Mon. Not. R. astr. Soc. (1988), **230**, 597–627

The diffusion of stars through phase space

James Binney *Department of Theoretical Physics, Keble Road, Oxford OX1 3NP, and Institute for Advanced Study, Princeton, NJ 08540, USA*

Cedric Lacey *Princeton University Observatory, Peyton Hall, Princeton, NJ 08544, USA, and Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA*

Accepted 1987 August 14. Received 1987 August 14; in original form 1987 July 14

Summary. Deviations of the potentials of stellar systems from integrability cause stars to diffuse through three-dimensional orbit space. The Fokker–Planck equation that describes this diffusion takes a particularly simple form when actions are used as orbit-space coordinates. The rate of diffusion is governed by a vector $\overline{\boldsymbol{\Delta}}$ and a tensor $\overline{\boldsymbol{\Delta}^2}$, which according to the circumstances of a particular problem should be calculated either from kinetic theory or from Hamiltonian perturbation theory. In many astrophysically interesting circumstances $\overline{\boldsymbol{\Delta}}$ is related to the divergence of the more readily calculated tensor $\overline{\boldsymbol{\Delta}^2}$. In addition to being computationally handy, this relationship ensures that the orbital diffusion described by the Fokker–Planck equation causes the system's entropy, derived from any H-function, to increase whenever the system is interacting with a hotter system of scatterers. An investigation of the heating of stellar discs in the light of these general results yields the following conclusions: (i) a population of stars with an initially Maxwellian peculiar-velocity distribution will remain Maxwellian as it diffuses through orbit space only if $\overline{\boldsymbol{\Delta}^2}$ is proportional to epicycle energy and the population's velocity dispersion grows as $\sqrt{t}$; (ii) the self-similar distribution functions that are the end-points in two and three dimensions of the star-cloud scattering process proposed by Spitzer and Schwarzschild, predict neither Maxwellian velocity distributions nor $\sigma \propto \sqrt{t}$; (iii) scattering by ephemeral spiral waves can account for the observed kinematics of the solar neighbourhood only if the waves have wavelengths in excess of 9 kpc and constantly drifting pattern speeds. However, even such frequency-modulated, global spirals cannot maintain $\sigma \propto \sqrt{t}$ above $\sigma_R > 40\,\mathrm{km\,s^{-1}}$, and they can account for the increase in the vertical dispersion with time only with the assistance of clouds. Only scatterers that are not confined to the Galactic disc are capable of simultaneously increasing the vertical and radial dispersions as $\sqrt{t}$ to dispersions in excess of $40\,\mathrm{km\,s^{-1}}$.

1 Introduction

The usual first step in the analysis of a large stellar system is the construction of a self-consistent mean-field model of the system. In this model stars move through a smooth potential on orbits that generally have three effective isolating integrals. In the system itself, by contrast, the potential is not absolutely smooth, and over long times the integrals of stars drift from their initial values. Hence, if it is desired to model the long-term evolution of the system, the mean-field model must be continually updated to take into account the effects of irregularities in the system's potential.

Studies of the evolution of spherical systems in response to irregular star–star interactions have generated an extensive literature (e.g. Goodman & Hut 1985). But comparatively little attention has been given to the study of the evolution of non-spherical systems such as flattened clusters (Goodman 1983) and galactic discs (Wielen & Fuchs 1986), even though such studies address questions of considerable observational interest. Two techniques have been employed in studies of spherical systems: (i) Monte-Carlo simulation, and (ii) orbit-averaged Fokker–Planck equations. In this paper we establish a framework for the study through orbit-averaged Fokker–Planck equations of the evolution of non-spherical systems, and illustrate the power of this framework by studying the secular heating of stellar discs such as that of our Galaxy.

At birth, disc stars have random velocities which are typical of the molecular complexes in which they are formed: less than $10\,\mathrm{km\,s^{-1}}$ (e.g. Wielen 1977). The fact that all components of the velocity dispersion appear to increase along the main sequence towards later spectral types, has long been regarded as compelling evidence that some sort of stellar acceleration mechanism operates in the disc. A mechanism which arises naturally, indeed unavoidably, from the inhomogeneous nature of galactic discs is stochastic scattering of stars by embedded, massive perturbers. Both molecular cloud complexes (Spitzer & Schwarzschild 1951, 1953) and transient spiral features (Barbanis & Woltjer 1967; Carlberg & Sellwood 1985) are likely accelerators of Population I stars.

The suggestion that inhomogeneities in the disc accelerate stars goes back to Spitzer & Schwarzschild (1951). The acceleration process discussed in that paper was entirely analogous to the thermalization of a population of energetic particles embedded in a cold plasma, with the massive molecular clouds playing the role of the energetic particles. Remarkable for its prescience, this work suggested the existence of gaseous agglomerations of the interstellar medium with typical masses of $5\times10^5\,M_\odot$ some quarter of a century before molecular cloud complexes were appreciated to be an important component of the galactic disc. Spitzer & Schwarzschild (1953) subsequently showed that in a differentially rotating disc, stars are accelerated even by clouds that have no random motion. They used the epicyclic approximation to show that star-cloud scattering in a razor-thin disc would cause the radial velocity dispersion to increase with time t as $t^{1/3}$. More recently, Wielen (1977) and Lacey (1984) have extended this approach to include the vertical structure of the disc, and numerical integrations of star–cloud encounters and N-body simulations (Icke 1982; Villumsen 1985) have been used to study the evolution of the stellar distribution function.

A very different point of view was introduced by Barbanis & Woltjer (1967), who investigated the possibility that discs are heated by spiral arms. They showed that only transient spiral patterns heat the disc, and recently this idea has been taken up again by Carlberg & Sellwood (1985), who interpret the rapid heating of numerically simulated discs such as those of Sellwood & Carlberg (1984), as effected by transient spiral features.

In reality, the distinction between the molecular cloud scatterers of Spitzer and Schwarzschild, and the spiral scatterers of Barbanis and Woltjer is to some extent an artificial one. Indeed, the spirals that heat most effectively are the most ephemeral arms and, as Julian & Toomre (1966) have shown, in a cool stellar disc just such arms grow naturally around any massive body such as a molecular cloud. Furthermore, observations of the molecular gas contents of our and external galaxies have shown that molecular clouds are not randomly distributed through the disc, but tend to be organized into spiral arms (Cohen *et al.* 1985). Thus heating by molecular clouds is inevitably augmented by spiral heating, and it is natural to seek to formulate the disc-heating problem in such a way that both heating processes can be studied simultaneously. This is a primary goal of this paper. An additional stimulus for this work is provided by the large stellar velocity samples that are now becoming available. Such samples enable us to determine the *shape* of the stellar velocity distribution in addition to its dispersion. To interpret such distributions theoretically, we need to follow the evolution of the entire galactic distribution function, and not just estimate its second velocity moments, as has been customary till now.

This paper is organized as follows. In Section 2 we derive an orbit-averaged Fokker–Planck equation that may be used to follow the secular evolution of any stellar system in which most orbits are regular (i.e. admit at least as many isolating integrals as the dimensionality of the orbits). This equation involves both a first-order diffusion vector $\overline{\mathbf{\Delta}}$ and a second-order diffusion tensor $\overline{\mathbf{\Delta}^2}$. In Section 3 we show how the tensor $\overline{\mathbf{\Delta}^2}$ is related to the autocorrelation function of an externally imposed stochastic perturbing potential, and in Appendix A we show that the vector $\overline{\mathbf{\Delta}}$ generated by such a potential is simply the action-space divergence of $\overline{\mathbf{\Delta}^2}$. Appendix B describes circumstances in which slightly more complex relations are obtained between $\overline{\mathbf{\Delta}}$ and $\overline{\mathbf{\Delta}^2}$, and in Section 3.2 we discuss the implications of these relations for the rate of entropy generation within the system. In Section 4 we use the epicycle approximation to show that the Schwarzschild distribution of solar-neighbourhood stellar velocities derives from a simple self-similar solution of the relevant Fokker–Planck equation, provided that $\overline{\mathbf{\Delta}^2}$ is proportional to the epicycle energies, and that in this solution the velocity dispersions increase as the square root of time. In Section 5 we use the approximations introduced by Spitzer & Schwarzschild (1953) to obtain the single independent diffusion coefficient involved in their planar diffusion problem, and display the self-similar form to which the solution of this problem tends at late times: in this solution the velocity dispersions grow as $t^{1/3}$ and the velocity distribution is significantly sub-Gaussian. In Appendix C we use the Spitzer–Schwarzschild approximations to derive the diffusion coefficients of the more realistic three-dimensional problem. From the general structure of these coefficients we show that in the three-dimensional case the vertical and radial velocity dispersions tend rapidly to the ratio $\sigma_z/\sigma_R \simeq 0.79$ and thereafter grow as $t^{1/4}$. We give approximate expressions for the self-similar solution to which the distribution function tends at late times.

The approximations introduced by Spitzer and Schwarzschild are only marginally valid for molecular clouds, and break down completely for larger scatterers such as transient spiral arms. Therefore, in Section 6 and Appendix D we employ angle-action coordinates to calculate the diffusion coefficients generated in a razor-thin planar disc by (i) a single transient spiral arm, and (ii) a population of compact clouds such as those discussed, less accurately, by Spitzer and Schwarzschild. These examples lead in Section 7 to the conclusion that neither molecular clouds nor transient spiral arms are on their own capable of accounting for the kinematics of the solar neighbourhood. However a combination of global frequency-modulated spiral structure and massive molecular clouds may be able to account for the structure of the solar neighbourhood provided (i) the radial velocity dispersion does not continue to increase as $\sqrt{t}$ above $\sigma_R = 40$ km s^{-1}, and (ii) the heating of the solar neighbourhood has been episodic. Alternatively, a significant contribution to the heat of the Galactic disc may come from scattering by objects such as halo black holes, or dwarf satellites of our Galaxy, that are on either highly inclined or highly eccentric orbits.

2 Orbit-averaged Fokker–Planck equation

Orbits in smooth galactic potentials generally admit three effective isolating integrals. One may show (e.g. Arnold 1978) that such orbits are closely confined to three-dimensional surfaces in six-dimensional phase space, namely the orbital tori. These tori may be labelled by any three independent isolating integrals, but the natural labels to use in any problem in which the potential changes slowly are some three actions J_i, since these are adiabatic invariants. Deviations of the potential of a real galaxy from perfect integrability and time-independence cause stars to move in phase space from one torus of constant $\mathbf{J}$ to another. However, if the deviations of the potential from integrability are small, the path traced in phase space by any star will keep close to some torus of constant $\mathbf{J}$ for at least several orbital times. On a longer time-scale, the effects of the potential's irregularities on the motion of the star are likely to accumulate, and the particular torus around which the trajectory snakes, will change. Thus we may picture the phase point of each star in a real galaxy as drifting slowly from one torus of constant $\mathbf{J}$ to another, but thoroughly exploring the current torus on a much shorter time-scale.

One may show (e.g. Lynden-Bell 1962 or Binney & Tremaine 1987, section 4.A) that the phase-space density of stars in a perfectly relaxed encounterless galaxy may be assumed to be a function $f(\mathbf{J})$ of the actions only. (We shall refer to this result as the *strengthened Jeans theorem.*) Clearly, if regularities in the potential cause stars to move from one torus to another, f will change. However, if the drift of stars across tori of constant $\mathbf{J}$ is slow compared to the time required for a star to explore its current torus fairly thoroughly, the phase-space density of stars f will remain at any given time a function of the actions only, though a function that changes slowly in time. Thus $f=f(\mathbf{J}, t)$.

We imagine that the steady drift of each star in phase space takes place in a series of discrete steps in which the star moves from actions $\mathbf{J}$ to actions $\mathbf{J}+\mathbf{\Delta}$, where $\mathbf{\Delta}$ is a small vector. Let $\delta P(\mathbf{J}, \mathbf{\Delta})$ be the probability that a star with actions $\mathbf{J}$ changes these actions by $\mathbf{\Delta}$ in the time interval δt. Then summing over all values of $\mathbf{\Delta}$ we have that the total probability in the given time interval that the star is scattered from the region of phase space associated with actions $\mathbf{J}$ is

$$\int \delta P(\mathbf{J}, \mathbf{\Delta})\, d^3\mathbf{\Delta}. \tag{2.1}$$

Since the volume of phase space associated with actions in the element $d^3\mathbf{J}$ is $(2\pi)^3 d^3\mathbf{J}$, the number of stars at time t in $d^3\mathbf{J}$ is

$$(2\pi)^3 d^3\mathbf{J} f(\mathbf{J}, t) \tag{2.2}$$

and the number that in time δt leave this volume is

$$(2\pi)^3 d^3\mathbf{J} f(\mathbf{J}, t) \int \delta P(\mathbf{J}, \mathbf{\Delta})\, d^3\mathbf{\Delta}. \tag{2.3}$$

Similarly, the number of stars that are scattered into this volume of phase space in the time interval δt is

$$(2\pi)^3 d^3\mathbf{J} \int f(\mathbf{J}-\mathbf{\Delta}, t)\, \delta P(\mathbf{J}-\mathbf{\Delta}, \mathbf{\Delta})\, d^3\mathbf{\Delta}. \tag{2.4}$$

Equating the difference between the number (2.4) scattered in and the number (2.3) leaving to the increase in the number in the given volume of phase space, and dividing through by $(2\pi)^3 d^3\mathbf{J}\,\delta t$, we obtain

$$\frac{\partial f}{\partial t} = \int f(\mathbf{J}-\mathbf{\Delta}, t)\, \dot{P}(\mathbf{J}-\mathbf{\Delta}, \mathbf{\Delta})\, d^3\mathbf{\Delta} - f(\mathbf{J}, t) \int \dot{P}(\mathbf{J}, \mathbf{\Delta})\, d^3\mathbf{\Delta}. \tag{2.5}$$

Since stars change their actions gradually, the rate $\dot{P}(\mathbf{J}, \mathbf{\Delta})$ is non-negligible only for small $\mathbf{\Delta}$.

Thus we may expand the product $f(\mathbf{J}, t)\,\dot{P}(\mathbf{J}, \boldsymbol{\Delta})$ in a power series in the J_i:

$$f(\mathbf{J}-\boldsymbol{\Delta}, t)\,\dot{P}(\mathbf{J}-\boldsymbol{\Delta}, \boldsymbol{\Delta}) = f(\mathbf{J})\,\dot{P}(\mathbf{J}, \boldsymbol{\Delta}) - \Delta_i \frac{\partial(f\dot{P})}{\partial J_i} + \tfrac{1}{2}\Delta_i\Delta_j \frac{\partial^2(f\dot{P})}{\partial J_i \partial J_j} + \dots \tag{2.6}$$

Substituting the first three★ terms of this series into (2.5), we find

$$\frac{\partial f}{\partial t} \simeq -\frac{\partial(f\overline{\Delta_i})}{\partial J_i} + \frac{1}{2}\frac{\partial^2(f\overline{\Delta_{ij}^2})}{\partial J_i \partial J_j}, \tag{2.7a}$$

where

$$\overline{\Delta_i}(\mathbf{J}) \equiv \int \Delta_i \dot{P}(\mathbf{J}, \boldsymbol{\Delta})\; d^3\boldsymbol{\Delta},$$

$$\overline{\Delta_{ij}^2}(\mathbf{J}) \equiv \int \Delta_i\Delta_j \dot{P}(\mathbf{J}, \boldsymbol{\Delta})\, d^3\boldsymbol{\Delta}. \tag{2.7b}$$

Equation (2.7a) is the orbit-averaged Fokker–Planck equation for the evolution of f, and we shall call the quantities $\overline{\Delta_i}$ and $\overline{\Delta_{ij}^2}$ the first- and second-order diffusion coefficients respectively. A less general orbit-averaged Fokker–Planck equation has been derived by Rosenbluth, MacDonald & Judd (1957).

