PRAMANA — journal of physics
© Indian Academy of Sciences
Vol. 73, No. 1
July 2009
pp. 193–214

The gravitational dynamics of galaxies

RAJARAM NITYANANDA
National Centre for Radio Astrophysics, Tata Institute of Fundamental Research, Post Bag 3, Ganeshkhind, Pune 411 007, India
E-mail: rajaram@ncra.tifr.res.in

Abstract. The broad area of galactic dynamics is presented for a physics audience, with the requisite astronomy background in outline, and focusing on gravitational effects. The basic underlying model is a large number of particles (which could be stars or dark matter) moving in their self-consistent gravitational potential. The effects of two-particle correlations/scattering, although weak, can be cumulative and hence important for a class of systems such as star clusters which are hence termed collisional. On the larger scale of galaxies, we have collisionless behaviour which is different and in some ways richer. The basic ideas and applications in both these regimes are described, and some issues highlighted in conclusion.

Keywords. Galaxy dynamics; gravitating systems; stellar dynamics.

PACS Nos 98.10+z; 98.52.-b

1. Introduction

1.1 *Scope and plan*

Physicists are usually attracted by the birth of the Universe and the deaths of stars. In contrast, the dynamics of galaxies does not provide the relativistic fireworks of the Big Bang and black holes. But it is a very active area of contemporary astronomy, with its own share of interesting questions and intriguing answers, certainly for astronomers but also from a physicist's point of view. This survey article introduces galactic dynamics to readers from outside astronomy with some background from classical mechanics and statistical physics. In spite of the restriction to processes dominated by Newtonian gravity, there is a wide variety of topics which come under its scope. Astronomical observations naturally form the underpinning of the subject and are therefore introduced at the level of broad trends – the practitioner of galactic astronomy will certainly find this part oversimplified.

The core of the article goes over the dynamics of star clusters and galaxies. Both come under the umbrella of the Newtonian N-body problem for large N, but there are important differences. A cluster of even a million stars is 'collisional' – there are processes analogous to those which establish a uniform density and a Maxwellian

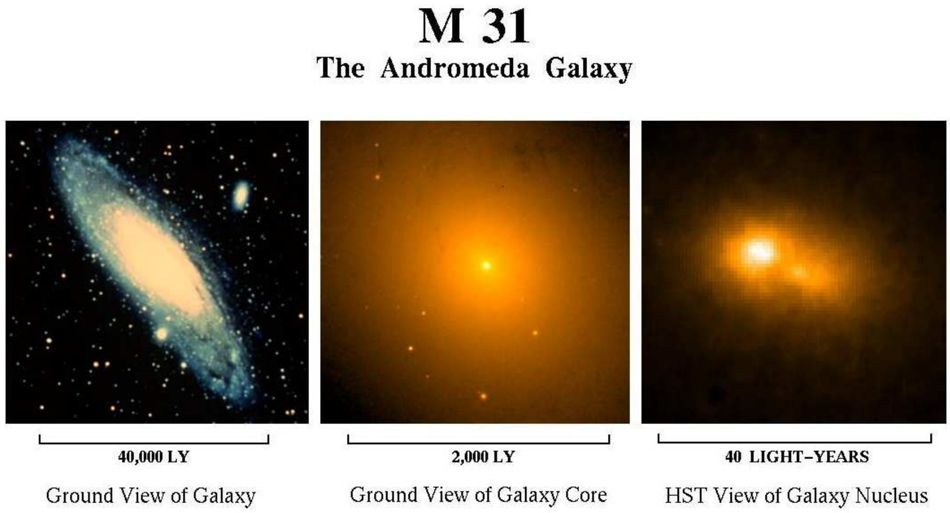

Figure 1. The Andromeda galaxy, our nearest neighbour, viewed at three different scales (courtesy the Hubble archive). Note the satellites on the largest scale. On the smallest scale, one sees the dense star cluster whose motions reveal the central black hole (ref. [11]).

distribution in laboratory gases. The consequences are rather different though – the theory leads us to expect that the system is driven further and further away from 'equilibrium'.

On the larger scale of a galaxy, however, these processes are negligible and there is a great multiplicity of possible steady states. The problem is now to try and understand which of these get selected in our Universe. While the so-called spheroidal or elliptical galaxies form a class more amenable to idealized models, we also have many disc galaxies which are also called spiral because of the spectacular patterns which they exhibit (figure 1). We will comment on the stellar dynamics of disc galaxies as well, although gas dynamics certainly plays a crucial role in these systems.

Clearly, steady states are clearly not enough – one has to understand how they are reached and hence the section on time-dependent phenomena. One significant area which is mentioned but not covered in any detail is the numerical simulation of gravitational N-body systems which has reached a high level of sophistication, and provided a much needed reality check given that analytic methods have their limitations on account of the simplifying assumptions made.

The length of this article clearly precludes both detail and comprehensiveness. The goal is to give an overview, with an admittedly personal slant, of the basic observational facts, theoretical ideas, and some open issues, while providing links to the literature for those who are inclined to go further. Accordingly, very few primary sources are cited, but they can usually be found in the review articles mentioned. The subject has an authoritative and close to canonical text [1] which has just come out in a second edition. The reader can safely assume that a reference to standard material – say to Jeans or the virial theorem – can be followed up there, and in the other texts/monographs [2].

1.2 *Galaxies – Some phenomenology*

It took till the early twentieth century to establish that our galaxy and its neighbours like Andromeda (figure 1) are collections of about 10^{11} stars, about a hundred thousand light years (10^{23} cm) in size. Much of the twentieth century went by before one could arrive at the number of galaxies of this size in the observable Universe – again around 10^{11}. A few galaxies are bigger and brighter than our own, but many more are smaller, going down to dwarf galaxies which could be ten thousand times less luminous. Nevertheless, galaxies do form a distinct and unique unit in the organization of the visible Universe. The distribution of luminosities, while broad, does seem to cut off at both the upper and lower end, though there are worries that we may be lulled into complacence by observational selection. In fact it is quite as reasonable to say that the Universe consists of galaxies as to say that the body consists of cells. This does not exclude interesting structures either within a galaxy, such as star clusters or made up of galaxies such as groups, etc. Galaxies are gravitationally bound systems, whose separations are greater than their sizes, very roughly by a factor of ten – though interacting pairs, small tight groups of a few, exist. Still larger entities like clusters of galaxies, superclusters, filaments, walls all made up of galaxies do exist and their formation is an important part of the subject of physical cosmology [3]. But galaxies retain their identities within these structures. Likewise, there are structures within galaxies which are bound and have a distinct identity – globular star clusters (figure 2) – but the majority of stars in a galaxy are not contained in these.

Hubble's morphological classification of galaxies in his famous tuning fork diagram (figure 3) has stood the test of time. We will need only broad distinction between spheroidal and disc galaxies (left and right of the picture), and that between barred and unbarred (the two branches on the right of the picture). One basic problem is that one sees an individual galaxy only in projection. This implies that we only measure the two-dimensional or surface density of starlight on an image. Spectroscopic measurements of stars or gas also give only one component of the velocity, along the line of sight, via the Doppler effect. Clearly, inferring the structures of galaxies in six dimensions of phase space from the three-dimensional observations will not be unique and will require assuming a model. In the case of disc galaxies, we do have examples ranging all the way from face on to edge on, which are consistent with a model of a thin disc of stars with circular symmetry to zero order (and spiral features to first order, accompanied by a central 'bulge'). Within the thin circular disc assumption, one can use the apparent elliptical outline to infer the angle of inclination between the plane of the galaxy and the plane of the sky (equivalently, between the normal to the galaxy and our viewing direction). We can then de-project the surface density. Since we have only one velocity component measured, yet another assumption is needed. Azimuthal symmetry plus a steady state (no expansion or contraction) is satisfied by assuming circular orbits for the stars and gas to a first approximation. So the kinematics

