experiments such as those shown in Figure 8.18 show that whenever an intruder passes close to the center of the target disk on a trajectory that is angled by less than about 30° from the symmetry axis of the disk, a striking ring is generated. If the intruder misses the center of the target just slightly, the dense center of the target galaxy is displaced from the center of the ring, as is observed in the Cartwheel Galaxy.

8.5.4 Shells and other fine structure

Figure 8.19 shows images of NGC 3923 and NGC 1344, which exhibit arclike

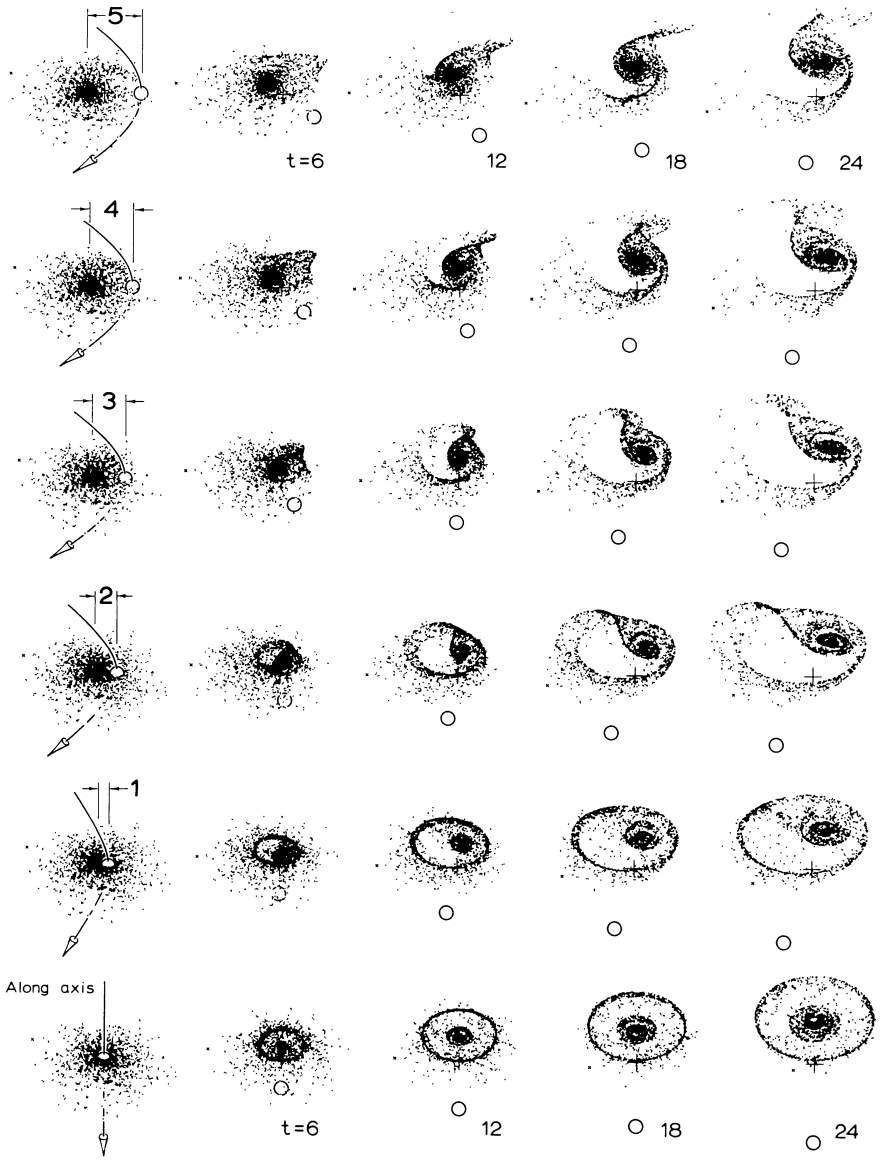

Figure 8.18 Six encounters between a disk of test particles orbiting a point mass and an intruder of half the mass, marked by an open circle. The relative orbit is parabolic, and the system is viewed from 45° above the disk. A ring is generated when the impact parameter is small compared to the size of the disk (bottom three rows). From Toomre (1978), with permission of Springer Science and Business Media.

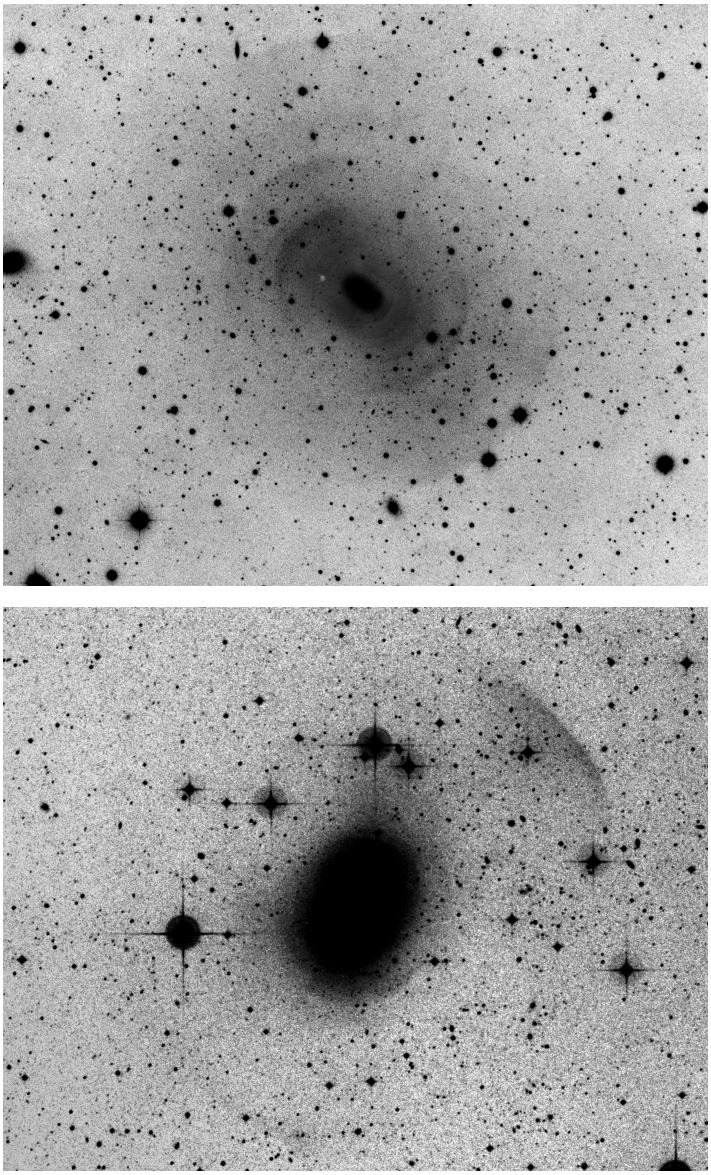

Figure 8.19 The giant elliptical galaxies NGC 3923 (top) and NGC 1344 (bottom) are surrounded by faint shells. The images have been processed to accentuate the shells using a high-pass spatial filter. Courtesy of David Malin, © Anglo–Australian Observatory.