3 The diffusion coefficients and their properties

Clearly, if the diffusion of orbits to be described by the Fokker–Planck equation is driven by discrete scattering events for which it is possible to calculate the probabilities $\dot{P}(\mathbf{J}, \boldsymbol{\Delta})$, it is straightforward to calculate the diffusion coefficients directly from equations (2.7b). For example, in so far as it is possible to consider the scattering of a star to occur at a point on its orbit, $\dot{P}$ may be calculated from standard kinetic theory (Chandrasekhar 1942; Spitzer & Schwarzschild 1953; Rosenbluth *et al.* 1957; Section 5 of this paper). However, this approach is unable to handle satisfactorily the scattering by spatially extended potential fluctuations. If the fluctuating gravitational field may be considered to be an externally imposed stochastic field, the diffusion coefficients may be calculated by using perturbation theory to follow the continuous evolution of one orbit into another.

We proceed by expressing the perturbing potential $\boldsymbol{\Phi}_1(\mathbf{x}, t)$ as a function of angle-action coordinates $(\boldsymbol{\theta}, \mathbf{J})$ for the phase-space of the unperturbed potential $\boldsymbol{\Phi}_0(\mathbf{x})$. We therefore write

$$\boldsymbol{\Phi}(\mathbf{x}, t) = \boldsymbol{\Phi}_0(\mathbf{x}) + \boldsymbol{\Phi}_1(\mathbf{x}, t) = \boldsymbol{\Phi}_0 + \sum_{n_1, n_2, n_3} \boldsymbol{\Psi}_{\mathbf{n}}(\mathbf{J}, t)\, \exp\,[i(\mathbf{n}\cdot\boldsymbol{\theta})]. \tag{3.1a}$$

where $|\boldsymbol{\Phi}_1/\boldsymbol{\Phi}_0| \ll 1$. Since $\boldsymbol{\Phi}_1$ is real, we have

$$\boldsymbol{\Psi}_{-\mathbf{n}} = \boldsymbol{\Psi}_{\mathbf{n}}^*. \tag{3.1b}$$

In the derivation of the Fokker–Planck equation (2.7a), successive changes were assumed to be uncorrelated and the diffusion coefficients $\overline{\Delta_{ij}^2}$ were defined such that for any sufficiently large time T, $T\overline{\Delta_{ij}^2}$ is the expectation value of the product $\Delta_i\Delta_j$ of the changes over time T in the actions of a star of undetermined initial angle coordinates $\boldsymbol{\theta}_0$. If we write the Hamiltonian in the form

$$H(\boldsymbol{\theta}, \mathbf{J}) = H_0(\mathbf{J}) + \boldsymbol{\Phi}_1(\boldsymbol{\theta}, \mathbf{J}) \qquad \text{with} \qquad \boldsymbol{\omega}_0(\mathbf{J}) \equiv \frac{\partial H_0}{\partial \mathbf{J}}, \tag{3.2}$$

★The second and third terms on the right of equation (2.6) have means that are of the same order because Δ_i alternates in sign, while Δ_i^2, though smaller, is always positive. Similarly, the averages of the fourth and fifth terms in the series are of the same order and smaller than the terms we retain.

then the half of Hamilton's equations which read

$$\dot{\mathbf{J}} = -\frac{\partial H}{\partial \boldsymbol{\theta}} = -\frac{\partial \Phi_1}{\partial \boldsymbol{\theta}}$$

$$= -i \sum_{\mathbf{n}} \mathbf{n} \Psi_{\mathbf{n}}(\mathbf{J}, t) \exp(i\mathbf{n} \cdot \boldsymbol{\theta}), \tag{3.3}$$

yield the actions as continuous functions of time. So to make the connection with our previous picture, we have to integrate equations (3.3) from initial conditions $(\boldsymbol{\theta}_0, \mathbf{J}_0)$ for a time T sufficiently long that changes in successive intervals of length T are uncorrelated. Since $|\Phi_1/\Phi_0| \ll 1$, we write

$$\mathbf{J}(t) = \mathbf{J}_0 + \boldsymbol{\Delta}_1(t) + \boldsymbol{\Delta}_2(t) + \ldots$$

$$\boldsymbol{\theta}(t) = \boldsymbol{\theta}_0 + \boldsymbol{\omega}_0 t + \boldsymbol{\theta}_1(t) + \ldots, \tag{3.4}$$

where $(\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \boldsymbol{\omega}_0 t,\ \mathbf{J} = \mathbf{J}_0)$ is the unperturbed orbit, the first-order perturbations $(\boldsymbol{\theta}_1, \boldsymbol{\Delta}_1)$ are obtained by integrating equations (A2) and (3.3) along this orbit, and $\boldsymbol{\Delta}_2$ is similarly obtained by integrating equations (3.3) along the first-order orbit. Thus

$$\boldsymbol{\Delta}_1(T) = -i \sum_{\mathbf{n}} \mathbf{n} \int_0^T \Psi_n(\mathbf{J}_0, t) \exp[i\mathbf{n} \cdot (\boldsymbol{\theta}_0 + \boldsymbol{\omega}_0 t)]\, dt. \tag{3.5}$$

Squaring this equation and averaging over all initial phases $\boldsymbol{\theta}_0$, we find that the only terms in the resulting double sum which survive are those for which $\mathbf{m} = -\mathbf{n}$. With the aid of equation (3.1b) we may therefore write

$$\langle \Delta_{1i} \Delta_{1j}(T) \rangle_\theta = \sum_{\mathbf{n}} n_i n_j \int_0^T dt \int_0^T dt'\, \Psi_{\mathbf{n}}(\mathbf{J}_0, t)\, \Psi^*_{\mathbf{n}}(\mathbf{J}_0, t') \exp[i\mathbf{n} \cdot \boldsymbol{\omega}_0 (t - t')], \tag{3.6}$$

where $\langle \ldots \rangle_\theta$ indicates an average over $\boldsymbol{\theta}_0$. Finally, we average over the random amplitudes $\Psi_{\mathbf{n}}$. We assume that the perturbing potential is a stationary random process, and hence that the autocorrelation function $c_{\mathbf{n}}$ of $\Psi_{\mathbf{n}}$ depends on t and t' only in the combination $t - t'$:

$$c_{\mathbf{n}}(\mathbf{J}, t - t') = \overline{\Psi_{\mathbf{n}}(\mathbf{J}, t)\, \Psi^*_{\mathbf{n}}(\mathbf{J}, t')},$$

$$c_{-\mathbf{n}}(\mathbf{J}, v) = c^*_{\mathbf{n}}(\mathbf{J}, v) = c_{\mathbf{n}}(\mathbf{J}, -v), \tag{3.7}$$

where an overline denotes both a θ-average and an ensemble average. Substituting equation (3.7) into the ensemble average of equation (3.6) and changing to new variables of integration $u \equiv t + t'$ and $v \equiv t - t'$, we obtain

$$\overline{\Delta_{1i} \Delta_{1j}} = \tfrac{1}{2} \sum_{\mathbf{n}} n_i n_j \int_{-T}^{T} dv\, c_{\mathbf{n}}(\mathbf{J}_0, v) \exp[i(\mathbf{n} \cdot \boldsymbol{\omega}_0) v] \int_{|v|}^{2T - |v|} dv$$

$$= \sum_{\mathbf{n}} n_i n_j \int_{-T}^{T} (T - |v|)\, c_{\mathbf{n}}(\mathbf{J}_0, v) \exp[i(\mathbf{n} \cdot \boldsymbol{\omega}_0) v]\, dv. \tag{3.8}$$

All the $c_{\mathbf{n}}(\mathbf{J}, v)$ are small for v greater than the autocorrelation time of Φ_1. Hence for T much larger than the autocorrelation time, the right side of equation (3.8) becomes proportional to T. Equating the coefficient of T in this expression to $\overline{\Delta^2_{ij}}$, we have, correct to second order

$$\overline{\Delta^2_{ij}} = \sum_{\mathbf{n}} n_i n_j \tilde{c}_{\mathbf{n}}(\mathbf{J}, \mathbf{n} \cdot \boldsymbol{\omega}), \tag{3.9a}$$

where $\tilde{c}_{\mathbf{n}}$ is the Fourier transform of the autocorrelation function;

$$\tilde{c}_{\mathbf{n}}(\mathbf{J}, \nu) \equiv \int_{-\infty}^{\infty} c_{\mathbf{n}}(\mathbf{J}, t) \exp(i\nu t)\, dt. \tag{3.9b}$$

Note that $\tilde{c}_{\mathbf{n}}$ satisfies the relations

$$\tilde{c}_{\mathbf{n}}(\mathbf{J}, \nu) = \tilde{c}^*_{\mathbf{n}}(\mathbf{J}, \nu) = \tilde{c}_{-\mathbf{n}}(\mathbf{J}, -\nu). \tag{3.9c}$$

3.1 RELATIONS BETWEEN THE DIFFUSION COEFFICIENTS

The calculation of the first-order diffusion coefficients $\overline{\Delta_i}$ is harder than that of $\overline{\Delta^2_{ij}}$ because we now have to go to second order in the perturbation in order to obtain results of equivalent accuracy. Fortunately this labour is in general unnecessary: in Appendix A we show that any externally applied random field generates diffusion coefficients which satisfy

$$\overline{\Delta_i} = \tfrac{1}{2} \frac{\partial \overline{\Delta^2_{ij}}}{\partial J_j}. \tag{3.10}$$

When we substitute equation (3.10) into equation (2.7a), the Fokker–Planck equation simplifies to

$$\frac{\partial f}{\partial t} = \tfrac{1}{2} \frac{\partial}{\partial J_i}\left(\overline{\Delta^2_{ij}} \frac{\partial f}{\partial J_j}\right). \tag{3.11}$$

The relation (3.10) does not apply if the scatterers suffer significant recoil when they scatter stars. However, we show in Appendix B that if the distribution function of such scatterers is thermal, $f \propto \exp(-\beta H)$, then the associated diffusion coefficients satisfy

$$\overline{\Delta_i} = \tfrac{1}{2}\left(\frac{\partial \overline{\Delta^2_{ij}}}{\partial J_j} - \beta \omega_j \overline{\Delta^2_{ij}}\right). \tag{3.12}$$

Thus in the high-temperature limit $\beta \to 0$, these diffusion coefficients also satisfy equation (3.10).

A useful restriction on the second-order diffusion coefficients is that the quadratic form

$$Q(\mathbf{x}) \equiv x_i \overline{\Delta^2_{ij}} x_j \geqslant 0 \tag{3.13}$$

is non-negative. To prove this, we observe that some transformation $\mathbf{J} \to \mathbf{J}' \equiv \mathbf{U} \cdot \mathbf{J}$, where $\mathbf{U}$ is an orthogonal matrix, will diagonalize the symmetric matrix $\overline{\mathbf{\Delta}^2}$, and that the eigenvalues λ_i of $\overline{\mathbf{\Delta}^2}$ are simply the mean-square changes in the components of $\mathbf{J}'$: $\lambda_i = \overline{(\Delta_i')^2}$.

3.2 ENTROPY GENERATION

Consider the rate of change of the H-function, $\mathscr{H} \equiv -\int C(f)\, d^3\mathbf{J}$, associated with some convex function $C(f)$ (Tremaine, Hénon & Lynden-Bell 1986):

$$\begin{aligned} \frac{d\mathscr{H}}{dt} &= -\int C'(f) \frac{\partial f}{\partial t}\, d^3\mathbf{J} \\ &= -\int C'(f)\left[-\frac{\partial (f\overline{\Delta_i})}{\partial J_i} + \tfrac{1}{2} \frac{\partial^2 (\overline{\Delta^2_{ij}} f)}{\partial J_i \partial J_j}\right] d^3\mathbf{J}. \end{aligned} \tag{3.14}$$

Eliminating the first-order coefficients with the aid of equation (3.12) and integrating by parts, this becomes

$$\frac{d\mathscr{H}}{dt}=\tfrac{1}{2}\int C''(f)\,\frac{\partial f}{\partial J_i}\,\overline{\Delta_{ij}^2}\left(\frac{\partial f}{\partial J_j}+\beta\omega_j f\right)\,d^3\mathbf{J}. \tag{3.15}$$

In the case $\beta=0$ of very massive scattering structures, the non-negativity of $d\mathscr{H}/dt$ is guaranteed by the condition (3.13) on the quadratic form Q and the convexity condition $C''\geqslant 0$. If, on the other hand, the scatterers are thermally distributed with non-zero inverse temperature β and the scattered stars are thermally distributed with inverse temperature β', we have

$$\frac{d\mathscr{H}}{dt}=\tfrac{1}{2}(\beta'-\beta)\beta'\int C''(f)\,f^2\omega_i\overline{\Delta_{ij}^2}\omega_j d^3\mathbf{J}, \tag{3.16}$$

which clearly has the same sign as $\beta'-\beta$. Thus the condition (3.13) ensures that any entropy of a heated system increases, while that of a cooled system decreases.

4 The Schwarzschild distribution as a solution of (3.11)

K. Schwarzschild (1907) found that the number density of stars near the Sun ($z\approx 0$) with peculiar velocities $\mathbf{v}$ in any velocity range $d^3\mathbf{v}$ is well fitted by the formula

$$f_0(\mathbf{v})\,d^3\mathbf{v}=\frac{n_0 d^3\mathbf{v}}{(2\pi)^{3/2}\sigma_R\sigma_\phi\sigma_z}\exp\left[-\left(\frac{v_R^2}{2\sigma_R^2}+\frac{v_\phi^2}{2\sigma_\phi^2}+\frac{v_z^2}{2\sigma_z^2}\right)\right], \tag{4.1}$$

where n_0, σ_R, σ_ϕ and σ_z are constants. Modern research (e.g. Bahcall 1984) has tended to confirm this conclusion (although the best available data still do not narrowly constrain the shapes of the velocity distributions of coeval stellar populations). Furthermore, Wielen (1977) presents persuasive evidence that the velocity dispersions σ_R and σ_z of a coeval stellar population tend to grow as $\sqrt{t}$. In view of the strengthened Jeans theorem and the Copernican principle, it is interesting to express Schwarzschild's velocity distribution in terms of isolating integrals, and to see under what circumstances it is a a self-similar solution of the Fokker–Planck equation (3.11) with the velocities scaling as $\sqrt{t}$.

We adopt as the three fundamental actions of orbits in an axisymmetric potential $\mathbf{\Phi}(R, z)$ the radial action J_r, the azimuthal action $J_a\equiv L_\phi$, and the latitudinal action J_l, which in a spherical potential reduces to the difference $L-|L_\phi|$ between the total angular momentum and its component parallel to the symmetry axis. The usual epicycle energies yield approximate expressions for J_r and J_l.

For small excursions from the plane, the z-energy $E_z\equiv\frac{1}{2}(v_z^2+\nu^2z^2)$, where $\nu^2=(\partial^2\mathbf{\Phi}/\partial z^2)$, is an approximate isolating integral and we have

$$J_l\simeq\frac{E_z}{\nu}=\frac{1}{2\nu}(v_z^2+\nu^2z^2). \tag{4.1b}$$

Furthermore, if R_g is the guiding-centre radius of an epicyclic orbit defined in terms of the galactic circular frequency $\Omega(R)$ by $L_\phi=R_g^2\Omega(R_g)$, then J_r is approximately related to the epicycle energy E_R by,

$$J_r\simeq\frac{E_R}{\varkappa}=\frac{1}{2\varkappa}[v_R^2+\varkappa^2(R-R_g)^2], \tag{4.2}$$

where $\varkappa^2=[(2\Omega/R)\,d(\Omega R^2)/dR]_{R_g}$ is the square of the epicycle frequency. Noting that the tangential component of velocity with respect to the local standard of rest may be written (e.g.

