

Common processes and problems in disc dynamics

Scott Tremaine

Canadian Institute for Theoretical Astrophysics, University of Toronto, Canada

1 Introduction

The recognition that astrophysical discs exist was a major intellectual achievement. As Lissauer stressed at this meeting, it was more than 40 years after Galileo discovered peculiar appendages to Saturn ('two servants for the old man, who help him to walk and never leave his side') before Huygens published, as an anagram, the first correct model of the Saturn system ('it is surrounded by a thin flat ring, nowhere touching, and inclined to the ecliptic'). The long delay was due in part to the limited angular resolution of the available telescopes, but also reflects the leap of imagination needed to grasp the true nature of the first known non-spherical celestial body.

Compared with this one example of an astrophysical disc known for over 300 years, the number and variety of discs that have been discovered or inferred in just the last 30 years is remarkable: (1) Saturn's rings have been joined by lesser ring systems around the other three giant planets, all discovered since 1977; (2) there is recent strong evidence that discs are associated with many protostars and young stars (reviewed by Snell), as well as with active galactic nuclei (reviewed by Malkan); (3) it was only in the late 1960's that accretion discs were recognized to be a central ingredient of many close binary star systems, in particular cataclysmic variables and many Galactic X-ray sources; (4) although it has long been known that the solar system formed from a disc, the analysis of realistic models of protoplanetary discs, and direct observations of similar discs (*e.g.* the β Pictoris disc), began only in the last few years; (5) it is likely that discs play a crucial role in collimating the jets discovered in double radio sources, SS433, and bipolar flows from young stars.

One of the themes of this meeting has been that common dynamical processes act in astrophysical discs of various types, and hence that many of the problems confronting astrophysicists dealing with different disc systems can be solved using similar tools. In keeping with this theme, I begin by reviewing a few of the processes that appear to be central to the behaviour of astrophysical discs, and that we are confident we understand. Then I will move on to discuss some of the problems that we do not understand fully as yet, but that we believe are both common in and important to several types of astrophysical disc.

2 Common processes

The fundamental process governing the evolution of astrophysical discs can be stated simply: *energy dissipation makes discs spread*. This process was already understood by Maxwell, in his Adams' Prize essay on Saturn's rings ('as E diminishes, the distribution of the rings must be altered, some of the outer rings moving outwards, while the inner rings move inwards'). Maxwell also estimated the spreading time of the rings assuming that their viscosity was that of water, and commented humorously on the likely fate of material in the inner part of the rings: 'As for the men of Saturn I should recommend them to go by tunnel when they cross the "line" ' (see Brush *et al.* 1983).

The physical reason why disc spreading is associated with lower energy is easy to demonstrate. In the absence of external torques, discs conserve their total angular momentum, and seek to reach the lowest energy state consistent with that angular momentum. If for simplicity we consider a ring that is centrifugally supported in a fixed gravitational potential $\Phi(r)$, the specific energy and angular momentum of a disc element at radius r are given by

$$E(r) = \frac{r}{2} \frac{d\Phi}{dr} + \Phi, \quad L(r) = \sqrt{r^3 \frac{d\Phi}{dr}},$$

and it follows that $dE/dL = \Omega(r)$ where $\Omega = (r^{-1}d\Phi/dr)^{1/2}$ is the angular speed. Thus consider a simple model of disc spreading in which unit mass at radius r_2 moves outward, by acquiring angular momentum dL from a unit mass at $r_1 < r_2$ that moves inward. The resulting net change in energy is $dE = dE_1 + dE_2 = [-(dE/dL)_1 + (dE/dL)_2]dL = [-\Omega(r_1) + \Omega(r_2)]dL$, which is negative since Ω is usually a decreasing function of radius. Thus disc spreading leads to a lower energy state. In general, disc spreading, outward angular momentum flow, and energy dissipation accompany one another in astrophysical discs.

Maxwell considered only dissipation due to molecular viscosity, but we now realize that there are many different mechanisms that a disc can exploit in order to lower its energy. (1) D. N. C. Lin has argued in his talk that turbulent viscosity driven by infall plays a major role in determining the structure of the protoplanetary disc. (2) Magnetic viscosity may be important in accretion discs, although our understanding of the generation and evolution of magnetic fields in discs remains in a primitive state (*cf.* Donner's talk). (3) Lissauer has shown how efficiently density waves transport angular momentum and liberate energy in Saturn's rings – the typical rate for a strong wave turns out to be about a Gigawatt, comparable to Niagara Falls – and Ruden has argued that a strong $m = 1$ density wave may be present in circumstellar discs. (4) Small bodies ('ringmoons') imbedded in or just outside a disc can both transport angular momentum and induce effects such as the formation of sharp edges (Borderies). (5) In addition, a variety of structures within discs, including wakes (Toomre), grooves (Sellwood), bars (Frank), and other non-axisymmetric instabilities (Savonije) can provide effective angular momentum transport. Moreover, in many cases it is energetically favourable for these structures to form precisely because they liberate free energy by redistributing the disc's angular momentum.