shells in the surface brightness on both sides of the galaxy; careful analysis reveals over 20 such shells in this galaxy. The fraction of otherwise smooth

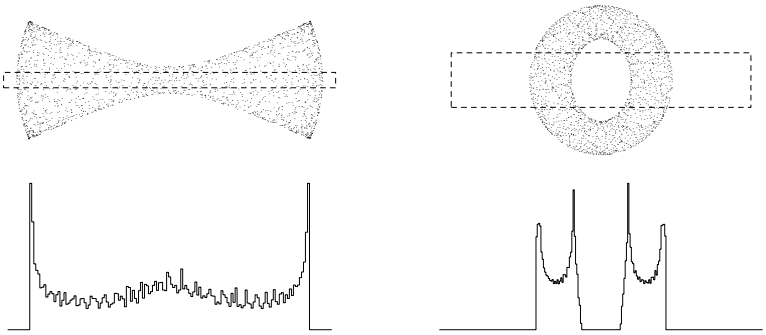

Figure 8.20 The box and loop orbits shown in Figure 3.8. The upper figures show stars at randomly chosen phases on the orbits to give a visual impression of the density distribution in the orbit, and the lower figures show the number of stars as a function of horizontal position within the boxes marked by dashed lines. Note the sharp cusps at the turning points of the orbits.

galaxies (ellipticals and lenticulars) that exhibit shells can be as large as 30–50%, depending on how closely one looks. Shells may also be present in spiral galaxies, but are camouflaged by spiral structure, dust, and irregular star formation in the disk. Spectra show that the shells are composed of stars, not gas. As is often the case in astronomy, the most famous examples of this phenomenon are atypical. The shells in NGC 3923 are exceptionally sharp and numerous, aligned with the major axis of the galaxy, and interleaved in radius (the shells from one side alternate in radius with those from the other side). In contrast, most shell galaxies contain $\lesssim 3$ detectable shells, and these are fainter, more diffuse, and have a less regular geometrical structure.

Other types of fine structure are also seen in elliptical galaxies (Kennicutt, Schweizer, & Barnes 1998), and are given names such as "loops", "ripples", "plumes", "jets", "X-structures", etc. The tidal streamers described in §8.3.3 are also a kind of fine structure.

Most fine structure in galaxies is formed by the same process that forms tidal streamers, namely the disruption of a stellar system that has high phase-space density—either a small, hot galaxy or a large, cold one.

For example, consider the fate of the stars in a small satellite that is disrupted by a host galaxy. Initially the disrupted stars will form a tidal streamer, but eventually the streamer will disperse. If the host potential is regular, the stars will finally spread into a cloud of particles that have similar actions but uniformly distributed angles. Such a cloud gives rise to surface-density distributions such as those shown in Figure 8.20 for box and loop orbits. A box orbit produces an X-shaped structure, while a loop orbit produces an annulus with sharp edges at its inner and outer turning points.

In projection, the edges of these particle distributions can appear as shells, as indicated in the figure.

Another simple example is the disruption of a disk galaxy that is on a radial orbit in a spherical potential (Quinn 1984). This process can be explored by releasing a cloud of test particles in a fixed potential (Figure 8.21). Since the angular momentum of the test particles is nearly zero, the motion can be followed in the two-dimensional (r, v_r) plane. Shells are formed at the turning points of the orbits, and they are interleaved in radius like the shells observed in NGC 3923. The rather special circumstances of the encounter (radial orbit, spherical host potential) are consistent with the observation that most shell galaxies do not exhibit the regular geometrical structure seen in this example.

The much more common case of the disruption of a galaxy on a non-radial orbit can also produce shells, such as those shown in Figure 8.22, but now the shells display the more complex geometry that is encountered in most shell galaxies (Hernquist & Quinn 1988, 1989).

More generally, shells arise when stars are confined to a subspace of lower dimensionality than the full six-dimensional phase space. The projection of this smooth manifold onto the two-dimensional plane of the sky gives rise to caustics, which can be classified using catastrophe theory (Tremaine 1999).

8.5.5 Starbursts

So far we have focused on the effects of mergers on a galaxy's stars, but the effects on its gas—if the galaxy has a gas disk—are even more dramatic. As Toomre & Toomre (1972) wrote, "Would not the violent mechanical agitation of a close tidal encounter—let alone an actual merger—already tend to bring *deep* into a galaxy a fairly *sudden* supply of fresh fuel in the form of interstellar material?" The Toomres' prescient question was answered by Larson & Tinsley (1978), who showed that many merger remnants had anomalous blue colors consistent with young, massive stars formed in a recent **starburst**—a short, intense period of rapid star formation at a rate far exceeding that of a normal galaxy. Since that time a wide variety of observations has confirmed that vigorous star formation occurs in merging galaxies. Among the most striking of these observations is the discovery of almost 10^3 blue objects in the Antennae (Figure 8.13), which appear to be young globular clusters formed in the merger (Whitmore & Schweizer 1995).

The observational link between mergers and star formation was cemented by the discovery of **starburst galaxies**. These are among the most luminous galaxies known, emitting up to $10^{12.5}\,L_\odot$, mostly at infrared wavelengths. This intense emission comes from young stars shrouded in dust and concentrated near the center of the galaxy. The emission is powered by extremely high star-formation rates, as large as $10^3\,\mathcal{M}_\odot\,\mathrm{yr}^{-1}$, compared to a few $\mathcal{M}_\odot\,\mathrm{yr}^{-1}$ in galaxies like the Milky Way. Starburst galaxies usually