Binney & Tremaine 1987, section 7.5)

$$v_\phi = -\left(\frac{\varkappa^2}{2\Omega}\right)_{R_g}(R-R_g), \tag{4.3}$$

we may rewrite E_R to leading order in $(R-R_g)/R_g$ as

$$E_R = \tfrac{1}{2}[v_R^2+\gamma^2 v_\phi^2],$$

$$\simeq \varkappa J_r \tag{4.4a}$$

where

$$\gamma \equiv \frac{2\Omega(R_g)}{\varkappa(R_g)}. \tag{4.4b}$$

In the epicycle approximation, $\sigma_\phi^2=\sigma_R^2/\gamma^2$, so $v_R^2/\sigma_R^2+v_\phi^2/\sigma_\phi^2=2E_R/\sigma_R^2$. Substituting this expression into equation (4.1), we see that Schwarzschild's empirical distribution is $f_0(\mathbf{v})=f_S(\mathbf{x}_0, \mathbf{v})$, where

$$f_S(\mathbf{x},\mathbf{v}) \equiv \frac{n_0}{(2\pi)^{2/3}\sigma_R\sigma_\phi\sigma_z}\exp\left[-\left(\frac{E_R}{\sigma_R^2}+\frac{E_z}{\sigma_z^2}\right)\right], \tag{4.5}$$

and $\mathbf{x}_0$ is the Sun's position vector. By Jeans' theorem, f_S is a solution of the collisionless Boltzmann equation provided the quantities n_0 and σ_i occurring in it depend on the phase-space coordinates only through integrals of motion. Furthermore, in a stellar disc in which $\sigma_\phi \ll v_c$, radius R is tightly correlated with L_ϕ; at any point in the disc we have $\overline{(L_\phi-\bar{L}_\phi)^2}\ll \bar{L}_\phi^2$. Therefore we assume $n_0=n_0(L_\phi)$ and similarly for the σ_i. With these assumptions, it is now straightforward to show that the derivatives of a distribution function of the form (4.5) that generates an exponential disc of scale-length R_d, satisfy at $E_R \lesssim \sigma_R^2$ and $E_z \lesssim \sigma_z^2$ the inequalities,

$$\left|\frac{\partial \ln f_S}{\partial L_\phi}\right| \lesssim \frac{1}{v_c R_d}; \quad \left|\frac{\partial \ln f_S}{\partial J_r}\right| \approx \frac{\varkappa}{\sigma_R^2}; \quad \left|\frac{\partial \ln f_S}{\partial J_l}\right| \approx \frac{\nu}{\sigma_z^2}, \tag{4.6}$$

and that the second derivatives of $\ln f_S$ satisfy similar relationships. Thus derivatives of f_S with respect to L_ϕ are smaller than those with respect to J_r or J_l by a factor $\sim(v/v_c)^2$, where v is the typical epicyclic speed. If scattering occurs at points on stellar orbits and produces comparable velocity changes Δv in the different directions, then the diffusion coefficients are of order

$$\overline{\Delta_{rr}^2} \sim \frac{v^2}{\varkappa^2}\overline{(\Delta v)^2}; \quad \overline{\Delta_{\phi\phi}^2} \sim R^2\overline{(\Delta v)^2} \sim \frac{v_c^2}{\Omega^2}\overline{(\Delta v)^2}$$

$$\overline{\Delta_{r\phi}^2} \lesssim \frac{Rv}{\varkappa}\overline{(\Delta v)^2} \sim \frac{v_c v}{\Omega\varkappa}\overline{(\Delta v)^2}. \tag{4.7}$$

Thus $\overline{\Delta_{\phi\phi}^2}/\overline{\Delta_{rr}^2} \sim (v_c/v)^2$, $\overline{\Delta_{r\phi}^2}/\overline{\Delta_{rr}^2} \lesssim v_c/v$, and similarly for $\overline{\Delta_{l\phi}^2}$. Consequently, the terms in the Fokker–Planck equation involving L_ϕ are smaller than the other terms by at least a factor $\sim v/v_c$, and may be neglected. Hence we may write

$$2\frac{\partial f}{\partial t} = \frac{\partial}{\partial E_R}\left(D_{RR}\frac{\partial f}{\partial E_R}+D_{Rz}\frac{\partial f}{\partial E_z}\right)+\frac{\partial}{\partial E_z}\left(D_{Rz}\frac{\partial f}{\partial E_R}+D_{zz}\frac{\partial f}{\partial E_z}\right) \tag{4.8a}$$

where

$$D_{RR}\equiv \varkappa^2\overline{\Delta_{rr}^2}; \quad D_{Rz}\equiv \varkappa\nu\overline{\Delta_{rl}^2}; \quad D_{zz}\equiv \nu^2\overline{\Delta_{ll}^2}. \tag{4.8b}$$

The Copernican principle suggests that Schwarzschild's distribution (4.5) with n_0 and the σ_i functions of t and L_ϕ only, should satisfy equation (4.8a). We write the solution of equation (4.8a) as

$$f(E_R, E_z, t) = \beta_R \beta_z \exp[-\beta_R E_R - \beta_z E_z], \tag{4.9}$$

where β_R and β_z are undetermined functions of time, since then the total number of stars per unit angular momentum, $N = (2\pi)^3 \int_0^\infty dJ_r \int_0^\infty dJ_l f$, is automatically time-independent. Substituting equation (4.9) into equation (4.8a), and cancelling through by f, we find

$$2\left(\frac{\dot\beta_R}{\beta_R} + \frac{\dot\beta_z}{\beta_z}\right) - 2(\dot\beta_R E_R + \dot\beta_z E_z) = -\beta_R\left(\frac{\partial D_{RR}}{\partial E_R} + \frac{\partial D_{Rz}}{\partial E_z}\right) - \beta_z\left(\frac{\partial D_{zz}}{\partial E_z} + \frac{\partial D_{Rz}}{\partial E_R}\right)$$
$$+ \beta_R^2 D_{RR} + 2\beta_R\beta_z D_{Rz} + \beta_z^2 D_{zz}. \tag{4.10}$$

This equation can be valid for all t, E_R and E_z only if the first parenthesis on the left side, which clearly neither vanishes nor depends on either E_R or E_z can be balanced by a similar term on the right. Furthermore, the term on the left that is linear in E_R and E_z must be balanced by a similar linear term on the right. Evidently either $D_{RR} \propto E_R$ or $D_{RR} \propto E_z$, and similarly for D_{zz}. The natural *ansatz* is

$$D_{RR} = K_R E_R; \quad D_{Rz} = M; \quad D_{zz} = 2K_z E_z, \tag{4.11}$$

where K_R, K_z and M are constants. If the diffusion coefficients are of this form, we obtain on equating the coefficients of E_R on each side of equation (4.10), $\dot\beta_R/\beta_R^2 = -K_R$, which implies

$$\sigma_R(t) = \frac{1}{\sqrt{\beta_R}} = \sqrt{K_R(t - t_0)} \qquad (t_0 \text{ is a constant}), \tag{4.12}$$

and similarly for $\sigma_z(t)$. it is straightforward to check that the first parenthesis on the left of equation (4.10) is correctly matched on the right only if $M = 0$.

Thus our analysis connects the conclusion that one naturally draws from solar neighbourhood studies and the Copernican principle, that an initially Gaussian velocity distribution will stay Gaussian, to the observational evidence presented by Wielen (1977) that the velocity dispersion of a coeval population grows as $\sqrt{t}$. Put another way, the conclusion of this section is that in order to explain *both* the excellent fit of a Gaussian distribution to the z-velocities of the K giants, *and* the observed evolution of the dispersions, we only have to explain why the diffusion coefficients are given by equations (4.11).

5 The Spitzer–Schwarzschild problem

As an illustration of the application of equation (3.11) to disc star heating, we first employ the approximations introduced by Spitzer & Schwarzschild (1953) to follow the evolution of the distribution function $f(J_r, L_\phi)$ of a razor-thin disc in which are embedded a number of compact scatterers such as molecular clouds. For a full discussion of these assumptions, see, e.g. Lacey (1984).

5.1 DIFFUSION IN A RAZOR-THIN DISC

We consider only star-cloud encounters with impact parameters $p \ll a$, where a is the amplitude of the star's radial excursions. In the epicycle approximation, we have

$$R - R_g = a\cos(\varkappa t + \chi) = a\cos\theta_r$$
$$v_R = \dot R = -\varkappa a \sin\theta_r$$
$$= -\sqrt{2E_R}\sin\theta_r, \tag{5.1a}$$

where χ is a constant and

$$\theta_r \equiv \varkappa t + \chi \tag{5.1b}$$

is the epicycle phase. Substituting the first of equations (5.1a) into (4.3), we have that

$$v_\phi = -\frac{1}{\gamma}\sqrt{2E_R}\cos\theta_r. \tag{5.1c}$$

In the approximation that each encounter occurs at one point on the epicycle, and the cloud is much more massive than the star, $v_R^2 + v_\phi^2$ is unchanged by an encounter, so the post-encounter velocity $\tilde{\mathbf{v}}$ is simply the pre-encounter velocity $\mathbf{v}$ rotated by the deflection angle $\psi_{\rm defl}$. Hence

$$\begin{aligned}\tilde{v}_R &= v_R \cos\psi_{\rm defl} + v_\phi \sin\psi_{\rm defl}\\ &= -\sqrt{2E_R}\left(\sin\theta_r\cos\psi_{\rm defl} + \frac{1}{\gamma}\cos\theta_r\sin\psi_{\rm defl}\right).\end{aligned} \tag{5.2}$$

From equations (5.1a) and (5.2) it now follows that the change in v_R^2 induced by the encounter is

$$\begin{aligned}\Delta v_R^2 &\equiv \tilde{v}_R^2 - v_R^2\\ &= 2E_R\left[\frac{1}{2\gamma}\sin 2\theta_r \sin 2\psi_{\rm defl} - \sin^2\psi_{\rm defl}\left(\sin^2\theta_r - \frac{1}{\gamma^2}\cos^2\theta_r\right)\right].\end{aligned} \tag{5.3}$$

Since $|\tilde{\mathbf{v}}|^2 = |\mathbf{v}|^2$, differencing equation (4.4a) yields

$$\Delta J_r = \frac{1}{2\varkappa}(1-\gamma^2)\,\Delta v_R^2. \tag{5.4}$$

Combining equations (5.3) and (5.4) we have

$$\Delta J_r = E_R\frac{1-\gamma^2}{\varkappa}\left[\frac{1}{2\gamma}\sin 2\theta_r \sin 2\psi_{\rm defl} - \sin^2\psi_{\rm defl}\left(\sin^2\theta_r - \frac{1}{\gamma^2}\cos^2\theta_r\right)\right]. \tag{5.5}$$

Notice that ΔJ_r can have either sign depending on the phase θ_r at which the encounter occurs. Since $\gamma > 1$, J_r diminishes when the encounter occurs near $\theta_r = 0$ and the star enters the encounter with tangentially directed peculiar velocity.

The deflection angle $\psi_{\rm defl}$ is a function of p and the speed v with which the star approaches the scattering cloud. By equations (5.1), the approach speed is

$$v = \sqrt{v_R^2 + v_\phi^2} = \sqrt{2E_R}\left(\sin^2\theta_r + \frac{1}{\gamma^2}\cos^2\theta_r\right)^{1/2}. \tag{5.6}$$

Since v is typically larger than $\approx 7\,\mathrm{km\,s^{-1}}$ even for a freshly made star, and the characteristic internal velocity dispersion of the clouds $\sigma_c \approx 10\,\mathrm{km\,s^{-1}}$, most star–cloud encounters cause only small deflections. We model the molecular clouds as Plummer spheres of mass M_c and scale length r_c, and estimate the deflection in the impulse approximation. In a small-angle deflection, the velocity gained is almost perpendicular to the velocity of approach, so

$$|\psi_{\rm defl}| \approx \frac{2GM_c p}{v^2(p^2 + r_c^2)}. \tag{5.7}$$

If there are N_c clouds per unit area of the disc, a star moving at speed v encounters clouds with impact parameters in the range $(p+dp,\, p)$ at a rate $2N_c v\,dp$. Hence integrating over impact parameters smaller than a, we have

$$\int_p 2N_c v\psi_{\text{defl}}^2\,dp = \frac{8G^2M_c^2N_c}{v^3}\int_0^a \frac{p^2\,dp}{(p^2+r_c^2)^2} \simeq \frac{2\pi G^2M_c^2N_c}{v^3 r_c}, \tag{5.8}$$

where we have approximated the integral by its value in the limit $a \to \infty$.

Squaring equation (5.4) and averaging over the epicycle phase θ_r, we have to leading order in ψ_{defl};

$$\overline{\Delta_{rr}^2} \simeq E_R^2 \frac{(1-\gamma^2)^2}{\varkappa^2\gamma^2}\frac{2}{\pi}\int_0^{2\pi}\left(\frac{2\pi G^2M_c^2N_c}{v^3 r_c}\right)\cos^2\theta_r \sin^2\theta_r\,d\theta_r. \tag{5.9}$$

With equation (5.7) this becomes,

$$\overline{\Delta_{rr}^2} = \sqrt{2E_R}\,\frac{2G^2M_c^2N_c}{\varkappa^2 r_c}\left(\frac{2A^2}{\varkappa^2}\right)\int_0^{2\pi}\frac{\cos^2\theta_r \sin^2\theta_r}{(\sin^2\theta_r+\gamma^{-2}\cos^2\theta_r)^{3/2}}\,d\theta_r$$

$$= \sqrt{2E_R}\,\frac{2G^2M_c^2N_c}{\varkappa^2 r_c}\,\mathcal{T}, \tag{5.10}$$

Here $A \equiv -\frac{1}{2}R_g(d\Omega/dR)_{R_g} = \Omega(1-\gamma^{-2})$ is Oort's constant and [Gradshteyn & Ryzhik 1965, equations (2.584.40) and (2.584.47)]

$$\mathcal{T} \equiv \left(\frac{4\Omega}{\varkappa}\right)^2\left[\left(1-\frac{A}{2\Omega}\right)K(\sqrt{A/\Omega}) - E(\sqrt{A/\Omega})\right], \tag{5.11}$$

where K and E are the usual complete elliptic integrals. In the solar neighbourhood, $A/\Omega \simeq 0.54$, $\Omega/\varkappa \simeq 0.74$ and $\mathcal{T} \simeq 0.43$.

We define a dimensionless time τ and a dimensionless energy ξ by

$$\tau \equiv \frac{4GM_cN_c\mathcal{T}}{U}t \quad \text{and} \quad \xi \equiv \frac{2E_R}{U^2}, \tag{5.12a}$$

where

$$U^2 \equiv \frac{GM_c}{r_c}. \tag{5.12b}$$

Notice that U is the characteristic internal velocity of the clouds, and $(2\pi GM_cN_c)t$ is the velocity that would be gained in time t by a test particle in falling towards a uniform sheet with the same surface density as that formed by the clouds. Substituting equation (5.10) and these definitions into equations (4.8) and retaining only terms involving E_R alone, we have

$$\frac{\partial f}{\partial \tau} = \frac{\partial}{\partial \xi}\left(\sqrt{\xi}\,\frac{\partial f}{\partial \xi}\right). \tag{5.13}$$

Equation (5.13) will henceforth be known as the Spitzer–Schwarzschild equation.

5.2 SELF-SIMILAR SOLUTIONS OF THE SPITZER–SCHWARZSCHILD EQUATION

Equation (5.13) admits self-similar solutions of the form

$$f_\alpha(\xi,\tau) = \tau^\alpha F(X), \quad \text{where} \quad X \equiv \xi^{3/2}/\tau, \tag{5.14}$$

and α is an arbitrary constant. On integrating equation (5.14) over all ξ (which is proportional to the action J_r), we have that the total number of stars in the system described by f_α is $\propto t^{\alpha+2/3}$. Thus for $\alpha \geqslant -\frac{2}{3}$, solutions to equation (5.14) describe discs in which the stellar birth rate is proportional to a power of time. Inserting equation (5.14) into equation (5.13) we find that F must satisfy the differential equation:

$$X\frac{d^2F}{dX^2}+\tfrac{2}{3}(1+\tfrac{2}{3}X)\frac{dF}{dX}-\tfrac{4}{9}\alpha F=0. \tag{5.15}$$

It is straightforward to find solutions to this equation of the form $F=\Sigma_n^\infty=0\, a_n X^{n+c}$, where $c=0$ or $\frac{1}{3}$. The required solution is that linear combination of these series that has a convergent integral $\int F\,d\xi$ for the total number of stars. At large X, this solution goes as $\exp(-\frac{4}{9}X)\,X^{-(\alpha+2/3)}$. In the interesting special case $\alpha=-\frac{2}{3}$ of a constant number of stars, the required solution to equation (5.15) is simply $F\propto\exp(-\frac{4}{9}X)$; that is, the physically interesting solution employs only the power series with $c=0$, since this series corresponds to zero flux of stars through $\xi=0$. Thus substituting from equations (4.4a), (5.12a) and (5.14) we see that at late times the distribution function of a coeval population tends to

$$f\propto t^{-2/3}\exp\left[\frac{-(v_R^2+\gamma^2v_\phi^2)^{3/2}}{9U^2GM_cN_c\mathcal{T}t}\right]. \tag{5.16}$$

Hence the velocity distribution is not Gaussian, and the dispersion grows as $t^{1/3}$.