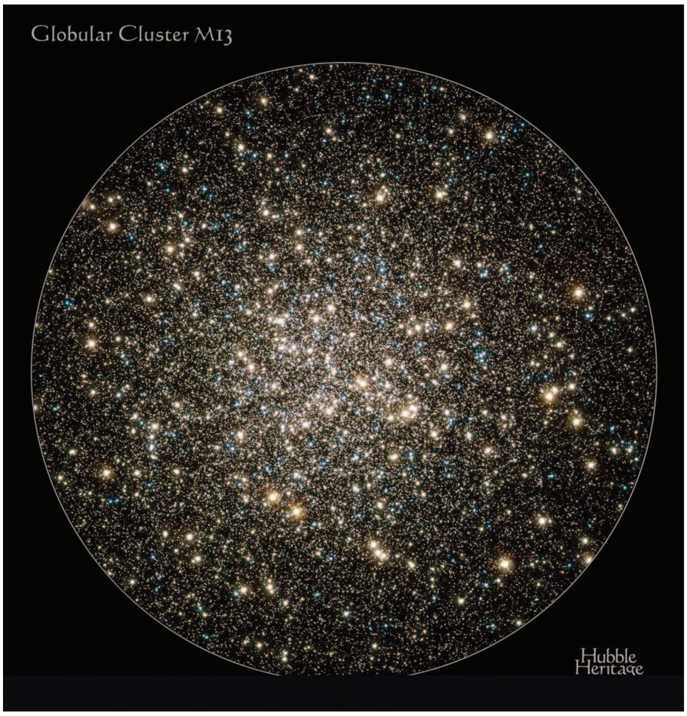

Figure 2. The centre of the globular cluster M13 (courtesy Hubble heritage project). The size is a few light years.

is described by the speed of circular motion as a function of distance from the centre $v_c(r)$, popularly known as the 'rotation curve'. It is possible to check the consistency of this assumption since we would be fitting a two-dimensional measurement of radial velocities with a function of one variable. Azimuthal symmetry seems grossly violated by the appearance of spiral galaxies in visible light. One should however note that these spectacular patterns manifest themselves most in the blue light of newly formed massive stars, while the great silent (actually underluminous) majority of older stars do not show such striking asymmetry and dominate the mass and potential. Twenty per cent is not a bad number for the deviation from full azimuthal symmetry.

One of the most remarkable discoveries of twentieth century astronomy came from studying the rotation curves of spiral galaxies. High school physics allows us to infer the distribution of mass with radius in the galaxy from this curve – $v_c^2(r)/r = GM(r)/r^2$. Here $M(r)$ refers to the mass within a sphere of radius r – using this for the inward gravitational force at r is strictly valid only if the density distribution is spherically symmetric, but the error for a non-spherical distribution is not large.

The difference between a disc galaxy and for example the solar system comes from the behaviour of $M(r)$ at large r. Since mass of the Sun dominates, we can take $M(r)$ to be a constant for all the planets and hence the rotation speed falls

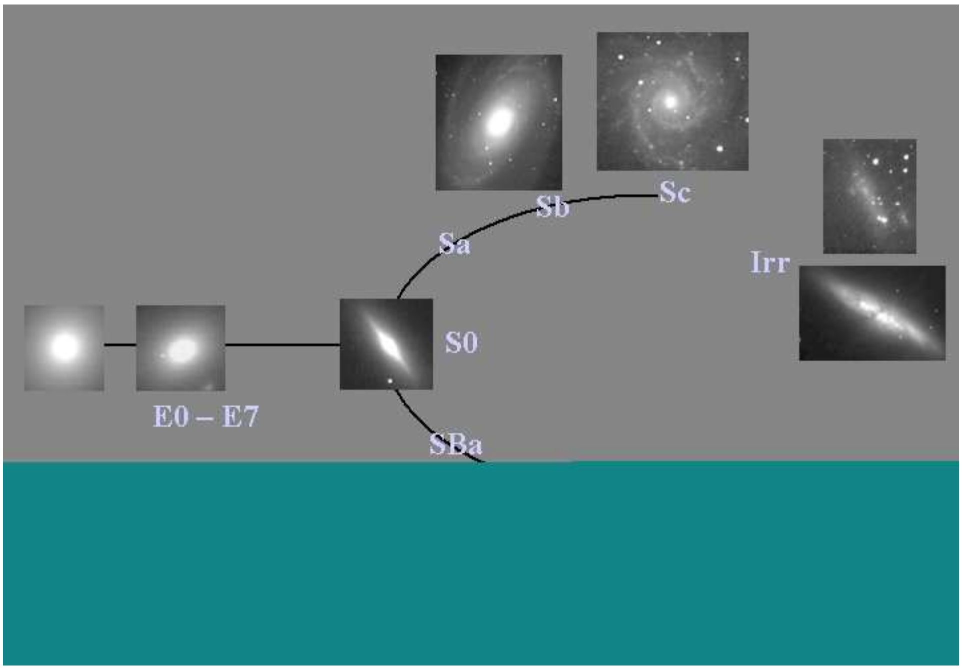

Figure 3. Hubble's 'tuning fork' classification of galaxies. The spheroidal systems are to the left and the discs, also known as spirals, to the right. Note the further classification of discs into unbarred (right, above) and barred (right, below).

off as $1/\sqrt{(r)}$. As late as the 1970s, it was felt that this law must hold in the outer regions of galaxies as well, since most of the visible light was seen from the inner regions. Of course, it is difficult to measure the Doppler shift if there is nothing to measure the Doppler shift of. So with hindsight it was a case of absence of evidence. The evidence assembled by Vera Rubin, her collaborators, and others [4] came from measurements of very faint spectral lines from gaseous material in the outer parts of galaxies, but also from the hydrogen gas emitting radiowaves at 21 cm – this is much more abundant in the outer parts of galaxies than stars. The rotation in the outer regions was very different from the naive expectation of $v_c \sim r^{-1/2}$. If anything, $v_c(r)$ was constant or mildly rising. This implies that $M(r)$ grows proportionally to r or faster. The matter making up this mass would have surface density $\sigma(r)$ falling as $1/r$ (the simple minded estimate is $\sigma(r) = M(r)/(\pi r^2)$). The observed surface density of starlight in a spiral galaxy falls off exponentially with distance from the centre – this is why the mass in stars alone would predict a falling, $r^{-1/2}$ rotation curve at large radii. This glaring discrepancy between gravitating and visible matter caused considerable unease when revealed in the first few cases, but is now the norm for galaxies. Astronomers have been forced to invoke 'dark matter' – matter distributed with a density going roughly as $\rho(r) \propto 1/r^2$ upto a length scale ($\sim$5) times larger scale than the bulk of the stars or gas. Clearly there has to be an outer cut-off to r to get finite mass.

One should be wary of things which explain just one observation, but dark matter has now explained many more things on a cosmological scale with a level of quantitative agreement that has led to wide acceptance. Acceptance does not imply that we know what dark matter is made of, even 25 years after, though there are some clues as to what it is not! For example, a million black holes in our Galaxy, each of a million solar masses, can safely be excluded on grounds of other effects they should produce. Particle physicists have embraced this astronomical crisis to their advantage – the lightest, most weakly interacting, and most stable of the unseen particles in theories going beyond the Standard Model are widely identified with dark matter – see the article by Mukhi and Roy in this special issue.