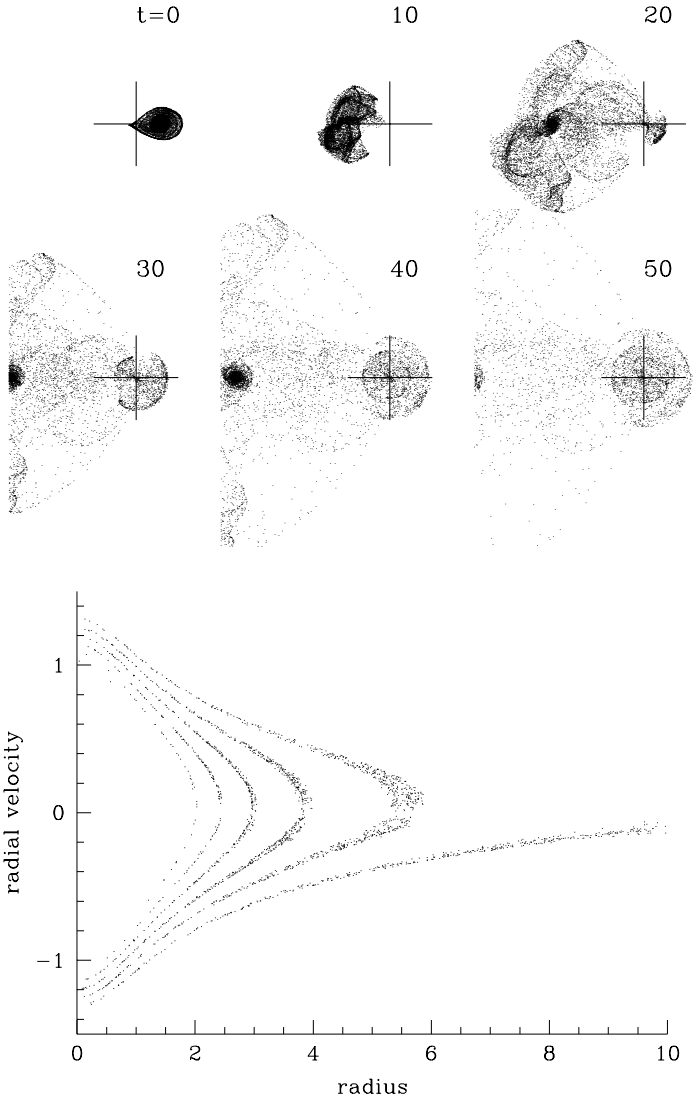

Figure 8.21 The disruption of a disk galaxy on a radial orbit in a spherical potential. The host galaxy is represented by a rigid, fixed Plummer potential (2.44a) with mass $M = 1$ and scale length $b = 1$, and the satellite has mass $m = 0.1$ and is modeled by a rigid Kuzmin-disk potential (2.68a) with scale length $a = 0.5b$, containing 10 000 test particles on initially circular orbits. Top: The cross marks the center of the host galaxy and the length of each arm of the cross is $5b$. The evolution is viewed from a direction normal to the plane of the disk galaxy (the x-y plane). The distribution of test particles is first shown just before the satellite reaches the center of the host, falling in from infinity along the positive x axis, and at intervals of 10 time units thereafter. Bottom: the projection of the test particles onto the radius-radial velocity plane at time $t = 50$ (Quinn 1984).

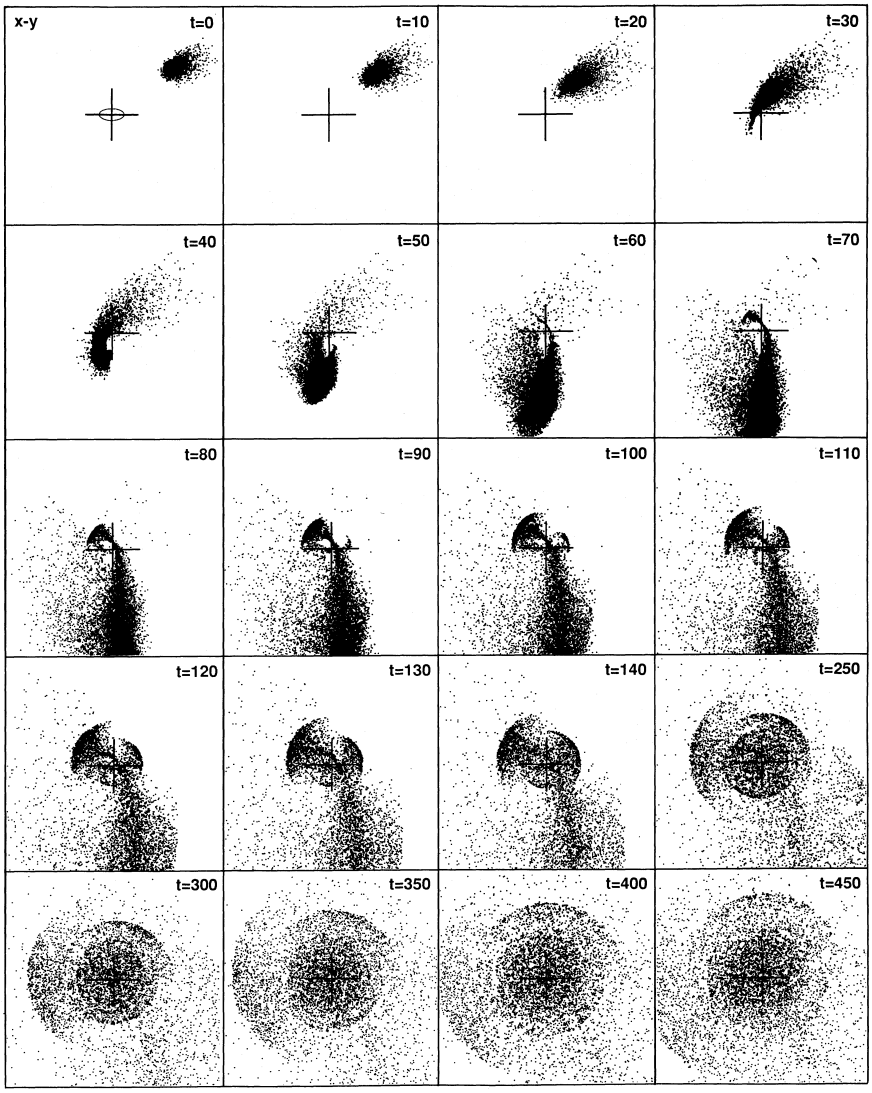

Figure 8.22 The formation of shells in the disruption of a small spherical galaxy on a non-radial orbit. The small ellipse in the first frame represents the approximate location and projected shape of the larger galaxy in the encounter. The simulation uses 20 000 test particles. From Hernquist & Quinn (1989), reproduced by permission of the AAS.

exhibit tidal streamers or other optical features indicative of a recent collision or merger (Sanders & Mirabel 1996; Kennicutt, Schweizer, & Barnes 1998; Kennicutt 1998). The presence of these short-lived features, and the rapid consumption of gas required by the high star-formation rate, imply

that starbursts last for only a few tens of Myr.