5.3 THREE-DIMENSIONAL SPITZER–SCHWARZSCHILD DIFFUSION

In Appendix C we use the usual Spitzer–Schwarzschild approximations to derive for a disc of finite thickness the second-order diffusion tensor **D** defined by equations (4.8b). We find that

$$\mathbf{D}=C\tilde{\mathbf{D}}(\eta),\qquad \text{where}\qquad \eta\equiv\arctan(E_z/E_R) \tag{5.17}$$

and the dimensional constant C is defined by equation (C8). The dashed curve in Fig. 1 shows for the case $\gamma=\sqrt{2}$ of a flat rotation curve the variation with η of the ratio λ_2/λ_1 of the two eigenvalues of $\tilde{\mathbf{D}}$. This ratio never rises above 0.1. Thus $\tilde{\mathbf{D}}$ is extremely anisotropic, permitting rapid diffusion along the curves that are everywhere tangent to the eigenvector of $\tilde{\mathbf{D}}$ associated with $\tilde{\mathbf{D}}$'s larger eigenvalue, but very slow diffusion perpendicular to these curves. Consequently, in a relaxed disc, contours of constant f will approximately coincide with these curves, which are shown in Fig. 2 for $\gamma=\sqrt{2}$.

We estimate the time required to establish this state of affairs as follows. With each of the curves shown in Fig. 2 we may associated a characteristic time for diffusion from end to end $t_d\approx(E_z/\cos\phi)^2/(\lambda_1 C)$, where E_z is the curve's intercept with the E_z axis, and the quantities $\phi\simeq 60°$ and $\lambda_1\simeq 2$ may be read from Figs 1 and 2. In the solar neighbourhood $C\approx(25\,\mathrm{km\,s^{-1}})^4/10^{10}\,\mathrm{yr}$ (e.g. Lacey 1984), so $t_d\approx 10^{10}\,\mathrm{yr}$ for $E_z\approx(21\,\mathrm{km\,s^{-1}})^2$, which is slightly smaller than the value of E_z to which Wielen's (1977) age–velocity–dispersion relationship seems to asymptote. Thus star–cloud scattering should be able to hold the isodensity contours of all but the hottest disc populations parallel to the curves of Fig. 2 independently of what mechanism is responsible for diffusing stars perpendicular to these curves.

The functional dependence $\tilde{\mathbf{D}}(\eta)$ causes the Fokker–Planck equation (4.8a) to admit self-similar solutions of the form

$$t^\alpha F(X_R,X_z)\qquad\text{where}\qquad X_i\equiv\frac{E_i}{\sqrt{Ct}} \tag{5.18a}$$

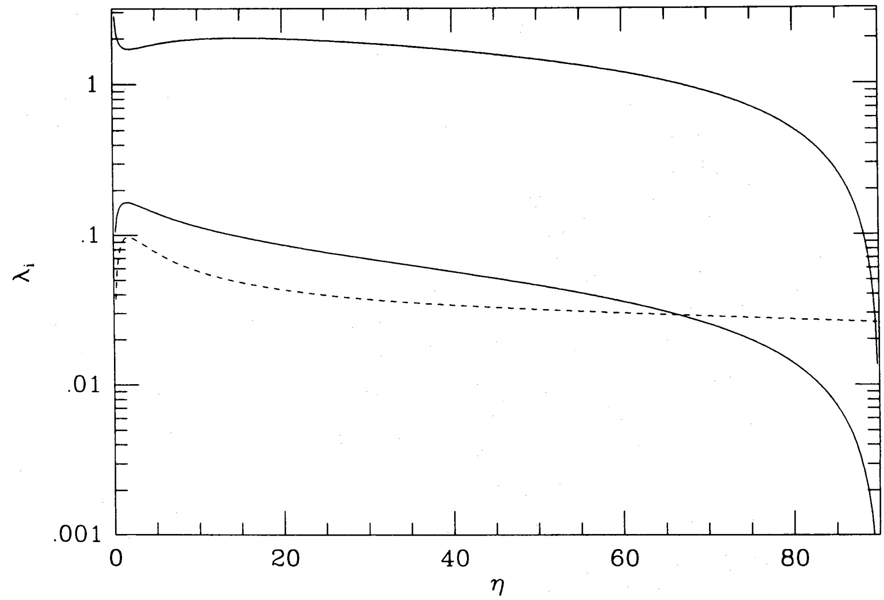


Figure 1. The full curves show as a function of the angle η defined by equation (5.17) the eigenvalues λ_1 and λ_2 of the dimensionless diffusion tensor $\widetilde{\mathbf{D}}$ generated by a thin layer of clouds in a disc with a perfectly flat rotation curve. The dashed line shows the ratio λ_2/λ_1.

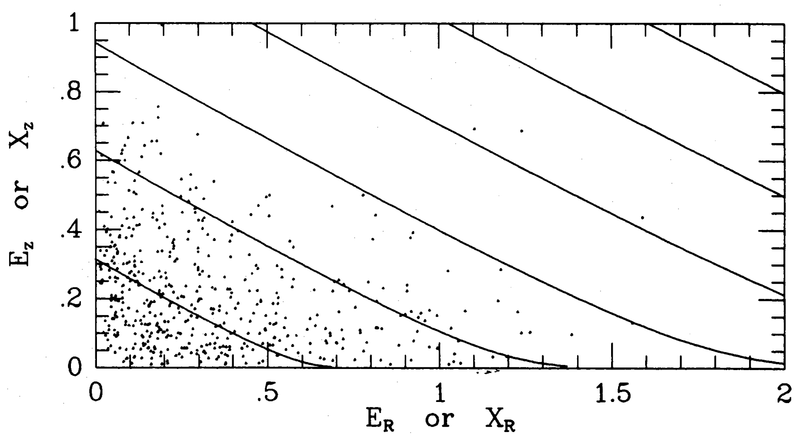


Figure 2. Curves in the (E_R, E_z) plane everywhere parallel to the eigenvector of $\widetilde{\mathbf{D}}$ that has the larger eigenvalue. In a steady-state disc, scattering by clouds would tend to establish a distribution function that is constant on these curves. Also shown are the positions of the points of a 500-particle Monte-Carlo simulation of the solution of equation (5.18b) in the case $\alpha = -1$.

and F satisfies

$$\frac{\partial}{\partial X_i}\left(\widetilde{D}_{ij}\frac{\partial F}{\partial X_j} + X_i F\right) = 2(1+\alpha)F. \tag{5.18b}$$

Integrating equation (5.18a) over all actions, we have that the number of stars in the system is $\propto t^{1+\alpha}$. We shall concentrate on the important special case $\alpha = -1$ of a coeval population.

The points shown in Fig. 2 are the final positions of the 500 particles of a Monte-Carlo simulation of the solution of equation (5.18b) in the case $\alpha = -1$. These particles have mean ratio

$\sigma_z/\sigma_R=\sqrt{\overline{X_z}/\overline{X_R}}=0.79\pm0.02$. Since all the integral curves of Fig. 2 are similar to one another, the hypothesis that these curves should coincide with isodensity contours, predicts that the square root of the ratio $\overline{E_z}/\overline{E_R}$ of the mean values of E_z and E_R along any such curve should be of order 0.79. Unfortunately, $\overline{E_z}/\overline{E_R}$ is zero along any curve of Fig. 2 because the latter asymptotes to the E_R axis. In reality the isodensity contours must depart from the curves of Fig. 2 a short distance above the E_R axis, since otherwise an infinite density gradient would arise above this axis, parallel to a direction in which the diffusion coefficient is non-zero. If the isodensity contours hit the E_R axis just below the point on the corresponding curve of Fig. 2 which has coordinate $\eta=\eta_{\min}$, then we find that $\sqrt{\overline{E_z}/\overline{E_R}}$ falls from 0.7888 at $\eta_{\min}=0.05$ to 0.65 at $\eta_{\min}=0.005$. These numbers may be compared with Lacey's (1984) analytic estimate $\sigma_z/\sigma_R=0.78$ and the value $\sigma_z/\sigma_R=0.6\pm0.04$ from Villumsen's (1985) numerical experiments. Taking our cue from our Monte-Carlo simulation, we henceforth adopt $\eta_{\min}=0.05$.

In the approximation that f is constant on each of the integral curves of Fig. 2, and there is negligible diffusion into or out of the wedge $\eta<\eta_{\min}$, equation (5.18b) may be solved as follows. If $\tilde{\mathbf{X}}(\eta)$, $0\leqslant\eta\leqslant(\pi/2)$, is some curve in Fig. 2, then we may construct coordinates (ξ,η) for the active portion of the X-plane by writing

$$\mathbf{X}(\xi,\eta)=\xi\tilde{\mathbf{X}}(\eta); \tag{5.19}$$

each curve in Fig. 2 is now a curve of constant ξ. Then setting $\alpha=-1$, integrating equation (5.18b) through the region of $\eta>\eta_{\min}$ between two adjacent curves of constant ξ and using the divergence theorem, we have in conventional vector notation

$$\int_{\xi+d\xi}\hat{\mathbf{n}}\cdot(\tilde{\mathbf{D}}\cdot\nabla F+\mathbf{X}F)\,dl-\int_{\xi}\hat{\mathbf{n}}\cdot(\tilde{\mathbf{D}}\cdot\nabla F+\mathbf{X}F)\,dl=0, \tag{5.20a}$$

where

$$dl\equiv\xi\left|\frac{d\tilde{\mathbf{X}}}{d\eta}\right|d\eta \tag{5.20b}$$

is an increment in distance along a curve of constant ξ, and

$$\hat{\mathbf{n}}\equiv\left|\frac{d\tilde{\mathbf{X}}}{d\eta}\right|^{-1}\left(\frac{d\tilde{\mathbf{X}}_z}{d\eta},\ -\frac{d\tilde{\mathbf{X}}_R}{d\eta}\right) \tag{5.20c}$$

is the upward-pointing unit normal to the curve. Since the tensor $\tilde{\mathbf{D}}$ is diagonal in a frame aligned with $\hat{\mathbf{n}}$, we obtain on dividing equation (5.20a) through by $d\xi$

$$\frac{d}{d\xi}\left[\int dl\,(\hat{\mathbf{n}}\cdot\tilde{\mathbf{D}}\cdot\hat{\mathbf{n}})(\hat{\mathbf{n}}\cdot\nabla)\,F(\xi)+\int dl(\hat{\mathbf{n}}\cdot\mathbf{X})\,F(\xi)\right]=0. \tag{5.21}$$

On substituting from equations (5.20b, c) for dl and $\hat{\mathbf{n}}$, this is seen to be of the form

$$\xi\frac{dF}{d\xi}+K\xi^2F=\text{constant}, \tag{5.22a}$$

where the constant K is defined by

$$K\equiv\int_{\eta_{\min}}^{\pi/2}d\eta\,(\tilde{X}_R\dot{\tilde{X}}_z-\tilde{X}_z\dot{\tilde{X}}_R)\Big/\int_{\eta_{\min}}^{\pi/2}d\eta\,(\hat{\mathbf{n}}\cdot\tilde{\mathbf{D}}\cdot\hat{\mathbf{n}})\,\frac{|\dot{\tilde{\mathbf{X}}}|^2}{\tilde{X}_R\dot{\tilde{X}}_z-\tilde{X}_z\dot{\tilde{X}}_R}. \tag{5.22b}$$

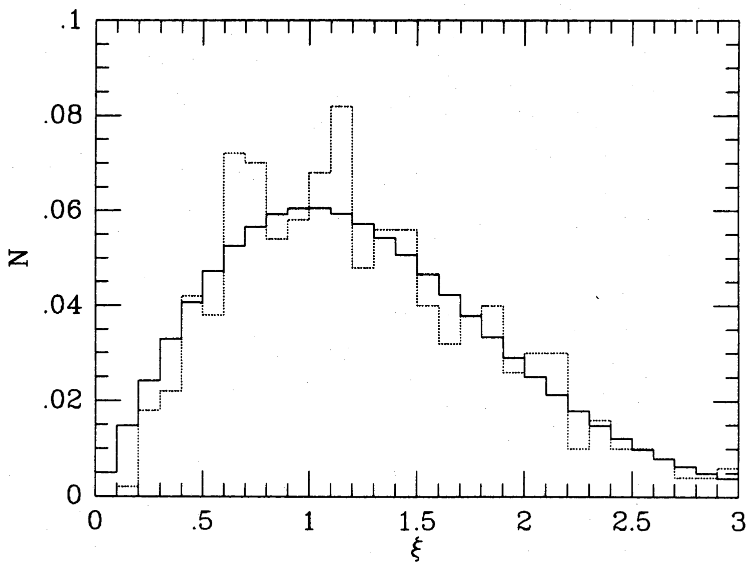


Figure 3. Histograms of the distribution of ξ values predicted by equation (5.23) (full lines) and the distribution of ξ values in our Monte-Carlo simulation of the solution to equation (5.18b) with $\alpha=-1$ (dotted lines). In the latter case particles were assigned ξ coordinates under the assumption $\eta_{\min}=0.05$.

By an appropriate choice of the curve $\tilde{\mathbf{X}}(\eta)$, and thus of a particular scaling for ξ, we may set $K=1$; for $\eta_{\min}=0.05$ we then have $X_z=0.314\,\xi$ on $X_R=0$. With $K=1$, the appropriate solution of equation (5.22a) is

$$F(\xi)=F_0\exp(\xi^2/2). \tag{5.23}$$

Thus three-dimensional Spitzer–Schwarzschild diffusions tends to establish a distribution function that is Gaussian in the *energies* rather than in the velocities. Fig. 3 compares the histogram of ξ values predicted by equation (5.23), $N(\xi)\propto\xi\exp(-\xi^2/2)$, with the ξ values assigned to the particles of our Monte-Carlo simulation when $\eta_{\min}=0.05$. The agreement is excellent.

6 Scattering by spiral structure

In Section 4 we showed that the observed stellar velocity-distribution near the Sun requires that the diffusion tensor $\overline{\Delta^2}(\mathbf{J})$ scales as $\mathbf{J}$, and in Section 5 we have shown that the Spitzer–Schwarzschild process is inconsistent with such behaviour. At this point it is natural to enquire whether this inconsistency arises from the rather marginal validity of the approximations introduced by Spitzer & Schwarzschild (1953). In Appendix D we show that when one drops the assumption that scattering occurs at a point on the orbit, one obtains for the planar problem a diffusion coefficient $\overline{\Delta^2_{rr}}$ that scales as J_r for epicycle amplitudes $a\lesssim r_c$, and as $J_r^{1/2}$ for $a\gtrsim r_c$. Thus to obtain the desired growth of $\overline{\Delta^2_{rr}}$ for stars with random velocities $\sim 40\,\mathrm{km\,s^{-1}}$ we would require clouds of size $r_c\gtrsim a=\sqrt{(2\bar{v}_R^2)}/\varkappa\simeq 1.5\,\mathrm{kpc}$. Unfortunately such extended clouds would be ripped apart by the shear of Galactic rotation. An additional effect is that a massive cloud orbiting in the disc will grow around itself a stellar wake in the form of a trailing spiral, which may considerably enhance scattering by the cloud (Julian & Toomre 1966; Julian 1967). Hence we are led to consider the possibility that the scattering objects are spiral density waves.

From equation (3.9a) it follows that a steady spiral wave heats stars only at the Lindblad resonances (Lynden-Bell & Kalnajs 1972): in this case $\boldsymbol{\Psi}_\mathbf{n}\propto\exp(-im\Omega_\mathrm{p}t)$, so the power spectrum $\tilde{c}_\mathbf{n}(\nu)\propto\delta(\nu-m\Omega_\mathrm{p})$ and there is no contribution to equation (3.9a) unless Ω_p satisfies a Lindblad condition $m\Omega_\mathrm{p}=l\varkappa+m\Omega$. Since the heating of discs is clearly not localized, we conclude that either (i) the disc is heated by large-scale spiral features that last only a dynamical time or so, and thus give rise to $\tilde{\mathcal{T}}_{\mathbf{n}}$ that are not strongly peaked around any particular frequency, or (ii) an infinite number of normal modes is excited, such that every star lies at a resonance of *some* mode. These two points of view would be equivalent if one were assured that the system had a complete set of normal modes, since any dynamically possible surface density $\Sigma(\mathbf{x}, t)$ could then be represented as a superposition of normal modes. However, we adopt the picture in which the spiral arms are ephemeral, since this picture is the more effective in practice.