2. Dynamical principles

2.1 *The Liouville–Poisson equation*

The dynamics of stars in galaxies is an extreme case of the N-body problem which is already rich enough in Newtonian gravity that we shall stay away from general relativity – in any case speeds in galaxies are a sedate $c/1000$ or less. For simplicity, we shall even take all the masses to be equal to some fixed number m for the most part. The large value of N is meant to imply that we are interested in the overall behaviour of the system, as in the kinetic theory of gases. A one-particle statistical distribution function of stars with respect to three-dimensional position and velocity $\mathrm{d}M = f^{(1)}(x,v)\mathrm{d}x\mathrm{d}v$, captures this information, and one expects that we will be able to predict this function at later times if it is given initially. (Since we nowhere use components, vector notation is supressed.) Clearly, this is already a 'coarse-grained' description, since the full information about an N-body system lives in a $6N$-dimensional phase space. This description neglects 'correlations' – information contained in, say, the two-particle distribution function $f^{(2)}(x_1,v_1,x_2,v_2)$ which might conspire in a way that would influence the behaviour of $f^{(1)}$ – which we denote by f from now on. Notice also that unlike the case of fluids, we do not coarse grain further to a density, velocity, pressure at a given point in three-dimensional 'real' space x.

The self-consistent potential $\phi(x,t)$ can be calculated, using Poisson equation $\nabla^2\phi(x,t) = 4\pi G\rho(x,t)$, from the real space mass density $\rho(x,t)$. This, in turn, can be computed from the one-particle distribution function f by integrating over v, $\rho = \int f\mathrm{d}v$. (The normalization we chose for is 'mass per unit volume in six-dimensional position velocity space', rather than the more usual choice of 'number of particles per unit volume in position–momentum space'.) The motion of the particles is clearly going to depend on this potential, which appears in the Liouville equation for the evolution of f (the partial derivative with respect to a vector indicates the gradient).

$$\frac{\partial f}{\partial t} + v\cdot\frac{\partial f}{\partial x} - \frac{\partial\phi(x,t)}{\partial x}\cdot\frac{\partial f}{\partial v} = 0. \tag{1}$$

The notion of self-consistency implies that the potential which evolves the distribution function has to be calculated from the same distribution function. Well after Jeans applied it to galaxies, this feature re-appeared in the Vlasov theory of plasma physics, since the Coulomb potential obeys the same Poisson equation. Given that ϕ is linear in f, the third term shows nonlinearity rearing its head – ugly or beautiful depending on the beholder. We deviate from standard practice in calling (1) the Liouville–Poisson equation rather than the more common CBE (for collisionless Boltzmann equation) which does no justice to the inventor of the collision term, or gravitational Vlasov equation which would be anachronistic.

2.2 *Two-body relaxation*

The Liouville–Poisson equation is not the whole story. The potential has been calculated from a 'coarse-grained' – which means smoothed out – distribution function, which would accurately describe a truly continuous fluid with self-gravity. Since our sytem is discrete, a scattering process in which momentum is exchanged between two stars which pass close to each other has to be put in quite separately from the common self-consistent potential in whose orbits they both move before and after the interaction. Jeans himself gave an early, rough, but instructive estimate of the importance of these encounters. Unbound two-body motion under an inverse square force has an exact solution in Newtonian dynamics, familiar from Rutherford scattering. For gravitational forces between two equal mass particles, the scattering angle θ in the centre of mass frame in terms of the impact parameter b and relative speed v is given by $\tan(\theta/2) = (2Gm/bv^2)$. For a given relative speed, there is a critical impact parameter $b_{\rm crit} \sim Gm/v^2$ below which one gets deflections greater than a radian, going upto π (backscattering) as b tends to zero. For $b \gg b_{\rm crit}$ the deflections are small. The Jeans estimate treats the stars as hard spheres with a velocity-dependent radius of $b_{\rm crit}$, and gives a formula for the 'two-body relaxation time' (i.e typical time for significant change in the magnitude/direction of the velocity of a given star due to collisions). This is the mean free path for this radius divided by v

$$\tau_{\rm relaxation} \sim \frac{1}{\pi b_{\rm cirt}^2 n v} \sim \frac{v^3}{G^2 m^2 n}. \tag{2}$$

A more illuminating form of this equation emerges if we put in simple estimates for the particle density n and the relative velocity v, in terms of the extensive parameters N and R, the number of particles and overall system size. The velocity scale is set by the virial theorem, which states that the kinetic energy of a system with inverse square forces which is neither expanding nor contracting equals half the negative of the potential energy. We then compare this relaxation time to $t_{\rm dynamical}$, the typical time it takes a star to travel a distance of the order of the system size R.

$$n \sim N/R^3 : v^2 \sim GNm/R; \quad t_{\rm dynamical} \sim R/v. \tag{3}$$

Putting these into the earlier formula (2) results in cancellation of all dimensional quantities, leaving the neat result

$$t_{\rm relaxation}/t_{\rm dynamical} \sim N. \tag{4}$$

This means that a star executes many orbits before the effects of two-body collisions become significant. This is quite in contrast to a gas in the laboratory, which under most circumstances has a mean free path which is the shortest length scale in the problem. (High vacuum systems, including many laboratory plasmas, are exceptions.) This formula gives a measure of how long we can neglect the effects of two-body encounters, compared to the effects of the mean potential (which manifest themselves on the dynamical time-scale). It is at first surprising that the more stars we put in, the more we can neglect collisions. But looking at the collisions as a consequence of the discreteness or 'graininess' of the system, it is actually satisfying to see their effects disappearing as N goes to infinity, keeping the total mass and physical size fixed. This is the so-called collisionless limit in which it is accurate to describe the system by eq. (1).

The estimate of eq. (2) is not correct in neglecting small angle scattering. It is true that collisions with impact parameters from, say $4b_{\rm crit}$ to $8b_{\rm crit}$ will produce scattering angles half as much as those from $2b_{\rm crit}$ to $4b_{\rm crit}$. The cumulative effect of many random scatterings depends on the square of this deflection. So we are talking of an effect which is four times less, per collision, in the first case. However, this is precisely offset by the fourfold increased cross-section. The final contribution is the same for each logarithmic interval of impact parameter! Rather than having a negligible effect, we find that the collisions for $b \gg b_{\rm crit}$ actually dominate, by a factor proportional to the 'Coulomb logarithm', $\ln(b_{\max}/b_{\rm crit})$. What is the upper impact parameter that renders this logarithm finite? In plasmas, it is the Debye length, beyond which the two-particle correlations act to shield the Coulomb force. In a gravitating system, we actually have 'antishielding' (also calculated by Jeans – see below!) and the appropriate number is of order the system size R. Being inside a logarithm, it does not affect the answer very much. The ratio $R/b_{\rm crit}$ is v^2R/Gm which is easily seen (from eq. (3)) to be just of order N, the total number of particles. A precise definition and careful calculation of the two-body relaxation time for a star cluster shows that it exceeds the dynamical time by approximately the factor $N/(20 \ln N)$ [5].

2.3 *Velocity space diffusion and dynamical friction*

The formalism appropriate to this kind of situation is best appreciated by going back to Boltzmann's formulation of the kinetic theory of gases, which has a 'collision term' on the right-hand side of the Liouville equation to describe the change in the distribution function caused by these encounters. For molecular collisions, the angle of scattering, and hence the change in velocity, can be large. The Boltzmann collision term giving the rate of change of $f(v)$ is therefore an integral operator. Schematically,

$$\left(\frac{\partial f}{\partial t}\right)_{\rm collisions} = \int\int {\rm d}v''{\rm d}v'(-f(v)f(v')K(v,v')+f(v')f(v')K(v',v'')).$$

It is understood that all the f's are evaluated at the same position x, since a molecule collides with nearby ones. The loss term with a negative sign accounts for all other molecules with velocity v' with which those with a given v could collide, producing v'' and v'''. The 'gain' term accounts for the 'inverse' collisions which produce molecules with velocity v (along with v') starting with v'' and v'''. One sees that this is both nonlocal in v and nonlinear in f, and the kernel K gives the rates for these collisions.