The reason for these very high star-formation rates is suggested by numerical simulations of mergers of disk galaxies that contain both gas and stars (Noguchi 1988; Mihos & Hernquist 1996). During the merger, both the gas and stars form strong bars. The gas bar leads the stellar bar, so the gravitational torque from the stars rapidly drains angular momentum from the gas. Remarkably, in a typical major merger the gas can lose up to 90% of its angular momentum in a fraction of an orbital period, thus settling into a dense rotating disk $\lesssim 0.5\,\mathrm{kpc}$ across, in which star formation is likely to be extremely rapid.

8.5.6 The merger rate

The rate at which galaxies merge is a fundamental point of comparison between observations and models of structure formation. The merger rate can be determined from observations using two quite different methods.

The first method is to count the fraction of galaxies showing obvious features of an ongoing or recent major merger, such as tidal tails or starbursts, and combine this fraction with an estimate of how long such features last to determine the merger rate per galaxy. The first attempt of this kind was made by Toomre (1977b), who pointed out that about 10 of the ~ 4000 NGC galaxies[13] show prominent tidal tails, and that these tails probably last no more than $\sim 0.5\,\mathrm{Gyr}$. Thus the rate of major mergers is probably about $10/4000/0.5\,\mathrm{Gyr} \simeq 0.005\,\mathrm{Gyr}^{-1}$ for a luminous galaxy. This estimate neglects two important biases, of opposite sign: first, not all major mergers yield visible tidal tails; second, mergers enhance the star-formation rate, so galaxies experiencing mergers are more luminous than quiescent galaxies, and hence will be over-represented in a flux-limited catalog like the NGC. A crude assumption is that these two biases cancel, leaving the original estimate of 0.005 major mergers per Gyr approximately correct.

A second approach is to count the fraction of galaxies with companions within a given radius, and combine this fraction with estimates of the rate of decay of the companion orbit by dynamical friction to obtain the merger rate (Tremaine 1981). The distribution of companions is described by the **galaxy-galaxy correlation function** $\xi(r)$, defined so that the probability of finding two galaxies in the volumes $\mathrm{d}^3\mathbf{r}_1$ and $\mathrm{d}^3\mathbf{r}_2$ separated by $r \equiv |\mathbf{r}_1 - \mathbf{r}_2|$ is

$$\mathrm{d}^2p = n_0^2[1 + \xi(r)]\,\mathrm{d}^3\mathbf{r}_1\mathrm{d}^3\mathbf{r}_2, \tag{8.136}$$

where $n_0\,\mathrm{d}^3\mathbf{r}$ is the probability of finding one galaxy in the volume $\mathrm{d}^3\mathbf{r}$. Over a wide range of separations and luminosities, the galaxy-galaxy correlation

[13] NGC stands for "New General Catalog", a catalog of galaxies, nebulae, and clusters compiled by Dreyer in 1888, which was a revision and expansion of Herschel's "General Catalog" of 1864. See BM Appendix B.

function can be described by a power law,

$$\xi(r) \simeq \left(\frac{r_0}{r}\right)^\gamma, \tag{8.137}$$

where (Hawkins et al. 2003)

$$r_0 = (7.2 \pm 0.4) h_7^{-1}\,\mathrm{Mpc} \quad ; \quad \gamma = 1.67 \pm 0.03, \tag{8.138}$$

over the range $1 \lesssim \xi \lesssim 10^3$. Thus the number density of companions at distance $r \ll r_0$ from a primary galaxy is

$$n(r) = n_0 \left(\frac{r_0}{r}\right)^\gamma. \tag{8.139}$$

The rate of decay of the orbital radius of a companion due to dynamical friction is given approximately by equation (8.16). Assuming that the primary galaxy and the companion have similar luminosity and velocity dispersion, σ, we have

$$\frac{\mathrm{d}r}{\mathrm{d}t} \simeq -0.4 f \sigma, \tag{8.140}$$

where we have set the Coulomb logarithm $\ln\Lambda \simeq 1$ according to the arguments that follow equation (8.17), and $f \gtrsim 1$ is a correction factor that arises because the relative orbit is likely to be elongated rather than circular, which accelerates the decay. For each primary galaxy, the current of merging companions through radius r is then

$$\dot{N}(r) \simeq \tfrac{1}{2} \times 4\pi r^2 n(r) \left|\frac{\mathrm{d}r}{\mathrm{d}t}\right| \simeq 0.8\pi f n_0 \sigma r_0^\gamma r^{2-\gamma}; \tag{8.141}$$

the factor $\frac{1}{2}$ is needed to avoid counting each galaxy twice, once as a primary galaxy and once as a companion. In a steady state, the current should be independent of r, and this constant number would represent the merger rate; the actual weak dependence on r seen in equation (8.141), $\dot{N} \propto r^{2-\gamma} \propto r^{0.3}$, probably arises because our approximation that both galaxies are isothermal spheres is not very accurate. We shall equate the merger rate $\dot{N}_{\mathrm{merge}}$ to $\dot{N}(r_{\mathrm{min}})$, where $r_{\mathrm{min}} = 20 h_7^{-1}\,\mathrm{kpc}$ is roughly the radius at which the stellar distributions of two luminous galaxies begin to merge.

The derived merger rate depends on the minimum luminosity of the companion galaxies that we consider—clearly, if we count minor mergers, the merger rate will be higher than if we count only major mergers. For the present estimate we shall consider only mergers of companions with luminosity $L > L_\star$, where $L_\star$ is the characteristic Schechter luminosity defined by equation (1.18). Using that equation, the average number density of galaxies more luminous than $L_\star$ is

$$n_0(L > L_\star) = \int_{L_\star}^\infty \mathrm{d}L\, \phi(L) = \phi_\star \int_1^\infty \mathrm{d}x\, x^\alpha \mathrm{e}^{-x}. \tag{8.142}$$

For the parameters given after equation (1.18), we find $n_0(L > L_\star) = 0.21\phi_\star \simeq 1.0\times10^{-3}h_7^3\,\mathrm{Mpc}^{-3}$. According to the Faber–Jackson law (1.21), the dispersion of an $L_\star$ galaxy is $\sigma_\star \simeq 200\,\mathrm{km\,s^{-1}}$. Thus

$$\begin{aligned}\dot N_{\rm merge}(L>L_\star) &\simeq 0.8\pi f n_0(L>L_\star)\sigma_\star r_0^\gamma r_{\rm min}^{2-\gamma}\\ &\approx 0.008h_7\frac{f}{2}\,\mathrm{Gyr}^{-1}.\end{aligned}\tag{8.143}$$

The approximate agreement of this estimate with the rate $0.005\,\mathrm{Gyr}^{-1}$ obtained by Toomre (1977b) provides encouraging evidence that our understanding of the merging process is sound. More recent determinations of the merger rate (Conselice 2006) are also roughly consistent with these crude estimates.