6.1 THE ANGLE REPRESENTATION OF A SINGLE SPIRAL WAVE

Consider the effects of the spiral potential

$$\mathbf{\Phi}_1 = \varepsilon(t) \exp[i(kR + m\phi)]. \tag{6.1}$$

We assume that the background potential $\mathbf{\Phi}_0$ generates a circular speed $v_c(L_\phi)$, and evaluate $\overline{\Delta^2_{rr}}$ for $J_l = 0$ and values of J_r sufficiently small that the epicycle approximation is valid. In the epicycle approximation, J_r is related to the radial energy E_R by equation (4.4a), and we have [see equations (5.1)]

$$R = R_g(L_\phi) + a(E_R)\cos\theta_r; \qquad \phi = \theta_\phi + \frac{\gamma a(E_R)}{R_g(L_\phi)}\sin\theta_r, \tag{6.2}$$

where $R_g \equiv L_\phi / v_c$ is the guiding-centre radius, $\varkappa$ is the usual epicycle frequency, γ is defined by equation (4.4b) and $a \equiv \sqrt{2E_R}/\varkappa$ is the epicycle amplitude. Hence

$$\mathbf{\Phi}_1 = \varepsilon(t) \exp(ikR_g) \exp(ika\cos\theta_r) \exp(im\phi) \exp\left(i\frac{m\gamma a}{R_g}\sin\theta_r\right). \tag{6.3}$$

Equation (8.511.4) of Gradshteyn & Ryzhik (1965)* enables us to rewrite $\mathbf{\Phi}_1$ as the product of two Fourier series in θ_r:

$$\mathbf{\Phi}_1 = \varepsilon(t)\exp(ikR_g)\sum_{n=-\infty}^{\infty} i^n J_n(ka)\exp(in\theta_r) \times \sum_{n'=-\infty}^{\infty} J_{n'}\left(\frac{m\gamma a}{R_g}\right)\exp(in'\theta_r)\exp(im\theta_\phi), \tag{6.4}$$

where the J_n, are now Bessel functions rather than actions. Replacing n' with the new summation variable $l \equiv n + n'$, and using formula (8.530.2) of Gradshteyn & Ryzhik (1965), equation (6.4) becomes

$$\mathbf{\Phi}_1 = \sum_{l=-\infty}^{\infty} \varepsilon(t) \exp[i(kR_g + l\alpha)]\, J_l(\mathscr{K}'a) \exp[i(l\theta_r + m\theta_\phi)], \tag{6.5a}$$

where

$$\alpha \equiv \arctan\left(\frac{m\gamma}{kR_g}\right) \quad \text{and} \quad \mathscr{K}' \equiv \sqrt{k^2 + \frac{m^2\gamma^2}{R_g^2}}. \tag{6.5b}$$

Comparing equation (6.5a) with equation (3.1a), we identify $\mathbf{\Psi}_{lm}$ as the coefficient of $\exp[i(l\theta_r + m\theta_\phi)]$ in equation (6.5a):

$$\mathbf{\Psi}_{lm} = \varepsilon(t)\, J_l(\mathscr{K}'a) \exp[i(kR_g + l\alpha)]. \tag{6.6}$$

*In the form $\exp(iz\cos\theta) = \Sigma_{-\infty}^{\infty} i^n J_n(z)\exp(in\theta)$. Inserting the identity $\sin\theta = \cos[(\theta - (\pi/2)]$ into this formula, we have $\exp(iz\sin\theta) = \Sigma_{-\infty}^{\infty} J_n(z)\exp(in\theta)$.

When we substitute this expression for Ψ_n and the corresponding result for the complex conjugate potential Φ_1^* into equation (3.9a) and approximate ω_ϕ with the circular frequency $\Omega(L_\phi)$, we find that the diffusion coefficient generated by the real spiral potential $\Phi_1+\Phi_1^*$ is

$$\overline{\Delta_{rr}^2}=2\sum_{l=-\infty}^{\infty} l^2 J_l^2(\mathcal{K}'a)\,\tilde{\varepsilon}(l\kappa+m\Omega), \tag{6.7}$$

where $\tilde{\varepsilon}(\nu)$ is the Fourier transform of the autocorrelation of $\varepsilon(t)$ and we have taken advantage of the identity $J_l(z)=(-)^l J_{-l}(z)$.

6.2 EFFECTIVENESS OF SPIRAL HEATING

By formula (8.536.2) of Gradshteyn & Ryzhik (1965) we have

$$\sum_{l=1}^{\infty} l^2 J_l^2(z)=\tfrac{1}{4}z^2. \tag{6.8}$$

Hence if the power spectrum $\tilde{\varepsilon}(\nu)$ of $\varepsilon(t)$ were independent of ν we would have $\overline{\Delta_{rr}^2}\propto(\mathcal{K}'a)^2\propto J_r$ in accordance with observation; i.e. a white-noise power spectrum generates a diffusion tensor of the required form

In practice we do not require that $\tilde{\varepsilon}$ be absolutely independent of ν: $|J_l(z)|$ declines steeply for $l>z$, so the first z terms in the series of equation (6.8) already sum to $\simeq\frac{1}{4}z^2$. Hence a power spectrum which falls off for $|\nu-m\Omega|>l_{\max}\kappa$ yields $\overline{\Delta_{rr}^2}\propto J_r$ for $a\lesssim l_{\max}/\mathcal{K}'$.

A spiral wave is likely to have a significant amplitude only in the radius range where the condition $\Omega-\kappa/m<\Omega_p<\Omega+\kappa/m$ is satisfied. Fig. 4 illustrates this situation. The width Γ of the peak in the power spectrum results from two effects: individual spiral patterns have finite growth and decay times, and successive spiral patterns may have different pattern speeds. It is hard to imagine any density perturbation forming or dispersing in less than the local epicycle period, so $\Gamma\lesssim\kappa$ seems a reasonable limit. In this case, illustrated in Fig. 4, we have $l_{\max}=1$, and so $\overline{\Delta_{rr}^2}\propto[J_1(\mathcal{K}'a)]^2$. The desired behaviour $\overline{\Delta_{rr}^2}\propto J_r$ is obtained only if $\mathcal{K}'a\lesssim 1$

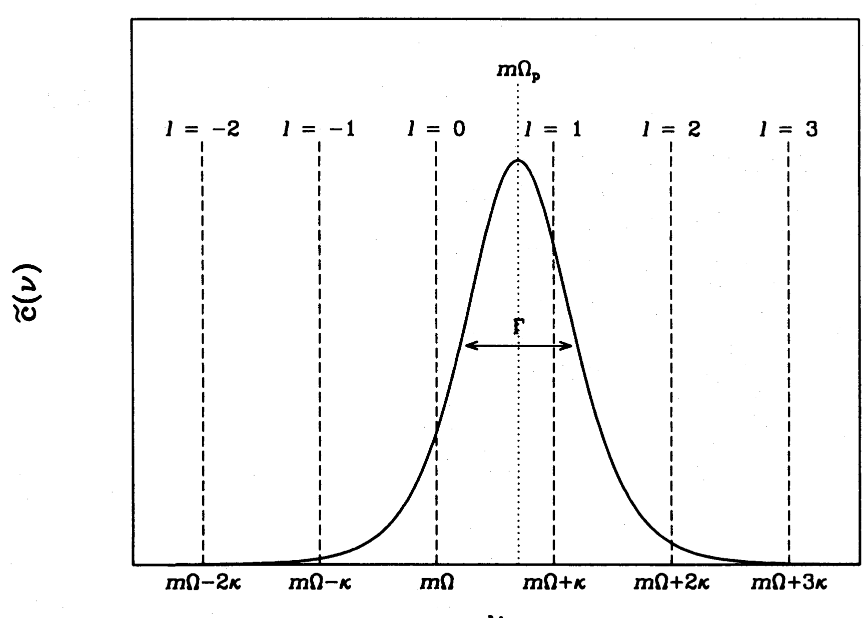


Figure 4. The power spectrum of a typical transient spiral perturbation.

(Carlberg & Sellwood 1985); by $z=1.1$, $dJ_1^2(z)/d(z^2)$ has fallen to half its value at the origin, and by $z=1.85$, $J_1(z)$ has started to *decrease* as a function of z.

For a typical solar neighbourhood star, with rms radial velocity $\overline{v_R^2}^{1/2}=40\ \mathrm{km\,s^{-1}}$, we have $a\simeq 1.5$ kpc, so that $\mathcal{K}'a\lesssim 1$ requires an effective wavelength $\lambda\equiv 2\pi/\mathcal{K}'\gtrsim 9$ kpc. Equivalently, m must be small ($m\lesssim 4$) and/or the spiral pitch angle large ($i\gtrsim 20°$ for $m=2$ and $i\gtrsim 40°$ for $m=3$).

A consideration which we have neglected is that the rms amplitude $\overline{\varepsilon^2(t)}^{1/2}$ of the spiral structure may vary secularly as the disc evolves. This possibility is discussed in detail by Carlberg & Sellwood (1985). The effect may be either to boost or depress the velocity dispersion of old stars relative to the case that $\overline{\varepsilon^2}$ is constant at its present value. However, with a suitable decline in the value of $\overline{\varepsilon^2}$ it is possible to reproduce the observed relation $\sigma\propto t^{1/2}$ with $\overline{\Delta_{rr}^2}/J_r$ a decreasing function of J_r – but even in this case the velocity distribution would in general be non-Gaussian.

6.3 VERTICAL HEATING BY SPIRALS

The previous discussion can be generalized to include the vertical motions of stars. The perturbing potential (6.1) is multiplied by a factor $\zeta(z)$ which gives its dependence on z. We assume that ζ is symmetric in z: $\zeta(-z)=\zeta(z)$. The Fourier coefficient of the potential Ψ_{lmR} is given by (6.6) multiplied by a factor F_n, where

$$F_n(J_l)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\zeta[z(J_l,\theta_l)]\exp(-in\theta_l)\,d\theta_l. \tag{6.9}$$

It follows from the symmetry of $\zeta(z)$ and of the z-motion that F_n is non-vanishing only for even n. The expressions for the diffusion coefficients are now of the form

$$\overline{\Delta_{ll}^2}=2\sum_{l=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}n^2J_l^2(\mathcal{K}'a)\,|F_n|^2\tilde{c}(l\varkappa+n\nu_z+m\Omega). \tag{6.10}$$

We consider the case of harmonic z-motion: $z=b\sin\theta_l$, $\theta_l=\nu_z t+\phi$; $J_l=E_z/\nu_z=\tfrac{1}{2}\nu_z b^2$. We further restrict ourselves to the case in which the thickness h of the spirals is small compared to their wavelength: $k_s h\ll 1$, where $k_s\equiv\sqrt{k^2+m^2/R^2}$. Then $\zeta(z)$ has the limiting behaviours

$$\zeta(z)\begin{cases}\approx 1-(k_s z^2/2h) & k_s|z|\ll k_s h\\ \propto\exp(-k_s|z|) & k_s h\ll k_s|z|\ll 1\end{cases}, \tag{6.11}$$

where we have adopted the normalization $\zeta(0)=1$. Evaluating F_n for $k_s b\ll 1$, we find that $\overline{\Delta_{rr}^2}$ is still given by equation (6.7), independent of E_z, while $\overline{\Delta_{rl}^2}$ and $\overline{\Delta_{ll}^2}$ are both proportional to E_z^2 when $b\ll h$, and proportional to E_z when $b\gg h$. If, in addition, $\mathcal{K}'a\ll 1$, then $\overline{\Delta_{rr}^2}$ and $\overline{\Delta_{rl}^2}$ are proportional to E_R and $\overline{\Delta_{ll}^2}$ is independent of E_R. The amplitudes of the diffusion coefficients depend on the power spectrum $\tilde{c}(\nu)$. The largest contributions to $\overline{\Delta_{rl}^2}$ and $\overline{\Delta_{ll}^2}$ come from the terms with $n=\pm 2$, and involve factors $\tilde{c}(m\Omega\pm 2\nu_z\pm\varkappa)$ and $\tilde{c}(m\Omega\pm 2\nu_z)$, respectively. For stars in the solar neighbourhood, $\nu_z/\varkappa\simeq 2$–3, so by the argument given in Section 6.2 regarding the local and width of the peak in $\tilde{c}(\nu)$, these factors are expected to be very much smaller than $\tilde{c}(m\Omega\pm\varkappa)$. Therefore we expect that $\overline{\Delta_{ll}^2}$, $\overline{\Delta_{rl}^2}\ll\overline{\Delta_{rr}^2}$ and conclude that vertical heating by large-scale spirals is negligible – a conclusion similar to that already reached by Carlberg (1984, 1987).

This conclusion may be somewhat mitigated if one includes higher harmonic components in the spiral potential, at multiples $q=2, 3, \ldots$ of the fundamental wavenumbers (k, m), as might for instance result from the non-linear response of the interstellar gas to the spiral potential. These would produce subsidiary peaks in the power spectrum centred at frequencies $qm\Omega_p$ with widths Γ, which could overlap the frequencies $qm\Omega\pm 2\nu_z$ for q of a few. The subsidiary peaks would however have amplitudes suppressed by several powers of q relative to the main peak, so the effect from any one such peak should be small. If the spiral patterns contain structure down to very small scales, then the cumulative effects of many harmonics might be significant for vertical heating, but then one is closer to the picture of scattering by small-scale 'clouds' than one involving large-scale 'arms'.

7 Conclusion

Lumps and bumps in the potentials of stellar systems cause the orbits of stars to evolve on time-scales that greatly exceed the orbital times. If, as is frequently the case, orbits in the potential of the underlying smooth stellar system may be coordinated into three-dimensional orbit space, this slow evolution of orbits manifests itself as a diffusion of stars through orbit space. In Section 2 we have shown that when actions are used as the coordinates of orbit space, the Fokker–Planck equation that governs this diffusion takes a particularly simple form.

The diffusion coefficients $\overline{\Delta}$ and $\overline{\Delta^2}$ of the general Fokker–Planck equation may be obtained either from kinetic theory – that is by supposing that stars move freely along mean-field orbits between abrupt scattering events – or via Hamiltonian perturbation theory from the Fourier transform $\tilde{c}_{\mathbf{n}}(\nu)$ of the autocorrelation function of the scattering potential. In Section 3 we used first-order perturbation theory to obtain a general relation between $\tilde{c}_{\mathbf{n}}(\nu)$ and the diffusion tensor $\overline{\Delta^2}$. In general, derivation by Hamiltonian perturbation theory of the corresponding relation between $\tilde{c}_{\mathbf{n}}$ and the first-order diffusion vector $\overline{\Delta}$ requires a second-order calculation. However in Appendices A and B we derive relationships between $\overline{\Delta}$ and $\overline{\Delta^2}$ that under many astronomically interesting circumstances permit one to calculated $\overline{\Delta}$ from $\overline{\Delta^2}$ without recourse to a tedious second-order calculation. Furthermore, in Section 3.1 we showed that these relationships together with the non-negativity of the quadratic form associated with $\overline{\Delta^2}$, ensure that any H-function yields a non-decreasing entropy for the system.

Two natural fields for the application of the results of Sections 2 and 3 are (i) the evolution of globular clusters, and (ii) the evolution of stellar discs. In Sections 4–6 we used our results to study the secular heating of stellar discs. In Section 4 we show, independently of any hypothesis as to what is responsible for heating of the discs of spiral galaxies, that a population of disc stars will retain a Gaussian distribution of peculiar velocities as it ages and heats if, and only if, the random velocities grow as the square root of time and the diffusion tensor $\overline{\Delta^2}$ scales with epicycle energy $E_{\rm rand}$ as $\overline{\Delta^2} \propto E_{\rm rand}$.