In stellar dynamics, we have seen that collisions in which the change in velocity is small dominate by about an order of magnitude over others. In this case, one of the final velocities in the loss term, and one of the initial velocities in the gain term, is close to v. This allows a Taylor expansion, which will clearly give a term proportional to f and one to the velocity space gradient of f. The resulting equation simplifies to the so-called Fokker–Planck form

$$\left(\frac{\partial f}{\partial t}_{\text{collisions}}\right) = -\frac{\partial J(v)}{\partial v},$$

$$J(v) = -D : \frac{\partial f}{\partial v} - F(v)(v - v_0)f(v).$$

The change in f at a given v due to small angle collisions is expressed as the divergence of a current, $J(v)$. The current has two terms. The first is a 'diffusive' term, which represents a spreading out in velocity space, from regions with a higher phase space density to lower. This diffusion has a simple physical interpretation, viz. random small kicks in velocity space due to a large number of collisions, each causing a small change in velocity. Clearly, this cannot be the only effect, since all stars would gain energy by random walking away from the origin in velocity space. We therefore need the second, 'convective' term in the current – f is carried along at a speed $-F(v)(v - v_0)$ in velocity space. The convective term describes a systematic drift towards the origin of velocity space (chosen as the centre of mass frame, moving at v_0, of the stars in a given region of real space). The name 'dynamical friction' coined by Chandrasekhar has stuck. It describes a tendency for collisions to 'drag' a star to the mean velocity of its neighbours at the given position, denoted by v_0 above. The systematic derivation of both terms due to Landau [6] for the case of Coulomb collisions in a plasma clearly brings out the close relation between diffusion and dynamical friction.

The original Fokker–Planck equation was derived for a heavy Brownian particle in a fluid made up of much lighter molecules, so that the assumption of many small impulses was a valid one [6]. For Brownian motion, the term is both linear and local. In the context of stellar dynamics and plasmas, it is neither. The schematic derivation sketched above already shows that F and D are functions of v, and v_0 is a function of x. They are all given by integrals over f, and so the locality in velocity is only apparent.

The basic Liouville equation conserves the phase space density. In fact, it expresses the constancy of f along the trajectory of a given particle in phase space. The dynamical friction term clearly increases phase space density, since by itself it would drag all particles at a given point in space to zero velocity relative to each other. The diffusive term does the opposite, since it spreads the particles out in velocity space. Given that volume in phase space is associated with entropy, one should make sure that the net effect increases. This was already fixed by Boltzmann's H-theorem, which applies to his general collision term and hence to the specialized Fokker–Planck form as well [6].

3. Collisional stellar dynamics: Globular and other star clusters

Globular clusters are nearly perfectly spherical sytems of typically a million stars, moving with relative velocities of a few kilometres per second, and extending for several parsecs around a dense core which can be a parsec (3×10^{16} m) in size (figure 2). Spitzer's 1987 book [5] is the standard reference for the basics, though the subject, like most others in astronomy, has seen great progress since that time. The dynamical time is thus about a million years, and the Coulomb logarithm is about twenty, giving a rough estimate of a few hundred million years for the relaxation time. This might seem long, until we realize that the globular clusters are 10^{10} years old, and hence have gone through many relaxation times in their evolution. It therefore makes sense to ask whether there is an equilibrium state to which they can relax. Based on experience with gases, one would expect something at constant temperature, with a density and hence pressure increasing inwards, to balance the force of gravity. This does not work, for at least two reasons. Firstly, the collisions tend to establish a Maxwellian distribution of velocities, and some fraction of this distribution lies at above the escape velocity at any given point in the cluster. There is hence a gradual process of 'evaporation' of stars from the cluster. A simple estimate would give a few per cent (the fraction in the tail) per relaxation time. Evaporating stars carry positive energy, and leave the system more tightly bound (and the binding energy per star even more negative). Going back to the simple estimates of eq. (3), one can see that the size goes down, the speed increases, and the number goes down. All these contribute to shortening the two-body relaxation time. In fact, beyond a point, the Fokker–Planck description breaks down and very interesting processes like three-body interactions with formation of binaries and ejection of the third body can happen [5].

A more subtle obstacle to equilibrium, unrelated to escape, showed up in a remarkable analysis by Antonov [7]. He considered the problem of a star cluster at constant temperature, confined to a spherical box of radius R, with self-gravity, imposing a criterion of entropy maximization at fixed energy for equilibrium. This is known in the trade as the finite isothermal sphere problem, and has a rather straightforward solution. What Antonov's predecessors missed was that the entropy of the so-called equilibrium solution ceases to be a true maximum of the entropy functional $S = -\int f \ln f \mathrm{d}r\mathrm{d}v$ as the radius increases for fixed central density and temperature. With increasing R, the ratio of the central density to that at the boundary goes up, and the instability occurs when it reaches a magic number of 709. With hindsight, arguments have been given [8,9] that separating the system into a 'core' and a 'halo' with most of the mass and the entropy in the halo, and a small hot tightly bound core, can give a higher entropy provided we give the halo enough room to expand. Notice that the Antonov analysis assumes the existence of two-body collisions tacitly – both by using the isothermal condition and by maximizing entropy. This instability appears anti-thermodynamic, because differences of 'temperature' appear to arise spontaneously. However, the equality of temperature of systems in contact in a maximum entropy state is itself a derived concept in statistical mechanics, assuming the additivity of energy between subsystems [6]. However, the long range of gravity does not allow the usual separation of the energy into individual energies for subsystems which are additive apart from small surface terms. Hence uniformity of temperature fails for a star cluster. It is also shown assuming the same additivity [6] that angular velocity should be constant in a system with fixed total angular momentum. This too fails for a star cluster. One can (in principle) put all the angular momentum into one star orbiting at a sufficiently large distance, and build a zero angular momentum, maximum entropy model for the rest with essentially the same energy as originally prescribed. This rather academic construction does reflect a well-known astrophysical effect – outward transport of angular momentum in rotating systems. The progenitors of globular clusters may well have had angular momentum which they seem to have got rid of by evaporation or at least moved to their outer parts.

An attempt to do statistical mechanics of the N-body system with a $1/r$ potential in a box of size R shows that even the microcanonical ensemble fails, because the phase volume at constant energy is infinite. The root cause is that even with fixed total energy, one can consider configurations in which two very close particles have a large negative potential energy, which is fed into the kinetic energy of the remainder. The resulting gain of entropy more than offsets the entropy lost by making this pair move around together, even for a three-body system. This becomes apparent from simple power counting. The momentum space part of the phase volume below total energy E for N free particles scales as a typical momentum raised the power of $3N$, i.e. as $E^{3N/2}$. We hence have a density of states scaling as $E^{3N/2-1}$. A bound pair with energy more negative than $-E_{\rm b}$ is confined to relative separations less than $GM^2/E_{\rm b}$ and has relative momentum of the order of $E_{\rm b}^{1/2}$. Putting these two facts together gives a phase volume below $-E_{\rm b}$ scaling as $E_{\rm b}^{-3/2}$ implying an $E_{\rm b}^{-5/2}$ tail in the density of states for the relative motion of this pair. Now consider $N+2$ stars in a box of size R, and let us look at configurations with any fixed total energy E, realized as $E = E_{\rm b}$ with large positive $E_{\rm b}$ for N stars and energy $-E_{\rm b}$ for the pair. The density of states for the combination is obtained by convolving the two results given above to get $\int E_{\rm b}^{-5/2}(E+E_{\rm b})^{3(N+1)/2-1}{\rm d}E_{\rm b}$ This already diverges for $N = 1$, i.e. a three-body system! (The $N+1$ in the free particle part allows for the centre of mass motion of the pair.) One can picture the small r divergence in the density of states at finite energy as follows. For free particles, the constant energy surface is a sphere in $3N$ momentum dimensions, with the spatial coordinates independently ranging over the box. With a $1/r$ attractive interaction, one sacrifices volume in coordinate space by restricting two particles to move close to each other, but makes up by using their binding energy to blow up the momentum space sphere.