One of the most important consequences of this redistribution process is embodied in Lynden-Bell & Pringle's (1974) remarkable formula for the effective temperature $T(r)$ of a steady-state, geometrically thin, optically thick, viscous accretion disc,

$$T(r)^4 = \frac{3GM\dot{M}}{8\pi r^3\sigma} \left(1 - \sqrt{r_*/r}\right),$$

where M is the central mass, \dot{M} is the accretion rate, σ is the Stefan-Boltzmann constant, and r_* is the inner boundary of the disc. The surface density of the disc, the equation of state, and the strength and properties of the viscosity do not enter the equation. The simplicity and power of this formula are central to our efforts to understand the spectral properties of unresolved discs in many different contexts (*cf.* talks by King, Malkan and Ruden).

Self-gravity plays an important role in many astrophysical discs. The central parameter governing the effects of self-gravity in discs is Toomre's (1964) Q parameter, usually defined as

$$Q = \frac{\sigma_R \kappa}{3.36 G \Sigma} \quad \text{for stellar discs} \quad Q = \frac{c \kappa}{\pi G \Sigma} \quad \text{for gaseous discs,}$$

where κ is the epicycle frequency, Σ is the surface density, σ_R is the radial velocity dispersion, and c is the sound speed. What might be called 'Toomre's law' states that: *discs with $Q < 1$ are unstable*. Although this result was derived only in a short-wavelength (WKB) approximation, numerical work has shown that the inequality $Q > 1$ provides an accurate necessary condition for stability of a wide range of disc models. In addition, Q has proved to be the most useful single thermometer that we have to measure the importance of self-gravity to disc dynamics.

Toomre's law can be rephrased in terms of other variables. The surface density $\Sigma(r) \approx \rho(r)h(r)$ where ρ is the mean density in the disc and h is its thickness; the epicycle frequency $\kappa \approx \Omega$, where Ω is the angular speed (this holds exactly for a Kepler disc and elsewhere is almost always correct within a factor of two). The angular speed is related to the mass interior to r , M_r , by $\Omega^2 \approx GM_r/r^3$ (this holds exactly for a spherical mass distribution and is approximately correct for a disc). If we define the mean density interior to r , $\rho_m(r) = M_r/(\frac{4}{3}\pi r^3)$, replace the dispersion σ_R or sound speed c by Ωh (approximately correct if the velocity ellipsoid is not very anisotropic and $Q \gtrsim 1$), and drop factors of order unity, then we find the simple result $Q \approx \rho_m(r)/\rho(r)$. In other words, discs are gravitationally unstable at radius r if the mean density in the disc at r exceeds the mean density of the system interior to r . In this form Toomre's law is roughly the inverse of the usual Roche limit, which states that a self-gravitating satellite of density ρ will be disrupted by tidal forces if $\rho_m/\rho \gtrsim 1$.

As this brief survey already shows, work on almost every different type of disc system has contributed to the development of disc dynamics: viscous spreading was first discussed in the context of Saturn's rings, then in models of the protoplanetary disc (Jeffreys 1924, Lüst 1952), but the complete mathematical formulation came from studies of accretion discs (Lynden-Bell & Pringle 1974). Toomre's original (1964) paper discussed stability of self-gravitating galactic discs, but he was to some extent anticipated by Safronov's (1960) study of the stability of the protoplanetary disc. Density and bending waves were investigated by C. C. Lin & Shu (1964) and Hunter & Toomre (1969) as models of spiral arms and warps in galaxies, but waves satisfying the Lin-Shu and Hunter-Toomre dispersion relations have now been exhibited much less ambiguously in Saturn's rings.

3 Common problems

Probably the single most embarrassing aspect of contemporary disc dynamics is that we do not understand why accretion discs accrete. Although accretion is almost certainly the result of energy dissipation and outward angular momentum flow, the nature of the dissipation remains mysterious. This problem was addressed by King in his review, but let me summarize some of the arguments again. Molecular viscosity is certainly negligible in accretion discs, and magnetic viscosity may be present but is not well-understood. Gravitational torques, density waves, and shocks can transport angular

momentum in inviscid discs, but so far there is no convincing model for accretion discs based on these processes.