In Section 5 we investigated how a stellar disc heats in consequence of the scattering of stars off molecular clouds on circular orbits. For a razor-thin disc we recover the classical result that the velocities grow as $t^{1/3}$, and we display the self-similar form to which the distribution function of a coeval stellar population in a two-dimensional disc would settle at late times. This proves to be exponential in the cubes of the velocities. In Section 5.3 and Appendix C we extend this analysis to discs of finite thickness. The diffusion tensor is in this case extremely anisotropic, with the consequence that the action-space density of stars rapidly becomes almost uniform along curves that are everywhere tangent to the direction of largest diffusivity. Once this regime is established, the velocity dispersions in a disc with a flat rotation curve satisfy $\sigma_z/\sigma_R \simeq 0.79$. On a time-scale considerably longer than that required to establish this velocity dispersion ratio, the distribution function of a coeval stellar population tends to a self-similar form in which the phase-space density falls off as a Gaussian in the *energies*, and the velocity dispersions grow as $t^{1/4}$.

Since the results of Section 5 accord very ill with observations of the solar neighbourhood, it is natural to enquire to what degree this conflict between theory and observation arises because clouds do not, in fact, scatter stars at a single point on the stellar epicycle, as is assumed in the picture introduced by Spitzer & Schwarzschild (1953). In Appendix D we have used the perturbation-theory results of Section 3 to calculate the diffusion coefficients for cloud scattering in a razor thin disc without making any locality assumption. For values of E_R such that the epicyclic amplitude $a \simeq 1.5\sigma_R/(40\text{ km s}^{-1})$ kpc is larger than the typical cloud size r_c, there is excellent agreement between our new results and those obtained by the method of Spitzer and Schwarzschild. For smaller epicycle amplitudes, however, the two calculations do give materially different answers. Indeed for values of E_R such that $a < r_c$, Appendix D yields a diffusion tensor $\overline{\Delta^2}$ that has exactly the scaling $\overline{\Delta^2} \propto E_R$ which observation seems to require for all values of E_R. However, for a scattering object with scale size $r_c \gtrsim 1.5$ kpc one cannot justify the assumption of Appendix D that the scatterer is spherically symmetric, since an initially spherical object of this size would soon shear out into a spiral arm. Hence from Appendix D we conclude with Lacey (1984) that the Spitzer–Schwarzschild process cannot, unaided, account for the age–velocity–dispersion relationship in the solar neighbourhood, but that there is a *prima facie* case that ephemeral spiral arms might be able to account for the observations.

We find that a sequence of sinusoidal spiral perturbations of effective wavenumber $\mathscr{K}'$ [equation (6.5b)] and having a temporal power spectrum with width Γ in angular frequency generates a diffusion tensor that scales as E_R for $\mathscr{K}' a \lesssim \max(1, \Gamma/\varkappa)$, where a is the epicycle amplitude. It seems unlikely that spiral features can evolve on time-scales shorter than the epicycle period, so $\Gamma \lesssim \varkappa$, and we obtain the desired scaling of the diffusion tensor only if the effective wavelengths $\lambda = 2\pi/\mathscr{K}'$ of the features responsible for the bulk of heating satisfy $\lambda \gtrsim 2\pi a$. For the solar neighbourhood this implies that spiral structure can drive the velocity dispersions of populations as $t^{1/2}$ up to 40 km s^{-1} only if the dominant features are of wavelength $\lambda \gtrsim 9$ kpc. In other words only global spirals could account for the observed age–velocity–dispersion relationship below 40 km s^{-1}. Significant amplitudes of smaller-scale features are excluded on the hypothesis that spiral structure heats the disc, since such features would tend to flatten a plot of σ versus $\sqrt{t}$ at late times.

While the functional form of $\sigma(t)$ is independent of Γ for $\Gamma < \varkappa$, the rate at which σ increases depends sensitively on $\Gamma/\varkappa$ and the dominant pattern speed Ω_p. If $\Gamma \ll \varkappa$, spiral heating is effective only near the Lindblad resonances associated with Ω_p. Thus spiral heating is effective over the whole disc only if $\Gamma \simeq \varkappa$. This could be achieved either by having $\Gamma \simeq \varkappa$ for a single wave, or by having a sequence of waves that covers a range $\delta\Omega_p \simeq \varkappa$ in pattern speeds. The first case seems implausible since one would not expect the disc regularly to stumble into a configuration that is unstable on a dynamical time. On the other hand, a sequence of slowly growing spirals with steadily shifting pattern speed may arise naturally: at each instant the set of the permitted pattern speeds of global spiral patterns would be discrete and the disc would heat rapidly in narrow resonant zones. The disc's resonant frequencies would then shift in response to this heating, and the spirals would start to heat different resonant zones. If these zones were to wander around the disc sufficiently rapidly such frequency modulated global spiral structure would be capable of accounting for the observed increase in σ_R near the Sun. Notice that this picture makes the observationally testable prediction that the growth of σ_R at the Sun should be episodic. On account of the long wavelengths of the spirals needed to account for solar neighbourhood observations, it is exceedingly improbable that spiral structure alone can account for the evolution of the vertical dispersion σ_z.

A simple physical argument explains why neither clouds nor short wavelength spirals can generate diffusion tensor that scales as E_{rand}. We have $\overline{\Delta^2} \propto (\Delta E_R)^2$, so our inference from observations of the solar neighbourhood that $\overline{\Delta^2} \propto E_R$ implies that $\mathbf{v} \cdot \Delta\mathbf{v} \propto \Delta E_R \propto \sqrt{E_R} \propto |\mathbf{v}|$, and thus that $\Delta\mathbf{v}$ is independent of $|\mathbf{v}|$. Compact scatterers such as clouds are ruled out because they imply $\Delta\mathbf{v} \propto 1/|\mathbf{v}|$. Spiral waves are ruled out because a wave excites only those stars which are in resonance with one of the frequencies generated by Doppler-shifting the wave to the star's rest-frame; for a fast-moving star these frequencies are inconveniently high, so the $\Delta\mathbf{v}$ generated by a wave is a steeply declining function of $|\mathbf{v}|$. What *can* produce changes $\Delta\mathbf{v}$ that are independent of

$|\mathbf{v}|$ is any potential perturbation that lasts significantly less than an epicycle period: for such a short-lived perturbation, the impulse approximation is valid and $\Delta\mathbf{v}$ is automatically independent of $|\mathbf{v}|$. The structures that give rise to these potential perturbations must move so fast that they are unlikely to be confined to the disc. Two possible candidates are (i) a massive halo object (Lacey & Ostriker 1985) and (ii) dwarf galaxies in the process of tidal disruption by our Galaxy (Rogers, Harding & Sadler 1981; Lance 1986). The existence of numerous high-velocity star streams near the Sun (Eggen 1977) strongly suggests that a significant proportion of the hottest portion of the disc has been contributed by now defunct satellites, rather than built up by accelerating stars formed at rest in the plane. However, it is not immediately apparent that this picture can account either for the approximately radius-independent scale heights observed in external galaxies (van der Kruit & Searle 1981, 1982) or the ratio $\sigma_z/\sigma_R \simeq 0.6$ measured for several solar neighbourhood populations. None the less, whatever agency is responsible for increasing the velocity dispersions in discs, the work of Section 5.3 suggests that scattering by clouds should be able to enforce a ratio $\sigma_z/\sigma_R \approx 0.8$ at least up to dispersions $\sigma_z \approx 20\ \mathrm{km\ s^{-1}}$.

Acknowledgments

We thank S. Balbus, J. Goodman, A. Toomre and S. D. Tremaine for stimulating discussions and R. Carlberg for comments on a draft of the paper. This work commenced while JJB was a Fairchild Scholar at the California Institute of Technology, which it is a pleasure to thank for its generous hospitality. Subsequent work at Princeton was supported by the NSF under grants PHY-8620266 and AST84-20352 and at Harvard by NASA grant number NAGW-931.

References

Arnold, V. I., 1978. *Mathematical Methods of Classical Mechanics,* Springer-Verlag, New York.
Bahcall, J. N., 1984. *Astrophys. J.,* **287,** 926.
Barbanis, B. & Woltjer, L., 1967. *Astrophys. J.,* **150,** 461.
Binney, J. J. & Tremaine, S. D., 1987. *Galactic Dynamics*, Princeton University Press.
Carlberg, R. G., 1984. In: *Formation and Evolution of Galaxies and Large Structures in the Universe*, p. 343, eds. Audouze, J., & van Tran Thanh, J., Reidel, Dordrecht, Holland.
Carlberg, R. G., 1987. *Astrophys. J.,* submitted.
Carlberg, R. G. & Sellwood, J. A., 1985. *Astrophys. J.,* **292,** 79.
Chandrasekhar, S., 1942. *Principles of Stellar Dynamics*, University of Chicago Press (reissued by Dover 1960).
Cohen, R. S., Grabelsky, D. A., May, J., Bronfman, L., Alvarez, H. & Thaddeus, P., 1985. *Astrophys J.,* **290,** L15.
Eggen, O. J., 1977. *Astrophys. J.,* **215,** 812.
Goodman, J., 1983. *PhD thesis*, Princeton University.
Goodman, J. & Hut, P. (eds), 1985. *Dynamics of Star Clusters, IAU Symp. No. 113*, Reidel, Dordrecht, Holland.
Gradshteyn, I. S. & Ryzhik, I. M., 1965. *Tables of Integrals, Series and Products*, Academic Press, New York.
Icke, V., 1982. *Astrophys. J.,* **254,** 517.
Julian, W. H., 1967. *Astrophys. J.,* **148,** 175.
Julian, W. H. & Toomre, A., 1966. *Astrophys. J.,* **146,** 810.
Lacey, C. G., 1984. *Mon. Not. R. astr. Soc.,* **208,** 687.
Lacey, C. G. & Ostriker, J. P., 1985. *Astrophys. J.,* **299,** 633.
Lance, K., 1986. *PhD thesis*, Australian National University.
Lynden-Bell, D., 1962. *Mon. Not. R. astr. Soc.,* **124,** 1.
Lynden-Bell, D. & Kalnajs, A. J., 1972. *Mon. Not. R. astr. Soc.,* **157,** 1.
Rogers, A., Harding, P. & Sadler, E., 1981. *Astrophys. J.,* **244,** 912.
Rosenbluth, M. N., MacDonald, W. M. & Judd, D. L., 1957. *Phys. Rev.,* **107,** 1.
Schwarzschild, K., 1907. *Göttingen Nachr.,* p. 614.
Sellwood, J. A. & Carlberg, R. G., 1984. *Astrophys. J.,* **282,** 61.
Spitzer, L. & Schwarzschild, M., 1951. *Astrophys. J.,* **114,** 385.
Spitzer, L. & Schwarzschild, M., 1953. *Astrophys. J.,* **118,** 106.

Tremaine, S., Hénon, M. & Lynden-Bell, D., 1986. *Mon. Not. R. astr. Soc.*, **219**, 285.
van der Kruit, P. C. & Searle, L., 1981. *Astr. Astrophys.*, **95**, 105.
van der Kruit, P. C. & Searle, L., 1982. *Astr. Astrophys.*, **110**, 61.
Villumsen, J. V., 1985. *Astrophys. J.*, **290**, 75.
Wielen, R., 1977. *Astr. Astrophys.*, **60**, 263.
Wielen, R. & Fuchs, B., 1986. In: *The Milky Way Galaxy, IAU Symp. No. 106*, p. 481, eds van Woerden, H., Allen, R. J. & Burton, W. B., Reidel, Dordrecht, Holland.

Appendix A: first-order diffusion coefficients

The second-order change in the actions Δ_2 is the difference between the integral of Hamilton's equations (3.3) along the first-order and the zeroth-order trajectories. Thus

$$\Delta_2(t) = -\int_0^t \left[\left(\Delta_1\cdot\frac{\partial}{\partial \mathbf{J}}+\boldsymbol{\theta}_1\cdot\frac{\partial}{\partial\boldsymbol{\theta}}\right)\frac{\partial\Phi_1}{\partial\boldsymbol{\theta}}\right]_{(\boldsymbol{\theta}=\boldsymbol{\theta}_0+\boldsymbol{\omega}_0 t,\,\mathbf{J}=\mathbf{J}_0)} dt'$$

$$= -i\sum_{\mathbf{n}} \mathbf{n}\int_0^t dt'\left[\Delta_1\cdot\frac{\partial\Psi_{\mathbf{n}}}{\partial\mathbf{J}_0}\bigg|_{t'} + i(\mathbf{n}\cdot\boldsymbol{\theta}_1)\,\Psi_{\mathbf{n}}(\mathbf{J}_0,t')\right]\exp\left[i\mathbf{n}\cdot(\boldsymbol{\theta}_0+\boldsymbol{\omega}_0 t')\right]. \tag{A1}$$

The first-order perturbation in $\boldsymbol{\theta}$ occurring in this equation may be calculated by observing that

$$\dot{\boldsymbol{\theta}} = \frac{\partial H}{\partial\mathbf{J}} = \boldsymbol{\omega}(\mathbf{J}) + \sum_{\mathbf{n}}\frac{\partial\Psi_{\mathbf{n}}}{\partial\mathbf{J}}\bigg|_t \exp(i\mathbf{n}\cdot\boldsymbol{\theta}). \tag{A2}$$

Subtracting the unperturbed equation of motion and integrating with respect to time, we have

$$\boldsymbol{\theta}_1(t) = \int_0^t\left\{\Delta_1\cdot\frac{\partial\boldsymbol{\omega}}{\partial\mathbf{J}_0} + \sum_{\mathbf{n}}\frac{\partial\Psi_{\mathbf{n}}}{\partial\mathbf{J}_0}\bigg|_{t'}\exp\left[i\mathbf{n}\cdot(\boldsymbol{\theta}_0+\boldsymbol{\omega}_0 t')\right]\right\}dt'$$

$$= -i\sum_{\mathbf{n}}\mathbf{n}\cdot\frac{\partial\boldsymbol{\omega}}{\partial\mathbf{J}_0}\int_0^t dt'\int_0^{t'} dt''\,\Psi_{\mathbf{n}}(\mathbf{J}_0,t'')\exp\left[i\mathbf{n}\cdot(\theta_0+\boldsymbol{\omega}_0 t'')\right]$$

$$+\sum_{\mathbf{n}}\int_0^t\frac{\partial\Psi_{\mathbf{n}}}{\partial\mathbf{J}_0}\bigg|_{t'}\exp\left[i\mathbf{n}\cdot(\boldsymbol{\theta}_0+\boldsymbol{\omega}_0 t')\right]dt'. \tag{A3}$$

Substituting equations (3.5) and (A3) into equation (A1) and averaging over the initial phase, we find

$$\langle\Delta_{2i}(T)\rangle_{\theta} = \sum_{\mathbf{n}} n_i\int_0^T dt\left\{\int_0^t dt'\,\mathbf{n}\cdot\frac{\partial}{\partial\mathbf{J}}\left[\Psi_{\mathbf{n}}(\mathbf{J}_0,t)\,\Psi^*_{\mathbf{n}}(\mathbf{J}_0,t')\right]\exp\left[i\mathbf{n}\cdot\boldsymbol{\omega}_0(t-t')\right]\right.$$

$$\left.+\mathbf{n}\cdot\frac{\partial}{\partial\mathbf{J}_0}(i\mathbf{n}\cdot\boldsymbol{\omega})\int_0^{t'} dt'\int_0^{t'} dt''\,\Psi_n(\mathbf{J}_0,t)\,\Psi^*_{\mathbf{n}}(\mathbf{J}_0,t'')\exp\left[i\mathbf{n}\cdot\boldsymbol{\omega}_0(t-t'')\right]\right\}. \tag{A4}$$

When we interchange the order of the t' and t'' integrations in the second term and perform the t' integral, we find that we can combine both terms to obtain

$$\langle\Delta_{2i}(T)\rangle_{\theta} = \frac{\partial}{\partial J_j}\sum_{\mathbf{n}} n_i n_j\int_0^T dt\int_0^t dt'\,\Psi_{\mathbf{n}}(\mathbf{J}_0,t)\,\Psi^*_{\mathbf{n}}(\mathbf{J}_0,t')\exp\left[i\mathbf{n}\cdot\boldsymbol{\omega}_0(t-t')\right). \tag{A5}$$

We carry the summation over **n** inside the integrals, and observe that the integrand is then manifestly invariant under the interchange $\mathbf{n} \rightarrow -\mathbf{n}$. However, with equation (3.1b) we have that the effect of this interchange is to exchange the roles of t and t' in the integrand. Consequently the specified integral over the portion of the (t, t') plane with $0 \leqslant t' \leqslant t$, $0 \leqslant t \leqslant T$ is equal to half the value of the integral of the same integrand over the entire square $0 \leqslant t \leqslant T$, $0 \leqslant t' \leqslant T$. But equation (3.6) shows that this integral is equal to $\langle \Delta_{1i} \Delta_{1j}(T) \rangle_\theta$. Hence to second-order accuracy

$$\frac{1}{2} \frac{\partial}{\partial J_j} \langle \Delta_{1i} \Delta_{1j}(T) \rangle_\theta = \langle \Delta_{2i}(T) \rangle_\theta = \langle \Delta_i(T) \rangle_\theta, \tag{A6}$$

where the second equality follows because from equation (3.5) we have that

$$\langle \Delta_1 \rangle_\theta = 0. \tag{A7}$$

Taking the ensemble average of both sides, and letting T become large, equation (3.10) now follows.