These phase space considerations are necessary but not sufficient for expecting the formation of binaries and their evolution to high binding energy – a dynamical mechanism is needed. Work on the interaction of binaries with single stars, including extensive numerical simulations amply confirmed the basic insight by Heggie [10] that a collision of a single star with an already tight binary drives the binary to even tighter binding and acts as a heat source for the intruder. (A very intuitive picture would be a slow particle hitting a rapidly rotating fan.) This leads to the concept of 'core collapse' of globular clusters in which a few tight binaries act as significant heat sources for the rest of the cluster. This is shown by a simple estimate – stars in a tight binary of two compact objects can move at a couple of hundred kilometres per second and hence supply enough energy to give a thousand stars speeds of a few kilometres per second, quite significant for the energetics of the cluster core.

As we go to clusters with a smaller number of stars, such as the so-called 'open' or 'galactic' clusters with a few thousand at most, the Pleiades being a well known example, the fluid description becomes increasingly suspect. The Coulomb logarithm becomes smaller, so the hard collisions (causing large changes in velocity) are more important in these clusters than in their more populous and globular cousins. They are also more weakly bound and susceptible to disruption as they move around in the galaxy and only live for about a few hundred million years after they are born.

The second outstanding discovery about galaxies in the twentieth century is the ubiquity of black holes right in their centres including our own galaxy and our friendly neighbour, Andromeda aka M31, and even its midget companion M32 (figure 1). The clusters surrounding these black holes can have more stars than the globulars, and the Fokker–Planck description should be better. There is now an interesting kind of 'evaporation', into the black hole. The influence of the central mass results in a central increase of density and random as well as rotational velocities of the stars, now dominated at small enough distances by the potential of the black hole itself. Unlike the globular clusters, these can have significant angular momentum. There are very interesting issues relating to these stellar systems which are the topic of much current work [11].

4. Collisionless stellar systems: Galaxies

4.1 *Steady state models*

We now come to systems like galaxies with so many particles that the collisional effects are negligible. The dynamical time is of order 10^8 years and multiplying by any reasonable number of particles gets one comfortably above the Hubble time which is of order 10^{10} years. Forty years ago, one would have thought of the particles as stars. But with the discovery of non-baryonic dark matter, making up most of the mass of most galaxies, the same model can be applied to this component. What this matter is really made of is not known but, provided it interacts mainly through gravity, we can sidestep the issue and go on to a general discussion of collisionless self-gravitating phase space fluids obeying eq. (1). The very first step is to discard our preconceptions about systems maximizing their entropy, or having a Maxwellian distribution of velocities, or even having the same pressure in all directions, all of which are consequences of the collision term. We should now think of a galaxy as being in *a* steady state, rather than in *the* equilibrium state. Without the collision term, a much richer variety of steady states is allowed, even after one fixes parameters like the total mass, binding energy or equivalently length scale since $E_{\rm b} \sim -GM^2/R$, and angular momentum. The first two can be removed as scale parameters by choosing suitable units of x and v. Angular momentum is then characterized by a dimensionless parameter. The variety of steady states allowed by the theory is actually a virtue – real galaxies come with a variety of shapes and velocity distributions.

The Liouville theorem is equivalent to the statement that f is constant along a phase space trajectory. If it is also constant in time at a given x, v (the steady state condition), this implies that f, considered as a function of x, v, is a conserved quantity. Let I_1, I_2, I_3 be conserved quantities for motion in the potential of the galaxy. The notation I is used because they are called 'integrals of the motion', and we usually have no more than three independent ones in three-space dimensions (i.e. if the Hamilton–Jacobi equation separates but only in one coordinate system). The condition that the distribution function depends on x and v via these is expressed by

$$f(X, v) = F(I_1, I_2, I_3). \tag{5}$$

Technically, these are required to be 'isolating' integrals and the action variables do the job. This simple-looking criterion (5) for a steady state distribution in phase space goes by the name of Jeans' theorem. The simplicity is deceptive for two reasons. First, even if we were given the potential, we have no algorithm for finding integrals of motion, which may not even exist – modern chaos theory tells us that 'any old potential' does not have them. Second, we are not given the potential – it has to be computed from the density which has to be computed from f.

Let us deal with the problem of self-consistency first. One can break the vicious circle by assuming a potential with some geometrical symmetry – for example, disc galaxies appear approximately axisymmetric. We have energy and the z component of angular momentum as integrals for such potentials. So assuming a distribution function which is a known function of E and L_z, we can set up Poisson's equation for the potential, which will have the potential on the right-hand side as well, since the density is an integral over f which contains the potential ϕ via the energy $E = v^2/2 + \phi$. One can imagine solving this nonlinear integro-differential equation by starting from an initial guess plus iteration, which if all goes well will converge. At least, we can see that a naive count of equations vs. unknowns works. The potential is a function of two variables, r and z, in an axisymmetric situation, and we have a function of the two integrals of motion as input. Although this procedure seems quite general, it already makes a prediction which can be tested if you happen to live in such a galaxy. Resolving the velocity into cylindrical polar components v_r, v_z, v_θ, we find that the z component of the angular momentum only contains the last of them, v_θ, while only the sum of squares of the first two occurs in the energy. It follows that the average value of the square of v_z will be the same as v_r. This is in clear disagreement with observations in our solar neighbourhood in the galaxy. This argument led to the interesting conclusion that the potential of our galaxy must possess, at least approximately, a 'third integral'. To see how this helps, consider a three-dimensional harmonic oscillator potential, which allows separately conserved energies in the x, y and z directions. A distribution function f which is a general function of E_x, E_y and E_z would allow $\langle v_r^2 \rangle$ not equal to $\langle v_z^2 \rangle$. Of course, the actual potential of a galaxy is very far from a three-dimensional harmonic oscillator, but this example illustrates the consequence of additional integrals of motion. Specialists in Hamiltonian chaos have undoubtedly seen and cited the Henon–Heiles potential which first made its appearance in an astronomical journal in precisely the context of the third integral in galaxies [12].

If we now go to a general potential with spherical symmetry, we can have three integrals, E, L_z, L^2. At first sight, one might think that making f a function of all three might break the spherical symmetry of the potential. We show that this is not necessarily so, using an admittedly artificial example. One can start from a galaxy in which f is only a function of E and L^2, to start with. Since L_z occurs only via L_z^2 in the distribution function, we have each orbit traversed by equal numbers of stars with positive and negative angular momentum about the z-axis, and a zero net angular momentum. Now let us reverse the sign of L_z for those stars which have a negative value. This does not change their contribution to the potential, and so the self-consistency is undisturbed. But the distribution function now depends on L_z, and in fact the system has a net angular momentum about z. This is a counterexample to the naive belief that such a system with net rotation about one axis cannot have a potential with spherical symmetry. This also cautions us against the intuitive idea that flattening must be caused by rotation – this presumes isotropic stresses coming from isotropic velocity distributions.

We saw that we had a nice balance of classical (i.e. symmetry-induced) integrals of motion available in an axisymmetric system, against the number needed to construct such a system. For spherical systems, there was an embarassment of riches – one could construct the same spherical density distribution in more than one way. We now examine a situation with lower spatial symmetry and hence only one classical integral avaiable, viz. the energy. If f depends only on E, we can clearly have a situation where $\langle v_x^2\rangle = \langle v_y^2\rangle = \langle v_z^2\rangle$ (isotropic stress tensor) and $\langle v_x\rangle = \langle v_y\rangle = \langle v_z\rangle$ (no bulk motion). One can guess intuitively that the 'pressure' obeys Pascals law and therefore the final galaxy will be spherical – indeed, there is a theorem due to Lichtenstein which proves this. In fact, the older models for spherical systems like globular clusters used a function of E alone (but not a Maxwellian, because that would have unbound stars).