Problems

8.1 [1] Two identical galaxies are initially at rest, at a large distance from one another. They are spherical, composed solely of identical stars, and their light distributions obey the Sérsic law (1.17) with Sérsic index m and effective radius $R_{\rm e}$. The galaxies fall together and merge. If the merger product also satisfies the Sérsic law with the same index, what is its effective radius?

8.2 [1] The derivation of the dynamical friction formula (8.1) assumes that the subject system is a point mass, but in many cases of interest the subject system is an extended body, such as a star cluster or satellite galaxy, characterized by a median radius $r_{\rm h}$. If the point of closest approach of the field star to the center of the subject body is $\lesssim r_{\rm h}$ then the deflection of the field-star orbit, and its contribution to the drag force, will be smaller than if the subject body were a point of the same total mass.

(a) Argue that the total drag force is largely unaffected by the non-zero size of the subject body if $r_{\rm h} \lesssim b_{90}$, where b_{90} is given by equation (3.51).

(b) If $r_{\rm h} \gtrsim b_{90}$, argue that encounters with impact parameter $\lesssim r_{\rm h}$ make a negligible contribution to the total drag force. Using the first of equations (L.11), argue that in this case the argument of the Coulomb logarithm is given by $\Lambda \simeq b_{\rm max}/r_{\rm h}$.

(c) Combine these conclusions to argue that the correct value of the argument of the Coulomb logarithm for a subject body of median radius $r_{\rm h}$ is approximately

$$\Lambda = \frac{b_{\rm max}}{\max(r_{\rm h}, GM/v_{\rm typ}^2)},\tag{8.144}$$

and that the fractional error in $\ln\Lambda$ that arises from using this expression is of order $(\ln\Lambda)^{-1}$.

8.3 [3] In the core of a certain flattened elliptical galaxy, the mean stellar velocity vanishes and the velocity distribution is Gaussian, with dispersion σ_z parallel to the galaxy's symmetry axis $\hat{\mathbf{e}}_z$, and dispersion $\sigma_\perp = \sigma_z/\sqrt{1-e^2} > \sigma_z$ in directions orthogonal to $\hat{\mathbf{e}}_z$. A massive body moves through the core at velocity $\mathbf{v} = v_z\hat{\mathbf{e}}_z + v_\perp\hat{\mathbf{e}}_\perp$, where $\hat{\mathbf{e}}_\perp\cdot\hat{\mathbf{e}}_z = 0$. Show that the frictional drag on the body may be written $\mathbf{F} = -\gamma_z v_z\hat{\mathbf{e}}_z - \gamma_\perp v_\perp\hat{\mathbf{e}}_\perp$, where

$$1 < \frac{\gamma_z}{\gamma_\perp} = \frac{I(1,\frac32)}{I(2,\frac12)}\ \ ;\ \ I(\mu,\nu)\equiv\int_1^\infty\frac{\mathrm{d}\lambda}{\lambda^\mu(\lambda-e^2)^\nu}\exp\Big[-\tfrac12\sigma_\perp^{-2}\Big(\frac{v_\perp^2}{\lambda}+\frac{v_z^2}{\lambda-e^2}\Big)\Big].\tag{8.145}$$

Hint: use the analogy between the Rosenbluth potential $h(\mathbf{v})$ and the gravitational potential (see discussion following eq. L.19) and equation (2.125).

8.4 [3] Chandrasekhar's dynamical friction formula can be derived using the linear response theory developed in §5.2.4 (Marochnik 1967; Kalnajs 1972b).

(a) Consider a point mass M traveling on the straight-line trajectory $\mathbf{x}_M(t) = \mathbf{v}_M t$ through a uniform stellar system. Show that the spatial Fourier transform (eq. 5.26) of the density response is given by

$$\overline{\rho}_{s1}(\mathbf{k},t) = M\int \mathrm{d}t'\,\overline{R}(\mathbf{k},t-t')\,\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{v}_M t'}, \tag{8.146}$$

where $\overline{R}(\mathbf{k},\tau)$ is the response function (eq. 5.27).

(b) As we showed in §5.2.4, an infinite homogeneous stellar system is unstable, so to avoid an infinite response we must suppress the self-gravity of the system when evaluating equation (8.146). The justification for this neglect is that the instability arises on scales comparable to the Jeans length, while the dominant contribution to dynamical friction comes from encounters at much smaller distances (page 576), for which the effects of self-gravity are small. To remove self-gravity, we simply replace the response function R in equation (8.146) by the polarization function P, which measures the response to a given total potential rather than a given external potential. With this substitution, use equation (5.55) to show that if the stellar system has a Maxwellian DF, the Fourier transform of the density response is

$$\overline{\rho}_{s1}(\mathbf{k},t) = 4\pi GM\rho\,\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{v}_M t}\int_0^\infty \mathrm{d}\tau\,\tau \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{v}_M\tau-(k\sigma\tau)^2/2}, \tag{8.147}$$

where ρ and σ are the density and velocity dispersion of the host system.

(c) Take the inverse Fourier transform of $\overline{\rho}_{s1}$, and evaluate the resulting integrals to show that the density response is

$$\rho_{s1}(\mathbf{x},t) = \frac{GM\rho}{\sigma^2 r}\exp\left(-\frac{v_M^2\sin^2\theta}{2\sigma^2}\right)\left[1-\mathrm{erf}\left(\frac{v_M\cos\theta}{\sqrt{2}\sigma}\right)\right], \tag{8.148}$$

where erf denotes the error function (Appendix C.3), $\mathbf{r} = \mathbf{x}-\mathbf{x}_M$, $r = |\mathbf{r}|$, and θ is the angle between $\mathbf{v}_M$ and $\mathbf{r}$. Hint: carry out the integral over $\mathbf{k}$ first, using polar coordinates in $\mathbf{k}$-space with the polar axis along the vector $\mathbf{r}+\mathbf{v}_M\tau$; then evaluate the integral over τ after transforming to the variable $u = 1/\tau$.

(d) Show that the gravitational force exerted on M by this density distribution is

$$\mathbf{F} = 2\pi\frac{G^2M^2\rho}{\sigma^2}\frac{\mathbf{v}_M}{v_M}\int\frac{\mathrm{d}r}{r}\int_{-1}^1 \mathrm{d}\mu\,\mu\exp\left[-\frac{v_M^2(1-\mu^2)}{2\sigma^2}\right]\left[1-\mathrm{erf}\left(\frac{v_M\mu}{\sqrt{2}\sigma}\right)\right], \tag{8.149}$$

where $\mu = \cos\theta$.