Appendix B: diffusion coefficients generated by recoiling scatterers

As Chandrasekhar (1942) has shown, the first- and second-order diffusion coefficients associated with a stellar population in thermal equilibrium are related in such a way that the stochastic acceleration of members of the population that is described by $\overline{\Delta^2}$ is balanced by the dynamical friction drag associated with $\overline{\Delta}$. It is interesting to derive this relationship from equation (2.7a).

Equation (2.7a) states that the rate of growth of f is equal to minus the divergence of the action-space flux

$$S_i \equiv -f \overline{\Delta_i} + \frac{1}{2} \frac{\partial f \overline{\Delta_{ij}^2}}{\partial J_j}. \tag{B1}$$

Now if the scattering objects have reached non-rotating thermal equilibrium, the equilibrium distribution function of the scattered stars will be of Gibbs' form

$$f_0 \propto \exp(-\beta H), \tag{B2}$$

where β is the inverse temperature of the scattering objects and H is the Hamiltonian governing the motions of individual stars and, by the principle of detailed balance, the flux **S** in equation (B1) will vanish. Substituting equation (B2) into equation (B1) with $\mathbf{S} = 0$, using equations (3.2) and dividing through by f_0, one obtains equation (3.12). Since the diffusion coefficients depend only on the distribution of scatterers, equation (3.12) must be satisfied independently of whether the scattered stars are, in fact, in thermal equilibrium with the scatterers.

In practical applications the scatterers may be neither thermally distributed nor so massive that it is obvious that their gravitational field may be assumed to be independent of the motion of the scattered stars, as was assumed in Section 3.1. A derivation of equation (3.10) from time-reversibility and kinetic theory will elucidate whether equation (3.10) is satisfied in such a case.

Let the phase-space coordinates $(x_1, \ldots, x_N, v_1 \ldots, v_N)$, where N is three times the number of particles in the system of scatterers, completely define the dynamical state of the system before it scatters a particular star in a given way from actions **J** to actions $\mathbf{J} + \mathbf{\Delta}$, and let : $x'_1, \ldots, x'_N, v'_1, \ldots, v'_N)$ define the state of the scattering system after the star has been scattered. Then in a time-reversed model of the entire stellar system, each such scattering will be associated with a scattering $(\mathbf{J} + \mathbf{\Delta}, \mathbf{x}', -\mathbf{v}') \rightarrow (\mathbf{J}, \mathbf{x}, -\mathbf{v})$. Of course the time-reversed system is distinguishable from its original because it will counter-rotate. However, if the original system has a symmetry plane containing its spin axis (and thus bears no spiral pattern), it will be statistically indistinguishable from a mirror image of the time-reversed system, although every transition $(\mathbf{J}, \mathbf{x}, \mathbf{v}) \to (\mathbf{J}+\boldsymbol{\Delta}, \mathbf{x}', \mathbf{v}')$ that occurs in the original will be accompanied by a transition $(\mathbf{J}+\boldsymbol{\Delta}, -\mathbf{x}', \mathbf{v}') \to (\mathbf{J}, -\mathbf{x}, \mathbf{v}')$ its time-reversed and mirrored double. Thus we may assert that the probabilities per unit time per unit volume of $2N$-dimensional phase space that a given star will take part in such transitions in either system are equal:

$$\dot{p}(\mathbf{J}+\boldsymbol{\Delta}, -\boldsymbol{\Delta}, -\mathbf{x}', \mathbf{v}', -\mathbf{x}, \mathbf{v}) = \dot{p}(\mathbf{J}, \boldsymbol{\Delta}, \mathbf{x}, \mathbf{v}, \mathbf{x}', \mathbf{v}'). \tag{B3}$$

Now the rates $\dot{P}$ entering equation (2.7b) are related to these rates by

$$\dot{P}(\mathbf{J}, \boldsymbol{\Delta}) = \int F(\mathbf{x}, \mathbf{v})\, \dot{p}(\mathbf{J}, \boldsymbol{\Delta}, \mathbf{x}, \mathbf{v}, \mathbf{x}', \mathbf{v}')\, d^N\mathbf{x}\, d^N\mathbf{v}\, d^N\mathbf{x}'\, d^N\mathbf{v}',$$

$$\dot{P}(\mathbf{J}+\boldsymbol{\Delta}, -\boldsymbol{\Delta}) = \int F(-\mathbf{x}', \mathbf{v}')\, \dot{p}(\mathbf{J}+\boldsymbol{\Delta}, -\boldsymbol{\Delta}, -\mathbf{x}', \mathbf{v}', -\mathbf{x}, \mathbf{v})\, d^N\mathbf{x}'\, d^N\mathbf{v}'\, d^N\mathbf{x}\, d^N\mathbf{v}, \tag{B4}$$

where $F(\mathbf{x}, \mathbf{v})\, d^N\mathbf{x}\, d^N\mathbf{v}$ is the probability that the scattering system will have coordinates in $d^N\mathbf{x}\, d^N\mathbf{v}$. Now if the scattering system is very massive, its coordinates $(\mathbf{x}, \mathbf{v})$ will be little changed by the scattering event, so unless the system is very cold in the sense that its phase-space density F is an extremely rapidly varying function of the phase-space coordinates, we have for massive scatterers $F(\mathbf{x}, \mathbf{v}) \simeq F(\mathbf{x}', \mathbf{v}')$, and in such a case it follows from equation (B4) that

$$\dot{P}(\mathbf{J}, \boldsymbol{\Delta}) = \dot{P}(\mathbf{J}+\boldsymbol{\Delta}, -\boldsymbol{\Delta}), \tag{B5}$$

which allows (2.7b) to be rewritten

$$\begin{aligned} \overline{\Delta_i} \equiv \int \Delta_i \dot{P}(\mathbf{J}, \boldsymbol{\Delta})\, d^3\boldsymbol{\Delta} &= \int \Delta_i \dot{P}(\mathbf{J}+\boldsymbol{\Delta}, -\boldsymbol{\Delta})\, d^3\boldsymbol{\Delta} \\ &= \int \Delta_i \left[\dot{P}(\mathbf{J}, -\boldsymbol{\Delta}) + \Delta_j \frac{\partial \dot{P}}{\partial J_j} + \dots \right] d^3\boldsymbol{\Delta} \\ &= -\overline{\Delta_i} + \frac{\partial \overline{\Delta^2_{ij}}}{\partial J_j} + \mathbf{O}(\overline{\Delta^3_i}). \end{aligned} \tag{B6}$$

This is equivalent to equation (3.10).

Appendix C: three-dimensional Spitzer–Schwarzschild diffusion

Since $E_R = \tfrac{1}{2}(v_R^2 + \gamma^2 v_\phi^2)$ and $E_z = \tfrac{1}{2}(v_z^2 + \nu^2 z^2)$, we have

$$\begin{aligned} \Delta E_R &= v_R \Delta v_R + \tfrac{1}{2}(\Delta v_R)^2 + \gamma^2[v_\phi \Delta v_\phi + \tfrac{1}{2}(\Delta v_\phi)^2] \\ \Delta E_z &= v_z \Delta v_z + \tfrac{1}{2}(\Delta v_z)^2. \end{aligned} \tag{C1}$$

Hence

$$\begin{aligned} (\Delta E_R)^2 &= v_R^2(\Delta v_R)^2 + 2\gamma^2 v_R v_\phi \Delta v_R \Delta v_\phi + \gamma^4 v_\phi^2 (\Delta v)^2 + \mathbf{O}[(\Delta v)^3] \\ \Delta E_R \Delta E_z &= v_R v_z \Delta v_R \Delta v_z + \gamma^2 v_\phi v_z \Delta v_\phi \Delta v_z + \mathbf{O}[(\Delta v)^3] \\ (\Delta E_z)^2 &= v_z^2 (\Delta v_z)^2 + \mathbf{O}[(\Delta v)^3]. \end{aligned} \tag{C2}$$

Let $\hat{\mathbf{a}}$ be a unit vector parallel to $\mathbf{v}$, and let $\hat{\mathbf{b}}$ and $\hat{\mathbf{c}}$ be any two mutually perpendicular unit vectors in the plane normal to $\hat{\mathbf{a}}$. We write

$$\Delta \mathbf{v} = \Delta v_\parallel \hat{\mathbf{a}} + \Delta v_{\perp b} \hat{\mathbf{b}} + \Delta v_{\perp c} \hat{\mathbf{c}}. \tag{C3}$$

We have to leading order in $\ln\Lambda$, for clouds at rest in the LSR.

$$\langle\Delta v_\parallel \Delta v_{\perp b}\rangle = \langle v_\parallel \Delta v_{\perp c}\rangle = \langle \Delta v_{\perp b}\Delta v_{\perp c}\rangle = 0,$$

$$\langle(\Delta v_{\perp b})^2\rangle = \langle(\Delta v_{\perp c})^2\rangle \equiv \tfrac{1}{2}\langle(\Delta v_\perp)^2\rangle = \frac{4\pi G^2 n_c M_c^2 \ln\Lambda}{v},$$

$$\langle(\Delta v_\parallel)^2\rangle = 0. \tag{C4}$$

Thus

$$\langle \Delta v_i \Delta v_j\rangle = \langle(\Delta v_\parallel)^2\rangle\, \hat{a}_i\hat{a}_j + \tfrac{1}{2}\langle(\Delta v_\perp)^2\rangle(\hat{b}_i\hat{b}_j + \hat{c}_i\hat{c}_j). \tag{C5}$$

But $\hat{a}_i\hat{a}_j + \hat{b}_i\hat{b}_j + \hat{c}_i\hat{c}_j = \delta_{ij}$ and $\hat{a}_i = \hat{v}_i/v$, so

$$\langle \Delta v_i \Delta v_j\rangle = \langle(\Delta v_\parallel)^2\rangle \frac{v_i v_j}{v^2} + \tfrac{1}{2}\langle(\Delta v_\perp)^2\rangle\left(\delta_{ij} - \frac{v_i v_j}{v^2}\right)$$

$$\simeq \tfrac{1}{2}\langle(\Delta v_\perp)^2\rangle\left(\delta_{ij} - \frac{v_i v_j}{v^2}\right). \tag{C6}$$

Substituting from equation (C6) into equations (C2), we obtain

$$\langle(\Delta E_R)^2\rangle = \frac{1}{2}\left[(v_R^2 + \gamma^4 v_\phi^2) - \frac{(v_R^2+\gamma^2 v_\phi^2)^2}{v^2}\right]\langle(\Delta v_\perp)^2\rangle$$

$$\langle \Delta E_R \Delta E_z\rangle = -\frac{1}{2}\frac{v_z^2(v_R^2+\gamma^2 v_\phi^2)}{v^2}\langle(\Delta v_\perp)^2\rangle$$

$$\langle(\Delta E_z)^2\rangle = \frac{1}{2}v_z^2\left(1 - \frac{v_z^2}{v^2}\right)\langle(\Delta v_\perp)^2\rangle. \tag{C7}$$

We assume that all encounters occur as stars cross the plane $z=0$; at such moments, $v_z = \sqrt{2E_z}$. The fraction of each orbital period spent in a cloud layer of half-thickness h_c is then $2\nu h_c/(\pi\sqrt{2E_z})$, so defining

$$N_c \equiv 2h_c n_c; \quad C \equiv \frac{8}{\pi} G^2 N_c M_c^2 \nu \ln\Lambda, \tag{C8}$$

the z-averaged diffusion coefficients are

$$\langle(\Delta E_R)^2\rangle_l = \frac{\pi C}{2\sqrt{2E_z}v}\left[(v_R^2+\gamma^4 v_\phi^2) - \frac{4E_R^2}{v^2}\right]$$

$$\langle \Delta E_R \Delta E_z\rangle_l = -\frac{\pi C}{2\sqrt{2E_z}v}\frac{4E_z E_R}{v^2}$$

$$\langle(\Delta E_z)^2\rangle_l = \frac{\pi C}{2\sqrt{2E_z}v} 2E_z\left(1 - \frac{2E_z}{v^2}\right). \tag{C9}$$

Since $v_R = -\sqrt{2E_R}\sin\theta_r$ and $v_\phi = -(\sqrt{2E_R}/\gamma)\cos\theta_r$, stars cross the plane with speed v given by,

$$v^2 = v_R^2 + v_\phi^2 + v_z^2 = 2(E_R+E_z)(1-k^2\cos^2\theta_r), \tag{C10a}$$

where

$$k^2 \equiv \frac{1-\gamma^{-2}}{1+E_z/E_R}. \tag{C10b}$$

Substituting these expressions into equations (C9) and averaging over θ_r, we obtain, finally, the independent elements of the diffusion tensor $D_{RR} = \langle(\Delta E_R)^2\rangle_{rl}$ etc. as

$$D_{RR} = C\sqrt{\frac{E_R}{E_z}}\frac{k}{\sqrt{1-\gamma^{-2}}}\left[\int_0^{\pi/2}\frac{1+(\gamma^2-1)\cos^2\theta_r}{\sqrt{1-k^2\cos^2\theta_r}}\,d\theta_r - \frac{k^2}{1-\gamma^{-2}}\int_0^{\pi/2}\frac{d\theta_r}{(1-k^2\cos^2\theta_r)^{3/2}}\right]$$

$$= C\sqrt{\frac{E_R}{E_z}}\frac{k}{\sqrt{1-\gamma^{-2}}}\left\{K(k)+\frac{\gamma^2-1}{k^2}[K(k)-E(k)] - \frac{k^2}{1-\gamma^{-2}}\frac{E(k)}{1-k^2}\right\} \quad \text{(C11a)}$$

$$D_{Rz} = -C\sqrt{\frac{E_z}{E_R}}\frac{k^3}{(1-\gamma^{-2})^{3/2}}\int_0^{\pi/2}\frac{d\theta_r}{(1-k^2\cos^2\theta_r)^{3/2}}$$

$$= -C\sqrt{\frac{E_z}{E_R}}\frac{k^3}{(1-\gamma^{-2})^{3/2}}\frac{E(k)}{1-k^2} \quad \text{(C11b)}$$

$$D_{zz} = C\sqrt{\frac{E_z}{E_R}}\frac{k}{\sqrt{1-\gamma^{-2}}}\left[\int_0^{\pi/2}\frac{d\theta_r}{\sqrt{1-k^2\cos^2\theta_r}} - \frac{E_z}{E_R}\frac{k^2}{1-\gamma^{-2}}\int_0^{\pi/2}\frac{d\theta_r}{(1-k^2\cos^2\theta_r)^{3/2}}\right]$$

$$= C\sqrt{\frac{E_z}{E_R}}\frac{k}{\sqrt{1-\gamma^{-2}}}\left[K(k) - \frac{E_R}{E_z}\frac{k^2}{1-\gamma^{-2}}\frac{E(k)}{1-k^2}\right]. \quad \text{(C11c)}$$

Appendix D: scattering by clouds in action-angle coordinates

We calculate the diffusion tensor for a star with guiding-centre radius R_g and epicyclic amplitude a. Let ϕ be the usual azimuthal coordinate, and define the locally Cartesian coordinates

$$x \equiv R - R_g; \quad y \equiv R_g\phi. \quad \text{(D1)}$$

The clouds are on circular orbits, whose coordinates $[x_c(t), y_c(t)]$ satisfy

$$x_c = x_0; \quad y_c = y_0 + (R_g\Omega - 2Ax_0)\,t, \quad \text{(D2)}$$

where x_0, y_0, Ω and A are all constants. Let the Fourier transform of the potential of a cloud centred on the origin be

$$\mathbf{\Phi}^{(c)}_{pq} = \frac{1}{\mathscr{L}^2}\iint_{-\infty}^{\infty}\mathbf{\Phi}^{(c)}(x, y)\exp[-i\mathscr{K}_0(px+qy)]\,dx\,dy, \quad \text{(D3)}$$

where $\mathscr{L}$ is an arbitrary constant and $\mathscr{K}_0 \equiv 2\pi/\mathscr{L}$. Then for convenience we work with the potential

$$\mathbf{\Phi}(x, y, t) = \sum_{p,q}\mathbf{\Phi}_{pq}(t)\exp[i\mathscr{K}_0(px+qy)], \quad \text{(D4a)}$$

where the sum is over integer values of p and q and

$$\mathbf{\Phi}_{pq}(t) = \mathbf{\Phi}^{(c)}_{pq}\exp\{-i\mathscr{K}_0[px_c(t)+qy_c(t)]$$

$$= \mathbf{\Phi}^{(c)}_{pq}\exp[\![-i\mathscr{K}_0\{px_0+q[y_0+(R_g\Omega-2Ax_0)t]\}]\!]. \quad \text{(D4b)}$$

This is the potential of a periodic array of clouds with centres at $(x_c+i\mathscr{L}, y_c+j\mathscr{L})$, with i and j integers. We choose $\mathscr{L}$ such that

$$M \equiv \frac{2\pi R_g}{\mathscr{L}} \quad \text{(D5)}$$

is an integer. The potential (D4a) is a sum of spirals of the form (6.1) with $k=\mathscr{K}_0 p$ and $m=\mathscr{K}_0 R_g q=Mq$. Thus, if we rewrite it in terms of the star's action-angle coordinates, the coefficient of $\exp[i(l\theta_r+m\theta_\phi)]$ is

$$\Psi_{l(qM)}(t)=\sum_p \Phi_{pq}(t)\exp(il\alpha)\,J_l(\mathscr{K}'a), \tag{D6}$$

where α and $\mathscr{K}'$ are defined by equations (6.5b).