All this is fine for a spherical cluster but leads to a difficulty when we no longer have spherical symmetry or even axial symmetry. This is quite well motivated after the inference from observations that the three-dimensional shape of some galaxies which appear elliptical on the sky is actually triaxial [1,2]. A general triaxial potential does not have integrals coming from spatial symmetry. So it appears as if we are only allowed to use functions of E in the Jeans theorem, forcing one back to spherical systems. Since triaxial galaxies exist, and successful numerical models of them also exist, there must be a way out of this dilemma. It was found in a class of ellipsoidal potentials with three integrals which could be used to build fairly realistic models of triaxial galaxies. Mumerically constructed models also show similar orbital structure to these special potentials [1,2]. It is remarkable that collisionless stellar systems actually appear to select density distributions and hence potentials which admit (at least approximately) additional integrals of motion not coming from spatial symmetries.

What happens if we set up a density distribution whose potential does not have three integrals? Since it does not allow a steady state, it presumably evolves till it reaches one. One cannot *a priori* rule out models, and galaxies, which are never in a steady state, for example, they could keep oscillating. Many authors have constructed such models [13]. In fact, in one-dimensional stellar dynamics (N-sheets interacting with a $|x_1 - x_2|$ potential which are delightfully easy to simulate), oscillating models seem quite robust, but in the more realistic three-dimensional case, they are rather fragile and do not seem to turn up from general initial conditions because of their rather flat density profiles. One cannot also rule out by pure logic, models in which Jeans theorem fails, the potential supports some regular orbits and some chaotic ones whose occupancy conspires to reproduce the required density – as with realistic three-dimensional oscillating models, one can only plead absence of evidence.

4.2 *Collisionless relaxation: Galaxy formation and interaction*

We know that the galaxies were formed at some epoch – they did not always exist as we see them now. It is true that physical processes going beyond stellar dynamics shaped them – the cooling and fragmentation of primordial gas clouds, the formation of stars, etc. There is, however, a purely dynamical scenario for attaining a steady state. Motivated by standard cosmology, in which galaxies form from primordial density fluctuations, one can take a collection of dark matter particles, which are bound, but initially expanding – a slightly overdense part of the Universe, small enough to think of in Newtonian terms, and with negative total energy. This initial condition implies that the particles turn around and re-collapse. Given deviations from perfect symmetry and from uniform density, all the particles do not arrive at the centre at the same time (which they would for a sphere of uniform density starting from rest). Early cosmologists assumed that they go on to form a bound system in a steady state, obeying the virial theorem – hence the term 'virialization'. Today, we know from numerical simulations that this does indeed happen, and that the whole process takes only a few dynamical times, albeit with some infall of latecomers continuing longer. The resulting density profile broadly matches that which would explain the rotation curves of galaxies. Well before all these developments, the process of reaching a steady state was rationalized in a rather appealing physical picture due to Lynden–Bell which he named 'violent relaxation' [14], again a name that has stuck. The violence is in contrast to two-body relaxation, which takes a very large number of dynamical times, and is therefore far too slow on the scale of galaxies. There are two essential ingredients to this picture which could loosely be called mixing in action and in angle variables. Assuming that the final potential has such variables, we can argue that the distribution function will actually lose its dependence on the angle variables by 'phase mixing'. This process depends on the potential being anharmonic, i.e. the period depending on energy (and other action variables in more than one dimension). An ensemble of particles with a small, let us say 20 per cent range of periods, will be spread out over 2π in orbital phase in five periods. The coarse grained distribution becomes a function of action variables alone (in one dimension, just the energy). This part of the mechanism is purely kinematic because it depends upon on the amplitude dependence of the orbital period. The second part includes the time dependence of the potential, which allows the energy (in one dimension) or in general the action variables to change for a given star/particle. Of course, the total energy is conserved, because the particles producing the potential fluctuations are exchanging energy with our test particle – this is sometimes called 'wave particle interaction'. The last step in Lynden-Bell's original model can be regarded as the limiting case of extreme violence. He divided the phase space into cells and shuffled them around (along with the particles contained), to mimic the mixing processes consistent with the conservation of energy and Liouville's theorem, and maximized the entropy. (One could say that Doctor Lynden-Bell's prescription was to shake the bottle!) The Liouville theorem acts like a Pauli exclusion principle since phase space elements which were distinct to start with conserve their fine grained phase volume and can never overlap. The occurence of a Fermi-like distribution function in a purely classical problem was one of the intriguing aspects of this picture. Because of the phase space constraint, there is no room for core collapse like phenomena. All this still leaves open the question as to whether we should be maximizing the usual entropy expression at all, given that there is no collision term. In fact, some authors have looked at all convex functionals of f [8].

Many decades and simulations later, the consensus is that the violence is never as strong as in the original extreme model, and various attempts have been made to come up with reasonable scenarios which reproduce the results of numerical simulations. One probably needs a violence parameter set by the substructure in both density and velocity in the initial conditions. In the cosmological context, these arise from the primordial density fluctuation spectrum, but purely as a problem of approach to a steady state in the statistical gravitational dynamics of N-bodies, it can be studied more generally. There has been an intriguing claim of quasi-universality in the density profiles which are the end points of collisionless relaxation, for which there is no strong theoretical argument at present. The root of the problem is that there are no small dimensionless parameters, only one time-scale $t_{\rm dynamical} \sim (GM/R^3)^{-1/2}$ which may as well be chosen as the unit of time, leaving us with at best general constraints and semi-quantitative pictures, drawing lessons from simulations rather than firmly predicting what they should give.

There is one more reason to pursue time-dependent stellar dynamics, well motivated by astronomical observations. Given that galaxies are part of larger structures and have relative motions other than those dictated by Hubble expansion, it is not a surprise that some fraction of them do come close enough to interact, and more of them did in the past. The 'island Universe' paradigm is oversimplified – observations and simulations have shown that interactions have had an important role in shaping galaxies as we see them today. Standard cosmology also leads us to expect that large structures form by merger of smaller ones, which collapsed earlier because of their higher overdensity.

Given two galaxies, each initially in a steady state, which then pass each other at not too high speeds and not too large distances, one expects their mutual gravitational interaction to leave its impact on each other, and on their relative motion. One might expect some conversion of the energy of orbital motion into energy of internal motions, making the interaction inelastic. The simplest possible assumption to make is that the time-scale of the encounter is much shorter than the dynamical time-scale within each galaxy. Under these conditions, the so-called 'impulse approximation' holds – each star in each galaxy is given a kick in velocity space calculable from the distance of the closest approach and relative speed. Much has been learnt from this simplified (and hence analytically tractable) picture in the early days of the subject though the tool of choice is now naturally numerical simulation, which allows one to explore a richer and more realistic parameter space [15]. The most spectacular things happen when the impulse approximation breaks down completely and the galaxies are greatly modified or even merge. One of the attractive ideas in the field was that the dichotomy between spiral and elliptical galaxies would be resolved by taking the latter to be products of mergers of the former in dense environments. Like many attractive ideas, this turns out to be more complex when examined in detail. But anyone examining the classic, early work of the Toomres will be left in no doubt of the power of the interaction idea in explaining quite naturally, some of the strangest galaxies which have been observed [16].

4.3 *Instabilities*

There is one regime of time-dependent collisionless stellar dynamics which can be attacked systematically and is relevant to understanding real galaxies – linear stability (or otherwise). Plasma physics has a long tradition of perturbing various collisionless equilibria – the classic one due to Landau is well known [6]. In the stable case, we have 'Landau damping' which is an early example of 'dissipation without dissipation' – the Vlasov equation which is nothing but a re-statement of Newton's laws and Maxwell's equations, when properly analysed shows how a disturbance to the distribution function can be damped. In some sense, this 'damping' arises from a similar cause as the anharmonic oscillator example given in the context of violent relaxation – there is a continuum of modes with different frequencies whose superposition can give rise to damping. But we also can have situations in which the disturbance can grow – the classic plasma physics example is the 'two-stream' instability – more generally, any situation when the distribution function is not a decreasing function of energy in some range.