(e) The upper limit to the integral over radius should be of order the size R of the host system, while the lower limit should be roughly the 90° deflection radius $b_{90} \approx GM/\sigma^2$ (eq. 3.51), since interior to this radius the perturbations to the orbits of passing stars are so large that linear response theory is invalid. With these limits, show that evaluation of the integrals in equation (8.149) yields the standard dynamical friction formula (8.7), with $\Lambda = R/b_{90}$.

8.5 [1] At some initial time the stellar streaming velocity $\overline{\mathbf{v}}(\mathbf{x})$ within an axisymmetric galaxy of density $\rho(R,z)$ constitutes circular rotation at angular frequency $\omega(R,z)$. The galaxy is then perturbed by the high-speed passage of a massive system. Show that within the impulse approximation the instantaneous change $\Delta\mathbf{v}$ in $\mathbf{v}$ that is produced by the encounter satisfies

$$\int \mathrm{d}^3\mathbf{x}\,\rho(\overline{\mathbf{v}}\cdot\Delta\mathbf{v}) = 0. \tag{8.150}$$

Hint: write $\overline{\mathbf{v}} = \omega R\,\hat{\mathbf{e}}_\phi$ and exploit the fact that $\Delta\mathbf{v}$ can be derived from a potential.

8.6 [2] Reproduce Figure 8.4.

8.7 [2] Consider a high-speed head-on encounter at relative velocity V. Assume that the perturber is spherical, with gravitational potential $\Phi(r)$, and let (R, z) be cylindrical coordinates such that the z axis coincides with the perturber's trajectory (i.e., the trajectory is $R = 0$, $z = Vt$).

(a) Show that the only non-zero component of the impulse to a star at (R, z) is

$$\Delta v_R = -\frac{2R}{V}\int_R^\infty \frac{\mathrm{d}r}{\sqrt{r^2 - R^2}}\frac{\mathrm{d}\Phi}{\mathrm{d}r}. \tag{8.151}$$

(b) If the perturber is a Plummer model, $\Phi = -GM/\sqrt{r^2 + b^2}$, of mass M and scale length b (§2.2.2c), show that the impulse is

$$\Delta v_R = -\frac{2GMR}{V(R^2 + b^2)}. \tag{8.152}$$

(c) If the perturber and the perturbed system are identical Plummer models, show that the energy per unit mass gained by each system in the encounter is

$$\Delta E = \frac{G^2M^2}{3V^2b^2}. \tag{8.153}$$

8.8 [1] Show that the probability $P(V)\,\mathrm{d}V$ that two stars drawn from Maxwellian distributions with one-dimensional dispersions σ_1 and σ_2 have relative speed in the interval $(V, V + \mathrm{d}V)$ is

$$P(V)\,\mathrm{d}V = (2\pi\sigma^2)^{-3/2}\exp\Big(-\frac{V^2}{2\sigma^2}\Big)V^2\mathrm{d}V, \quad \sigma^2 = \sigma_1^2 + \sigma_2^2. \tag{8.154}$$

In words, the relative speed distribution is Maxwellian, with squared dispersion equal to the sum of the squared dispersions of the two populations.

8.9 [1] Is there more angular momentum in the orbit of the Magellanic Clouds around our Galaxy or in the spin of the disk of our Galaxy?

8.10 [3] Show that the Lagrange points L_4 and L_5 in the restricted three-body problem

(a) each form an equilateral triangle with the two masses M and m.

(b) are stable if and only if the mass ratio $m/(M + m) < \frac{1}{2} - \sqrt{\frac{23}{108}} = 0.03852$ (Szebehely 1967). Hint: start with equation (3.116); do not use equations (3.127) since these assume that $\Phi_{xy} = 0$, which is not true in this case.

(c) Are the Lagrange points L_1, L_2, L_3 stable?

8.11 [3] A frictionless railroad crosses a valley that separates two flat plateaus of equal height. At $t = 0$, two cars, each of mass m, are sent off with the same speed v and separation d from the horizontal stretch of track on one side of the valley. Show that when the cars emerge onto the horizontal stretch of track on the other side of the valley, they have zero relative velocity and their separation is unchanged.

Discuss the relation between this system and disk shocking of globular clusters; in particular, why does passage through the disk heat the cluster but leave the relative velocity of the cars unchanged? Hint: consider adding a spring of rest length d and stiffness ω^2/m between the two cars.

8.12 [2] A satellite system of mass m is in a circular orbit around a point-mass host $M \gg m$. Let (x, y, z) be Cartesian coordinates with $\hat{\mathbf{e}}_x$ pointing along the line joining the two masses and $\hat{\mathbf{e}}_z$ normal to the orbital plane. The distance of the tidal surface from m along the x axis is r_J (eq. 8.91). Show that the distance of this surface from m along the y- and z-axes is $\frac{2}{3}r_\mathrm{J}$ and $(3^{2/3} - 3^{1/3})r_\mathrm{J}$, respectively. Thus, the tidal surface is not spherical.

8.13 [2] In the distant-tide approximation, the tidal field around a freely falling satellite of a host galaxy can be written in the form $-\boldsymbol{\nabla}\Phi_{\rm t} = -\sum_{i,j=1}^{3} \hat{\mathbf{e}}_i \Phi_{ij} x_j$, where $\{x_j\}$ are non-rotating Cartesian coordinates centered on the body (see §8.2.1). The tidal field is said to be **compressive** along axis i if $\hat{\mathbf{e}}_i \cdot \boldsymbol{\nabla}\Phi_{\rm t} > 0$, that is, if the tidal force points towards the center of the satellite.

(a) If the host galaxy is spherical with density $\rho_{\rm h}(R)$ at a distance R from its center, prove that the tidal force is compressive along all three axes if and only if $\rho_{\rm h} > \frac{2}{3}\overline{\rho}_{\rm h}$, where $\overline{\rho}_{\rm h}$ is the mean density of the host interior to R (cf. eq. 8.92).

(b) If the density of the host is $\rho(R) \propto R^{-\gamma}$, prove that the tidal force is compressive in all directions if and only if $\gamma < 1$. In a host with this property tidal disruption cannot occur, no matter how small the mass of the satellite may be. How is this result consistent with the discussion of tidal disruption in §8.3?