Lastly, but most persistently, it has often been argued that turbulent viscosity must be present in accretion discs, based on the following reasoning. Plane shear flows spontaneously become turbulent if the Reynolds number $Re \approx r^2\Omega/\nu$ (r is the system size, ν is the kinematic viscosity, Ω is the shear in a planar system or the angular speed in a disc) exceeds about 10^3 . If other forms of viscosity are negligible, the Reynolds number in an accretion disc would be extremely high if the flow were laminar; hence, it is argued, turbulence must develop, with sufficient vigour to reduce Re to $\lesssim 10^3$.

There are several reasons to be suspicious of this chain of argument. First, inviscid discs are strongly stabilized by angular momentum gradients (*i.e.* the Rayleigh stability criterion $d(r^2\Omega)/dr > 0$ is satisfied, or, in the language of stellar dynamics, the epicycle frequency is real, $\kappa^2 > 0$). Second, linear stability analyses of differentially rotating discs without self-gravity (as described by Glatzel and Papaloizou) generally do not show instability unless there is a reflecting boundary condition at the edge of the disc, and hence imply that only global, not local, instabilities are present. (There is a simple physical argument that suggests why a reflecting boundary induces instability. Waves inside co-rotation have negative energy density and waves outside have positive energy density. The reflecting boundary establishes a cavity between the boundary and the forbidden zone around co-rotation; leakage through the forbidden zone excites waves of the opposite energy density and hence the wave amplitude in the cavity must grow by energy conservation.) A third reason to disbelieve arguments that high Re is always forbidden is that we have a counter-example: Saturn's rings have $Re \approx 10^{14}$! Thus the weight of theoretical evidence strongly suggests that thin, isolated, Keplerian discs without self-gravity are locally stable (although turbulence can develop in discs subjected to a rain of infalling material; see D. N. C. Lin's review). Obviously, alternatives to turbulent viscosity as the angular momentum transport process in accretion discs deserve to be investigated very thoroughly.

It is possible that the question of whether accretion discs are necessarily turbulent will only be answered when the techniques of computational fluid dynamics have improved to the point where we can investigate the high Reynolds number, high Mach number flows characteristic of these systems. However, an intriguing alternative described here by Fridman is to use laboratory studies of shallow water flows, in which the speed of surface waves replaces the sound speed of the gas, to study supersonic flows in disc geometries. I hope that experiments of this kind will be exploited further.

The second common problem that I would like to discuss is the stability of self-gravitating discs. This field of study is almost exactly a quarter-century old: in fact, 1964 and 1965 saw the publication of three seminal papers that together introduced many of the concepts that have proved to be central to the stability of self-gravitating discs. Toomre (1964) invented the Q parameter as a measure of axisymmetric stability; C. C. Lin & Shu (1964) introduced the concept of density waves described by a WKB dispersion relation; and Goldreich & Lynden-Bell (1965) showed that leading density waves can undergo strong transient amplification due to self-gravity as they are sheared into trailing waves by the disc's differential rotation ('swing amplification'). Since that time, perhaps the most important new result has been the discovery that a wide range of discs are subject to violent 'bar-like' ($m = 2$) instabilities that persist even when Q

exceeds unity and the disc is safely stable to axisymmetric disturbances (Hohl 1971, Ostriker & Peebles 1973).

The elucidation of the physics governing the bar instability has proved to be a challenging task. An arsenal of techniques has been brought to bear – including fast Fourier transform and tree-based potential solvers, ‘quiet starts’ to suppress Poisson noise, and the use of ‘softened’ gravity to suppress axisymmetric instabilities in cold discs – but it has turned out to be surprisingly difficult to carry out reliable stability studies of self-gravitating discs. One problem is that high spatial resolution is needed to follow the dynamics in the central regions, which have an important influence on disc stability; another problem is that many numerical techniques lead to spurious feedback from trailing to leading waves, which thereafter are sheared and swing amplified into larger trailing waves. Thus it is very encouraging that N -body simulations and linear normal mode calculations now find the same unstable modes in many cases (*e.g.* Sellwood & Athanassoula 1986).

The past few years have seen rapid progress in the sophistication and accuracy of numerical models of self-gravitating discs. These advances have been accompanied by semi-analytic models of the bar instability (Toomre 1981) that are based on the concepts of density waves, swing amplification, and feedback loops and that appear to provide remarkably accurate predictions of the stability properties of many disc models. An excellent summary of our present understanding of the linear stability of self-gravitating discs has been given at this meeting by Papaloizou.