We expect any galaxy model with a claim to astrophysical relevance to be linearly stable. If not, this state could only be attained by artificially fine-tuned initial conditions. Galaxies differ from plasmas – the interaction is attractive, there is no neutralizing background, and one starts with a finite system at the outset. In a pioneering study, Jeans (again!) set aside the point about the neutralizing background and the finite system and gave an early analysis of gravitational instability in a putative uniform infinite system of density ρ_0. The result, at long wavelengths, can simply be viewed as a plasma frequency squared with the wrong sign $-\omega_{\mathrm{p}}^2 = 4\pi n_{\mathrm{e}} e^2/m_{\mathrm{e}}$ is replaced by $\omega_{\mathrm{j}}^2 = -4\pi G\rho_0$ (since e^2 goes to $-Gm^2$ and $\rho_0 = nm$). Including the restoring force due to gas pressure (which gives rise to sound waves at speed c_{s} acting alone), the dispersion relation now has the form $\omega^2 = c_z^2 k^2 - 4\pi G\rho_0$. The restoring force, measured by ω^2, can now go negative at $k < k_j$ where $k_j^2 = 4\pi G\rho_0/c_{\mathrm{s}}^2$, i.e. long wavelengths. Naturally, the absolute value of the pure imaginary frequency $|\omega|$ is then to be regarded as an instability growth rate, since the $\exp(i\omega t)$ modes become $\exp(\pm|\omega|t)$. The argument is flawed (and is known in the stellar dynamics trade as the 'Jeans swindle') because there is no compensating background and hence no uniform system in equilibrium which is being perturbed. Nevertheless, the time-scale and critical wave vector computed from the above dispersion relation have some significance on purely dimensional grounds. In fact, the dynamical time of equation is $\sim(G\rho_0)^{-1/2}$, and the inverse of k_j becomes, in the order of magnitude, the system size when one puts in the estimates from (1). The virial theorem ensures that the random (or systematic) velocities adjust themselves in the steady state to values such that the Jeans wavelength is of the order of the system size, and longer scales are irrelevant.

For a model of a finite galaxy, which is already a steady state solution to the Liouville–Poisson equation, one should ask if small perturbations decay or grow. The analysis of stability of collisionless stellar systems is in one way more involved than in homogeneous plasmas since we no longer have the luxury of doing each wave vector separately, which an infinite uniform system would offer us. Whole treatises exist on this area [17]. If the unperturbed system is axisymmetric, solutions can be sought with an angular dependence $\exp(im\phi)$. The case of $m = 2$ seems to offer one of the few generalizations in the subject, which goes by the name of the 'Ostriker–Peebles' [18] criterion. It was found that rapidly rotating stellar systems are unstable. At a rather small ratio $T/W = 0.14$ of kinetic energy in ordered rotational motion to the total energy, instability to formation of a bar-like structure sets in. The absence of violent instability in most observed galaxies led these authors to propose a massive halo on purely theoretical grounds – it contributes to W but not to T and stabilizes the disc. If we associate the kinetic energy of random motions with 'heat', then one can regard a state with small random velocities as 'cold', with a low entropy and hence a high free energy which can be tapped by the instability. This leads to a configuration which has lower free energy. At this level of generality, there are analogous phenomena for rotating masses of incompressible fluid, for which Chandrasekhar's lectures which became the book [19] are a standard reference. The bar instability is then related to that found by Jacobi – the rotating and oblate Maclaurin spheroid breaks symmetry around its axis and goes to a triaxial shape at large enough rotational energy.

Even if a flattened rapidly rotating disc of stars satisfies the criterion for stability against the bar mode, the question of local stability arises. Unlike the Jeans analysis, it is now possible to honestly perturb a rotating disk with sound speed c_{s}^2, surface density σ and epicyclic frequency κ. The 'epicyclic' should cause no apprehension of a Ptolemaic revival, being just the term used in galactic dynamics for the frequency of small radial oscillations about a stable circular orbit. Conservation of angular momentum per unit mass $h = r^2\mathrm{d}\theta/\mathrm{d}t$ implies accompanying oscillations in the tangential direction leading by 90°. κ can be determined from the second derivative of the effective potential (per unit mass) for radial motion, $V_{\mathrm{eff}}(r) = V_{\mathrm{grav}}(r) + h^2/2r^2$ for a general central potential $V(r)$ (per unit mass) and it is given by $\kappa^2 = (1/2r^3)(\mathrm{d}/\mathrm{d}r)(r^3\mathrm{d}V_{\mathrm{grav}}/\mathrm{d}r)$. A small perturbation of radial velocity at any wave vector would give rise to equally spaced rings of extra and deficient surface density, and would oscillate at the epicyclic frequency if there were no gas pressure and no self-gravity of perturbation. The gas pressure term is as earlier, but the restoring force due to the self-gravity term in two dimensions takes the form $-2\pi G\rho_0|k|$. The $|k|$ term is clearly needed to make up the dimensions of σ to those of ρ but the physical reason for this different wave vector dependence in two dimensions is that the force due to a line of excess density on a sheet (say one half wavelength of our sinusoidal perturbation) falls off inversely with distance, unlike the corresponding sheet in three dimensions whose force is independent of distance. The resulting dispersion relation for gas disks was derived by Safronov in the context of solar system formation $\omega^2 = c_{\mathrm{s}}^2k^2 - 2\pi G\rho_0|k| + k^2$. The main properties of this relation are that (a) for surface densities less than $\sigma_{\mathrm{crit}} = c(c_{\mathrm{s}}\kappa/\pi G)$, there is no instability – the quadratic expression for ω^2 has no real roots. And as the surface density is increased to $\rho_{\rm crit}$ the instability first appears at the double root $k_{\rm crit} = \kappa/c_{\rm s}$. We have chosen the case of gas for its relative simplicity. When one is dealing with stars, as one must in applications to real galaxies, we must keep in mind that stars have random velocities like a gas but they are collisionless and cannot support sound waves, so the k^2 term is not appropriate. The effect of the self-gravity averages to zero for stars whose typical amplitude of epicyclic oscillation $v_{\rm rms,radial}/\kappa$ is significantly greater than the inverse wavenumber k^{-1} of the gravitational potential. The full treatment was provided in the 1960s by Lin and Shu, and by Kalnajs (we refer to Toomre's review [20] and the book [21] for details). It is found that the frequency of small radial density/velocity perturbations goes back to κ at short wavelengths (high k) as well. Given this basic difference between gas and stars, we should be grateful to the famous and widely used result of Toomre [20] for the critical surface density of a stellar disc. It is what we get for a gas disc if we are prepared to replace $c_{\rm s}$ by $v_{\rm rms,radial}$ (and less convincingly, π by 3.358!). Real galaxies have both gas and stars, with the gas component carrying less mass but being more responsive because of its lower random velocity. The local instability work is closely coupled with attempts to construct a theory of the spiral structure of disc galaxies based on density waves. These are spiral patterns in the density and the potential with say $m = 2$ which rotate at a fixed angular velocity, so that stars (and gas) at small radii actually overtake a spiral arm and leave it. One general statement property is that that spiral waves carry angular momentum outwards. This is a rich and fascinating topic on which the last word has yet to be said, not for lack of trying [20,21]. Nature has found it far easier to make spiral galaxies than astrophysicists have.

5. Concluding remarks

This parting section carries the statutory warning of reflecting a personal viewpoint and taste, even more than the rest of the article. The phrase 'statistical mechanics of self-gravitating systems' has been avoided, out of respect for the non-existence of the standard ensembles for the collisional systems, and the rich variety of equilibria for collisionless systems. Less timid authors do talk about the statistical mechanics of such systems – the book [22], the review [9] or the recent article [23]. One can certainly explore general issues raised by many-particle systems with a $1/r$ attractive potential obeying statistical mechanics, using short distance cut-offs and finite boxes to make things well-defined, but it is not clear that the results have applications to astronomy. Perhaps the most prominent claim to a general formulation which is so applicable is the ambitious attempt [22] at a synthesis, invoking all the terminology of statistical mechanics such as grand ensembles etc. which has given rise to an entire school in the context of cosmological clustering. At its base, however, is a heuristic ansatz [24]. The cosmological clustering problem is outside the main scope of this survey, but it is worth remarking that equilibrium ideas do not apply in this case. The equations of motion for cosmological clustering in an expanding Universe [3] can be put in a form which is explicitly dissipative, or another form with an explicitly time-dependent interaction [25].