Equation (3.9a) relates the desired diffusion coefficients to the temporal power spectrum of the $\Psi_{l(qM)}$. The Fourier transform with respect to time of $\Phi_{pq}(t)$ is

$$\tilde{\Phi}_{pq}(\nu)=2\pi\Phi^{(c)}_{pq}\delta[\nu-q\mathscr{K}_0(R_g\Omega-2Ax_0)]\exp[-i\mathscr{K}_0(px_0+qy_0)]. \tag{D7}$$

Thus

$$\tilde{\Phi}_{pq}(\nu)\,\tilde{\Phi}^*_{p'q}(\nu')=\frac{(2\pi)^2}{2A|q|\mathscr{K}_0}\Phi^{(c)}_{pq}\Phi^{(c)*}_{p'q}\,\delta(\nu-\nu')$$

$$\times\,\delta\left(x_0+\frac{\nu-qM\Omega}{2Aq\mathscr{K}_0}\right)\exp[-i\mathscr{K}_0x_0(p-p')]\qquad(q\neq 0). \tag{D8}$$

Averaging over all possible initial cloud positions (x_0, y_0), with $|x_0|$ and $|y_0|$ in the interval $(0,\tfrac{1}{2}\mathscr{L})$, this yields

$$\overline{\tilde{\Phi}_{pq}(\nu)\,\tilde{\Phi}^*_{p'q}(\nu')}=\frac{1}{2A|q|}\Phi^{(c)}_{pq}\Phi^{(c)*}_{p'q}\,\theta\left(1-\left|\frac{\nu-qM\Omega}{2\pi Aq}\right|\right).$$

$$\times\exp\left[i(p-p')\frac{\nu-qM\Omega}{2Aq}\right]2\pi\delta(\nu-\nu'),\qquad(q\neq 0). \tag{D9}$$

Therefore

$$\tilde{c}_{l(qM)}(J_r,L_\phi,\nu)=\frac{\overline{\tilde{\Psi}_{l(qM)}(\nu)\,\tilde{\Psi}^*_{l(qM)}(\nu')}}{2\pi\delta(\nu-\nu')}$$

$$=\frac{1}{2\pi\delta(\nu-\nu')}\sum_{pp'}\overline{\tilde{\Phi}_{pq}(\nu)\,\tilde{\Phi}^*_{p'q}(\nu')}$$

$$\times\exp\{il[\alpha(p)-\alpha(p')]\}\,J_l[\mathscr{K}'(p)\,a]\,J_l[\mathscr{K}'(p')\,a]$$

$$=\frac{1}{2A|q|}\theta\left(1-\left|\frac{\nu-qM\Omega}{2\pi Aq}\right|\right)$$

$$\times\left|\sum_p\Phi^{(c)}_{pq}\exp\left[ip\left(\frac{\nu-qM\Omega}{2Aq}\right)\right]\exp(il\alpha)\,J_l(\mathscr{K}'a)\right|^2\qquad(q\neq 0). \tag{D10a}$$

Similarly

$$\tilde{c}_{l0}(\nu)=2\pi\delta(\nu)\sum_p|\Phi^{(c)}_{p0}|^2\,J_l^2(\mathscr{K}'a). \tag{D10b}$$

The spatial Fourier transform of the potential of a spherical cloud of mass M_c and scale length r_c is

$$\Phi^{(c)}_{pq} = \frac{2\pi}{\mathscr{L}^2}\int_0^\infty \Phi^{(c)}(r)\, J_0(\mathscr{K}r)\, r\, dr$$

$$= -\frac{GM_c}{\mathscr{L}}\frac{f_c(\mathscr{K}r_c)}{\mathscr{K}/\mathscr{K}_0}, \tag{D11a}$$

where the form factor $f_c(x)\to 1$ as $x\to 0$ and $f_c(x)\to 0$ for $x\to\infty$. For example, the Plummer potential, $\Phi^{(c)}(r) = -GM_c/\sqrt{r^2+r_c^2}$, yields $f_c(x)=\exp(-x)$.

The power spectrum of $N_c\mathscr{L}^2$ randomly distributed clouds per cell is simply $N_c\mathscr{L}^2$ times that of a single cloud. Hence the power spectrum of such an ensemble is

$$\tilde{c}_{l(qM)}(l\varkappa + qM\Omega) = \frac{G^2M_c^2N_c}{2A}\frac{|P_{lq}|^2}{|q|}\;\theta\left(1-\frac{|l|}{|q|}\frac{\varkappa}{2\pi A}\right) \quad (q\neq 0), \tag{D12a}$$

where

$$P_{lq} \equiv \sum_{p=-\infty}^{\infty}\frac{f_c(\mathscr{K}r_c)}{\mathscr{K}/\mathscr{K}_0}\cos(l\psi)\, J_l(\mathscr{K}'a),$$

$$\psi \equiv \frac{\varkappa}{2A}\frac{p}{q} - \arctan\left(\frac{p}{\gamma q}\right) \quad \text{and} \quad \mathscr{K}' = \sqrt{p^2+\gamma^2q^2}\,\mathscr{K}_0. \tag{D12b}$$

Since $|P_{-l,q}| = |P_{l,-q}| = |P_{lq}|$, we have $\tilde{c}_{-l,qM} = \tilde{c}_{l,-qM} = \tilde{c}_{l,qM}$. Substituting equations (D10) and (D12) into equation (3.9) and using these identities to simplify the summations, we have

$$\overline{\Delta^2_{rr}} = \sum_{l=-\infty}^{\infty}\sum_{q=-\infty}^{\infty} l^2\tilde{c}_{l,qM}(l\varkappa + qM\Omega)$$

$$= \frac{G^2M_c^2N_c}{A}\sum_{q=1}^{\infty}\frac{2}{q}\sum_{l=1}^{\eta q} l^2|P_{lq}|^2$$

$$\overline{\Delta^2_{r\phi}} = \sum_{l=-\infty}^{\infty}\sum_{q=-\infty}^{\infty} l(qM)\,\tilde{c}_{l,qM}(l\varkappa + qM\Omega)$$

$$= 0 \tag{D13a}$$

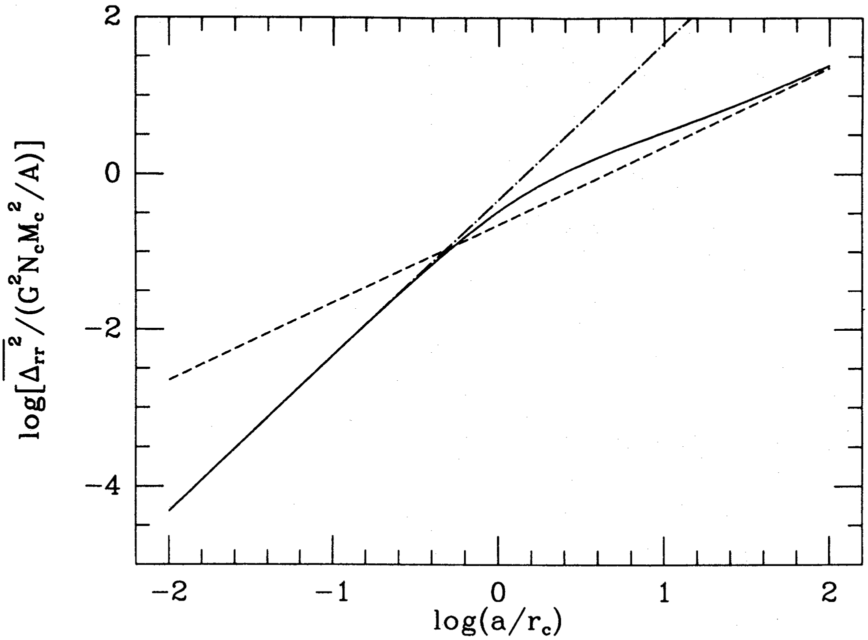
$$\overline{\Delta^2_{\phi\phi}} = \sum_{l=-\infty}^{\infty}\sum_{q=-\infty}^{\infty} (qM)^2\,\tilde{c}_{l,qM}(l\varkappa + qM\Omega)$$

$$= \frac{G^2M_c^2N_c}{A}\mathscr{K}_0^2R_g^2\sum_{q=1}^{\infty} q\sum_{l=-\eta q}^{\eta q}|P_{lq}|^2,$$

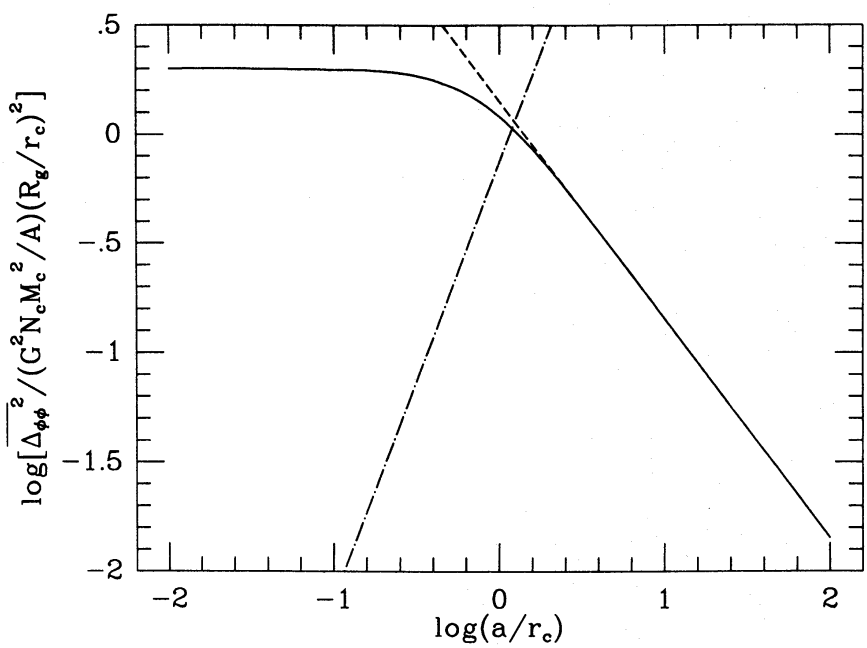
where

$$\eta \equiv \frac{2\pi A}{\varkappa}. \tag{D13b}$$

In Fig. D1 we plot for Plummer-model clouds and v_c = constant numerical estimates of the values of $\overline{\Delta^2_{rr}}$ and $\overline{\Delta^2_{\phi\phi}}$ in the limit $\mathscr{K}_0\to 0$ together with estimates of their asymptotic behaviour in the limits $a\to 0$ and $a\to\infty$. The asymptotes for the limit $a\to\infty$ have been calculated using



(a)



(b)

Figure D1. (a) Diffusion coefficient $\overline{\Delta^2_{rr}}$ as a function of epicyclic amplitude a for two-dimensional Spitzer–Schwarzschild diffusion. Solid curve: equation (D13a). Dashed line: the same quantity calculated with the approximations of Spitzer & Schwarzschild. Dotted line: asymptote for small a calculated by the method of Julian & Toomre (1966). (b) The corresponding curves for $\overline{\Delta^2_{\phi\phi}}$.

the approach of Spitzer & Schwarzschild (1953): $\overline{\Delta^2_{rr}}$ follows from equation (5.10) and we have

$$\overline{\Delta^2_{\phi\phi}} = \frac{2G^2M_c^2N_c}{A}\,\frac{(R_g/r_c)^2}{a/r_c}\,\frac{2\Omega}{\varkappa}\,[K(\sqrt{A/\Omega}) - E(\sqrt{A/\Omega})]. \tag{D14}$$

Both curves are seen to approach these asymptotes as expected. The asymptotes for the limit $a \to 0$ have been calculated by using first-order perturbation theory to follow individual orbits in the (x, y) plane (*cf.* Julian & Toomre 1966). We have

$$\overline{\Delta^2_{rr}} = \frac{2G^2M_c^2N_c}{A}\left(\frac{a}{r_c}\right)^2 G_1(\gamma)$$

$$\overline{\Delta^2_{\phi\phi}} = \frac{2G^2M_c^2N_c}{A}\left(\frac{R_g}{t_c}\right)^2\left(\frac{a}{r_c}\right)^2\frac{\gamma^2}{(\gamma^2-1)^2}G_{-1}(\gamma), \tag{D15a}$$

where

$$G_p(\gamma) \equiv \int_0^\infty \frac{x^p}{(1+x^2)^2}\left[F\left(\frac{\gamma\sqrt{1+x^2}}{(\gamma^2-1)x}\right)\right]^2 dx$$

$$F(y) \equiv y[K_1(y)+(\gamma^2-1)\,yK_0(y)]. \tag{D15b}$$

Here the K_i are modified Bessel functions. Fig. D1 shows that while the value of $\overline{\Delta^2_{rr}}$ obtained from equations (D14a) rapidly approaches the corresponding asymptote of equations (D15a), for small a/r_c the curves for $\overline{\Delta^2_{\phi\phi}}$ do not agree. The origin of this discrepancy lies in the singularity of the angular momentum exchange for encounters at impact parameter x_0 as $x_0 \to 0$. The (x, y) plane calculation proceeds by calculating ΔL_ϕ for complete encounters, and then averaging its square over x_0. Unfortunately, encounters at $x_0 = 0$ are never completed, so strictly speaking, the point $x_0 = 0$ should be excluded from the averaging integral and these encounters treated separately. Of course, a separate treatment would be unnecessary if ΔL_ϕ were finite at $x_0 = 0$. However, the angular momentum exchange between the star and a cloud on an orbit with $x_0 = 0$ diverges as the integration time T as $T \to \infty$. When this divergent quantity is squared, averaged over x_0, divided by T and T is taken to infinity, a finite contribution to $\overline{\Delta^2_{\phi\phi}}$ arises, which the action-angle calculation automatically includes. However this contribution is of little consequence since the angular momentum exchange calculated with perturbation theory is unreliable for all sufficiently small x_0: at late times along such orbits the angular coordinate y differs by an arbitrarily large amount from that along the unperturbed trajectory (e.g. Icke 1982), and thus no convergent perturbation series for $y(t)$ is possible. Fortunately this phenomenon does not materially affect the reliability of the calculation of $\overline{\Delta^2_{rr}}$.