8.14 [3] This problem investigates how orbits that lie far beyond the Jacobi radius can remain bound to a satellite. We consider a satellite on a circular orbit, using Hill's approximation (§8.3.2) and restrict our attention to the orbital plane of the satellite, $z = 0$. Since the orbits in question are much larger than the Jacobi radius, the gravitational field of the satellite is weak. Thus we may assume that the orbit is described approximately by the solution (8.99) over timescales of order the epicycle period $T_r = 2\pi/\kappa_0$, with constants of motion $x_{\rm g}$, y_{g0}, X, Y, and ψ that change slowly due to perturbations from the satellite.

(a) Show that the guiding-center radius $x_{\rm g}$ changes at a rate

$$\dot{x}_{\rm g} = \frac{1}{2B_0}\frac{\partial \Phi_{\rm s}}{\partial y}, \tag{8.155}$$

where $\Phi_{\rm s}(\mathbf{x}) = -Gm/(x^2+y^2)^{1/2}$ is the potential from the satellite. Hint: use equations (8.97), (8.101), and (8.102).

(b) If the perturbations from the satellite are weak, and $|x_{\rm g}| \ll |y_{\rm g}|$ (assumptions we will justify below) then the term $\partial\Phi_{\rm s}/\partial y$ in equation (8.155) can be replaced by its average over an epicycle period at fixed values of the constants of motion; that is

$$\frac{\partial \Phi_{\rm s}}{\partial y} \Rightarrow \left\langle \frac{\partial \Phi_{\rm s}}{\partial y} \right\rangle \equiv \frac{Gm}{2\pi}\int_0^{2\pi} {\rm d}\tau \frac{y_{\rm g} - Y\sin\tau}{[(x_{\rm g}+X\cos\tau)^2 + (y_{\rm g}-Y\sin\tau)^2]^{3/2}}. \tag{8.156}$$

In the limit where $|x_{\rm g}| \ll |y_{\rm g}| \ll X$, that is, where the distance of the guiding center from the satellite is much less than the epicycle size, show that

$$\left\langle \frac{\partial \Phi_{\rm s}}{\partial y} \right\rangle = -\frac{Gmy_{\rm g}}{X^3} W\left(\frac{\Omega_0}{\kappa_0}\right), \quad \text{where} \quad W(u) \equiv \frac{2}{\pi}\int_0^{\pi/2} {\rm d}\tau \frac{8u^2\sin^2\tau - \cos^2\tau}{[\cos^2\tau + 4u^2\sin^2\tau]^{5/2}}. \tag{8.157}$$

Hint: expand equation (8.156) in a Taylor series, and use equation (8.100). The function $W(u)$ varies from 0.10032 for $u = 1$ (Keplerian orbits) to 0.22662 for $u = 2^{-1/2}$ (flat circular-speed curve) to 0.5 for $u = \frac{1}{2}$ (harmonic oscillator).

(c) Differentiating the equation for $y_{\rm g}(t)$ in (8.99) and using the assumption that the time derivatives of the constants of motion are small yields $\dot{y}_{\rm g} = -2A_0 x_{\rm g}$. Using this result and equation (8.155) show that the equation of motion for the guiding center is

$$\ddot{y}_{\rm g} = -\frac{A_0}{B_0}\left\langle \frac{\partial \Phi_{\rm s}}{\partial y} \right\rangle. \tag{8.158}$$

Interpret this result in terms of the "effective mass" introduced in Box 3.3.

(d) Show that the motion of the guiding center is given by

$$x_{\rm g}(t) = X_{\rm g}\cos(\omega t + \alpha) \quad ; \quad y_{\rm g}(t) = Y_{\rm g}\sin(\omega t + \alpha), \tag{8.159}$$

where $X_{\rm g}$ and α are arbitrary and

$$\frac{X_{\rm g}}{Y_{\rm g}} = -\frac{\omega}{2A_0} \quad ; \quad \omega^2 = \frac{GmA_0}{-B_0X^3} W\left(\frac{\Omega_0}{\kappa_0}\right) = \frac{4\Omega_0 A_0^2}{-B_0}\left(\frac{r_{\rm J}}{X}\right)^3 W\left(\frac{\Omega_0}{\kappa_0}\right); \tag{8.160}$$

the final form has been derived with the use of equation (8.106). In most galactic potentials $A_0 > 0$ and $B_0 < 0$, so $\omega^2 > 0$ and ω is real. Thus the guiding center oscillates around the satellite; if the orbit lies far outside the tidal radius ($X \gg r_J$) then (i) the oscillation is slow in the sense that $\omega \ll \Omega_0$, and (ii) the excursions in y_g are much larger than the excursions in x_g, consistent with the assumptions we made in deriving this result. The approximations we have used also require that $Y_g \ll Y$, that is, the amplitude of the guiding-center oscillations must be smaller than the epicycle amplitude.

(e) In Problem 5.1, we showed that a solid ring orbiting a planet is unstable. This calculation neglected tidal forces. Would a solid ring that is much larger than the Jacobi radius be stable?

8.15 [1] (a) Derive the energy integral (8.104) for the sheared sheet in two ways, first by multiplying the equations of motion (8.97) by $\dot{x}$, $\dot{y}$, $\dot{z}$ respectively, adding, and integrating; second by finding the Lagrangian and Hamiltonian that yield the equations of motion.

(b) Assume that we impose periodic boundary conditions on the sheared sheet, by identifying $y + 2\pi R$ with y. Find the angle-action variables for the case $\Phi_s = 0$, and relate these to the angle-action variables in the epicycle approximation (§3.5.3b).

8.16 [1] A spherical host galaxy contains two small satellites having masses m_1 and m_2. The satellites travel on nearly circular orbits with nearly the same orbital radius and plane.

(a) Argue that their interactions can be described using Hill's approximation (8.97) in the form

$$\ddot{x}_1 - 2\Omega_0\dot{y}_1 - 4\Omega_0 A_0 x_1 = -\frac{\partial\Phi_{12}}{\partial x_1}\ ;\ \ddot{y}_1 + 2\Omega_0\dot{x}_1 = -\frac{\partial\Phi_{12}}{\partial y_1}\ ;\ \ddot{z}_1 + \Omega_0^2 z_1 = -\frac{\partial\Phi_{12}}{\partial z_1}, \quad (8.161)$$

where $\Phi_{12} = -Gm_2/|\mathbf{x}_1 - \mathbf{x}_2|$. Here $\mathbf{x}_i \equiv (x_i, y_i, z_i)$ is the position of satellite i, $i = 1, 2$. The equation of motion for satellite 2 is obtained by interchanging the indices 1 and 2.

(b) In this approximation, what is the trajectory of the center of mass of the two satellites, $\mathbf{x}_{cm} \equiv (m_1\mathbf{x}_1 + m_2\mathbf{x}_2)/(m_1 + m_2)$?