My own favourite physics issue regarding collisionless stellar systems [26] is the overall landscape of stable solutions, and the initial conditions which lead to them. Because of the dissipative nature of the evolution, it is possible that the final states might admit of some broad classification based on those features of the initial conditions which they retain. This would in some sense be a thermodynamic goal but with many (but hopefully not too many) more parameters characterizing the final state than the usual energy and volume which we use for a laboratory gas. The phase space density constraint must enter in some way, i.e. whether the system is 'hot' or 'cold' in the sense explained earlier. With a proper choice of units, mass, energy and size are eliminated, everything else becomes dimensionless, angular momentum to start with but perhaps even some features of its distribution, and some aspects of the density profile and its anisotropy. Are we (meaning both observers and simulators) seeing a somewhat limited subset of what the Liouville–Poisson system allows because cosmological initial conditions have been kind to us? Perhaps the whole steady state paradigm is a simplification, and we have slow evolution or decay of some features even on cosmological time-scales? One cannot rule out collective effects with time-scales filling in the desert between the dynamical/crossing time and the Hubble time. Anything lasting a few tens of dynamical times or longer (e.g. Arnold diffusion in a potential which is not quite integrable) could give rise to pseudosteady states for billions of years.

In real galaxies, such hypothetical processes compete with real ones such as encounters with satellites and neighbours and gas infall which produce evolutionary effects on the Hubble time-scale. The other factor which is significant for realistic models of galaxies is the role of a central black hole whose presence is now recognized to be the rule rather than the exception. While it has a small fraction of the total galaxy mass, the influence on the dynamics of the central regions cannot be ignored. For example, the potential of a self-consistent triaxial model has stars near the centre on 'box orbits' which are so called because they broadly resemble the Lissajous figures of a three-dimensional harmonic oscillator with incommensurate frequencies. All stars on such orbits sooner or later visit the vicinity of the central mass which has quite a different kind of potential, and this must be taken into account. In such ways do real galaxies go beyond the realm of our idealized models.

Acknowledgments

I would like to thank C J Jog, M Das and A Thyagaraja who went through the manuscript in detail and made many helpful suggestions. My special thanks to S Sridhar for all the interactions we have had on this subject over many years. This article was mainly written during visits to the Chennai Mathematical Institute, Chennai, the National Centre for Biological Sciences, Bangalore, and the Canadian Institute for Theoretical Astrophysics, Toronto, and I thank all three institutions for their hospitality.

References

[1] J Binney and S Tremaine, *Galactic dynamics*, 2nd edn (Princeton University Press, Princeton, NJ, USA, 2008)
[2] G Bertin, *Introduction to galaxy dynamics* (Cambridge University Press, 1992)
A M Fridman and V L Polyachenko, *Physics of gravitating systems. I. Equilibrium and stability* (Springer, 1984)
[3] P J E Peebles, *Physical cosmology* (Princeton, 1992)
T Padmanabhan, *Theoretical astrophysics* (Cambridge, 2004) Vol. III
[4] V C Rubin, *Dark matter, in spiral galaxies* (Scientific American, 1983) 248, 88
[5] L Spitzer, *Dynamical evolution of globular clusters* (Princeton University Press, Princeton, NJ, 1987)
[6] L P Pitaevskii and E M Lifshitz, *Physical kinetics* (Pergamon, 1988) for the collision term, the Fokker–Planck form, and Landau damping
L D Landau and E M Lifshitz, *Statistical physics* (Pergamon, 1992) for the logic behind equality of temperature and angular velocity in thermal equilibrium
[7] V A Antonov, Vestnik Leningrad University, **7** 135 (1962), translation in *Dynamics of globular clusters, IAU symposium 113* 525 (Hut and Goodman ed.) (1985)
See also D Lynden-Bell and R Wood, *R. Mon. Not. R. Astron. Soc.* **138**, 495 (1968) and the review in ref. [21]
[8] A V Antonov, S N Nuritdinov and I P Osipkov, in an untranslated conference proceedings (Alma Ata, 1973) seem to have been the first to propose that all convex functionals of the distribution function should increase
A more accessible, detailed and much later paper is S Tremaine, M Henon and R Lynden-Bell, *Mon. Not. R. Astron. Soc.* (1987). However, note that there is no real H-theorem for stellar systems
[9] T Padmanabhan, *Phys. Rep.* **188**, 285 (1990)
[10] D C Heggie, *Mon. Not. R. Astron. Soc.* **173**, 729 (1975)
[11] R Bender *et al*, *Ap. J* **631**, 280 (2005). A study of M31 with the Hubble space telescope, and a typical example of current studies of black holes at the centres of galaxies
[12] M Henon and C Heiles, *Astron. J.* **73** (1964). Modestly entitled *The applicability of the third integral of motion: Some numerical experiments*. This ended up as a highly cited classic
[13] V A Antonov and S N Nuritdinov, *Vestnik Leningrad* **7**, 133 (19) is an early example of radial oscillations
S Sridhar and R Nityananda, *Mon. Not. R. Astron. Soc.* **245**, 713 (1990) has oscillations with many degrees of freedom, and also contains references to earlier work
[14] D Lynden-Bell, *Mon. Not. R. Astron. Soc.* **136**, 101 (1965)
[15] J E Barnes and L Hernquist, *Ann. Rev. Astron. Astrophys.* **30**, 705 (1990) for numerical simulations
S M Alladin and S V S Narasimhan, *Phys. Rep.* **92**, 339 (1982) for a review of the results from the impulse approximation
J E Barnes and L Hernquist, *Ann. Rev. Astron. Astrophys.* **30**, 705 (1990) for numerical simulations
Julio F Navarro, Carlos S Frenk and Simon D M White, *Astrophys. J.* **463**, 563 for the universality proposal
[16] A Toomre and J Toomre, *Sci. Am.* **229**, 38 (1973)
[17] J C B Papaloizou, *Stability of collisionless stellar systems* (Reidel, Dordrecht, 1994); see also references 1 and 2

[18] J P Ostriker and J Peebles, *Astrophys. J.* **186**, 467 (1973)
[19] S Chandrasekhar, *Ellipsoidal figures of equilibrium* (Dover, 1987)
[20] A Toomre, *Ann. Rev. Astron. Astrophys.* **15**, 437 (1977)
[21] G Bertin and C C Lin, *Spiral structure in galaxies – a density wave theory* (MIT Press, 1996)
[22] W C Saslaw, *Gravitational physics of stellar and galactic systems* (Cambridge, 1987)
[23] H J de Vega and N G Sanchez, *Statistical mechanics of the self-gravitating gas: Thermodynamic limit, unstabilities and phase diagrams*, astro.ph..1600D [23]
[24] W C Saslaw and A J S Hamilton, *Ap. J.* **276**, 13 (1984) attempted to apply thermodynamic ideas to galaxy clustering with some phenomenological success
[25] R Nityananda and T Padmanabhan, *Mon. Not. R. Astron. Soc.* **271**, 976 (1994) for the dissipative formulation of the scale-free clustering problem
S Bharadwaj, *Ap. J.* **428**, 419 (1994) for an explicitly time-dependent formulation of clustering which is not restricted to the scale-free case
[26] R Nityananda, Collisionless relaxation of dark matter, in: *Proceedings of the international Conference on Non-accelerator Particle Physics* edited by R Cowsik (World Scientific, 1995)