(c) Show that determining the motion of the two satellites can be reduced to solving the equation of motion for a single particle with position $\mathbf{x} \equiv \mathbf{x}_2 - \mathbf{x}_1$.

8.17 [2] This problem analyzes the sheared sheet (§8.3.2) as a model for the kinematics of the solar neighborhood or other stellar disks. For simplicity, we restrict ourselves to a two-dimensional disk, ignoring motion in the z-coordinate, although the results are easily generalized to three-dimensional disks.

(a) Show that in the absence of local mass concentrations (that is, if the satellite potential $\Phi_s = 0$) the equations of motion (8.97) are invariant under the transformation

$$x \to x + \Delta x \quad ; \quad y \to y - 2A_0\Delta x t. \quad (8.162)$$

Describe the physical meaning of this symmetry.

(b) According to the Jeans theorem, the equilibrium DF $f(x, y, \dot{x}, \dot{y})$ can depend only on the integrals of motion $E_\parallel$ and L (eq. 8.101). Show that the only combination of these integrals that is invariant under the transformation (8.162) is the epicycle energy defined in equation (8.103). Thus argue that if the disk is smooth on small scales, the DF must have the form $f(E_x)$.

(c) For a DF of this form, show that the surface density is independent of position, the mean radial velocity vanishes, the mean azimuthal or y-velocity is $-2A_0x$, and the ratio of the dispersions in the azimuthal and radial directions is

$$\frac{\sigma_y^2}{\sigma_x^2} = \frac{\int \mathrm{d}\dot{x}\mathrm{d}\dot{y}\, f(E_x)(\dot{y} + 2A_0x)^2}{\int \mathrm{d}\dot{x}\mathrm{d}\dot{y}\, f(E_x)\dot{x}^2} = \frac{\kappa_0^2}{4\Omega_0^2}, \quad (8.163)$$

a result already derived by different methods in equation (3.100).

(d) Show that if $f(E_x) \propto \exp(-E_x^2/\sigma_0^2)$ then

$$f(x,y,\dot{x},\dot{y}) = \frac{\Sigma}{\pi\sigma_0^2}\frac{\Omega_0}{\kappa_0}\exp\left[-\frac{\dot{x}^2}{2\sigma_0^2} - \frac{2\Omega_0^2(\dot{y}+2A_0x)^2}{\kappa_0^2\sigma_0^2}\right]. \tag{8.164}$$

Show that this is the analog of the Schwarzschild DF introduced in §4.4.3.

(e) Does this DF exhibit asymmetric drift (§4.8.2a)?

8.18 [2] Assume that the Sun travels in a circular orbit in the Galactic plane. Let (x,y,z) be rotating Cartesian coordinates centered on the Sun, with $\hat{\mathbf{e}}_x$ pointing away from the Galactic center and $\hat{\mathbf{e}}_z$ pointing to the north Galactic pole.

(a) Show that the zero-velocity surfaces in the combined gravitational field of the Sun and the Galaxy are given by

$$2A(B-A)x^2 + (A^2 - B^2 + 2\pi G\rho_0)z^2 - \frac{G\mathcal{M}_\odot}{r} = \text{constant}, \tag{8.165}$$

where ρ_0 is the density in the solar neighborhood, A and B are Oort's constants, and $r^2 = x^2+y^2+z^2$. Hint: see Problem 3.18.

(b) Let $x_{\rm J}$, $y_{\rm J}$, $z_{\rm J}$ be the intersections of the Sun's tidal surface with the coordinate axes. Evaluate these quantities in parsecs, using the parameters in Tables 1.1 and 1.2.

8.19 [2] Reproduce Figure 8.8.

8.20 [3] The goal of this problem is to determine the epicycle amplitude induced in a star as it passes a molecular cloud, in the shear-dominated regime. We use the equations of motion (8.97) and neglect motion perpendicular to the x–y plane. We assume that the cloud is at the origin and that the star is initially on a circular orbit with impact parameter b, so $\mathbf{x}(t) = (b, -2A_0bt)$. If the cloud potential is $\phi(\mathbf{x}) = -Gm/(x^2+y^2)^{1/2}$ and its mass is sufficiently small that the right sides of the equations of motion can be evaluated along the unperturbed stellar orbit, show that after the encounter the epicycle amplitude is (Julian & Toomre 1966)

$$X = \frac{Gm\Omega_0}{\kappa_0 A_0^2 b^2}\left[K_0\left(\frac{\kappa_0}{2A_0}\right) + \frac{\kappa_0}{2\Omega_0}K_1\left(\frac{\kappa_0}{2A_0}\right)\right], \tag{8.166}$$

where K_ν is a modified Bessel function (Appendix C.7). Thus, derive the correction factor f in equation (8.121b).

8.21 [2] In this problem we estimate the rate of growth of epicycle energy in the dispersion-dominated regime. Consider a star traveling on a nearly circular orbit in the equatorial plane of a razor-thin galaxy. At time zero, the star is instantaneously deflected by the gravitational field from a nearby molecular cloud that is itself on a perfectly circular orbit. The star is traveling at speed v with respect to the cloud, and the encounter deflects it through an angle η onto a new, nearly circular orbit within the galactic plane. Show that the deflection changes the star's epicycle energy by an amount

$$\Delta E_x = E_x(\gamma^2-1)\big[\sin^2\eta(\sin^2\alpha - \gamma^{-2}\cos^2\alpha) - \frac{1}{2\gamma}\sin 2\eta\sin 2\alpha\big], \tag{8.167}$$

where $\gamma = 2\Omega/\kappa$ and α is the epicycle phase (see eqs. 3.91 and 3.93, or 8.99). At the radius of the star's orbit, there are n clouds per unit area, each having mass m. The mass distribution in the clouds can be represented by a Plummer model with scale length b, which is much smaller than the star's epicycle radius. Using the impulse approximation, assuming that the relative velocity is dominated by the velocity dispersion of the stars, and assuming that the deflection angle η is small, show that the expectation value of the rate of change of epicycle energy is

$$\dot{E}_x = \frac{\sqrt{2}G^2m^2n}{b\sqrt{E_x}}(\gamma^2-1)\int_0^{\pi/2}\mathrm{d}\alpha\,\frac{\sin^2\alpha - \gamma^{-2}\cos^2\alpha}{(\sin^2\alpha + \gamma^{-2}\cos^2\alpha)^{3/2}}. \tag{8.168}$$

Verify that $\dot{E}_x > 0$ for $\gamma > 1$. What happens to stars in a galaxy with $\gamma < 1$?