As an outsider, my impression is that with these advances the problem of the stability of self-gravitating discs has largely been solved; while obstacles are still present and new phenomena certainly remain to be investigated (such as the closely related edge instabilities described by Toomre, the groove instabilities described by Sellwood, and the instabilities driven by vortensity gradients described by Papaloizou), what is left to be done would be described in military terms as ‘mopping up’. Of course, there are many cases from military history where an announcement of this kind not only demoralized the supposedly victorious troops but also was followed by enemy victory, but I hope that neither of these unfortunate events will occur, and that the stability of self-gravitating discs will soon be regarded as a solved problem.

A closely related subject in disc dynamics is spiral structure theory, which has been intertwined with the study of disc stability ever since C. C. Lin & Shu's (1964) recognition that spiral structure could be treated as a density wave. Their work led to many efforts to demonstrate that theoretical disc models admit long-lived normal modes of spiral form, and to compare the observed shapes and kinematics of spiral arms to these modes, in the hope that ultimately the fits would yield information on the surface density and velocity dispersion in observed discs. It is striking that the presentations at this meeting have contained few direct comparisons of this kind – except, of course, Lissauer's analyses of forced density waves in Saturn's rings – and I think that this is a symptom of fundamental changes in spiral structure theory.

Almost all dynamicists agree that most spiral structure reflects an underlying density wave in the old stars that comprise the bulk of the mass in the galactic disc. However, it now seems likely that in most cases it is an oversimplification to identify the spiral pattern with the most unstable normal mode of the disc, which is the pattern that would develop if a smooth disc evolved in isolation. Discs are subject to a variety of

gravitational disturbances – bars (Athanasoula's talk), encounters with other galaxies (Sundin's video paper), and molecular clouds (Toomre's talk), to mention just three – and their response to these disturbances is generally spiral, if there is differential rotation, and often very strong, if Q is not too large. Only in the case where the perturber mainly drives a single mode do we expect the spiral pattern to match the pattern of a discrete mode of the disc.

One consequence of this point of view is a new answer to the old winding problem: why does differential rotation not make the spiral arms wind up? As originally envisaged by many theorists, the answer to this question in density wave theory was that the pattern is composed of one or more modes, and discrete modes are stationary in a rotating frame and hence do not wind up. It now seems more likely that the spiral pattern remains stationary only if the disturbance is persistent (*e.g.* a central bar), and *does* wind up, if it arises from a transient disturbance (*e.g.* an encounter with another galaxy). The detailed agreement of neutral hydrogen kinematics with the predictions of models based on gas flow through a stationary density wave has sometimes been used to argue that spiral patterns must be stationary, but even density waves that are winding up can often produce a kinematic signature in the gas that closely resembles observations.

An important consequence of this change in perspective is that there is now less emphasis on *a priori* predictions of the spiral structure that should be present in a given galactic disc. Instead, we argue that there are many different mechanisms that can excite spiral structure, that we understand in broad outline how spirals originate, but that the prediction of the strength, shape and origin of the spiral pattern in a given galaxy is usually too difficult a task. Instead, it is more fruitful to concentrate, as Sancisi did in his talk, on the effects of the spiral pattern on the disc – on star formation, the energy balance of the interstellar medium, angular momentum transfer, heating the disc stars, and so forth – since these are important for galactic evolution and relatively independent of the detailed mechanism by which the spiral was formed.

A final interesting dynamical puzzle is that many types of astrophysical disc are warped. The only warps whose origin is well-understood are those associated with planetary rings (Lissauer, Borderies), but accretion discs and spiral galaxies also appear to be warped. In her talk, Sparke has advanced the promising idea that galactic warps are discrete bending modes in a disc embedded in a flattened dark halo. The basis for this idea is that the angular momentum vectors of the disc and halo are unlikely to be aligned, since the disc material has a different history from the halo material and hence has been subjected to different torques in the process of galaxy formation (Efstathiou's talk). Since the disc material is dissipative, it settles to a state in which each ring of material precesses at the same rate in the combined field of the halo and the other disc material, and this rate is simply the pattern speed of Sparke's mode. This idea dates back at least to papers by Toomre (1983) and Dekel & Shlosman (1983), and in fact has even earlier roots: the shape of Sparke's mode is simply the invariable surface first discussed by Laplace in studies of solar system dynamics. Sparke has argued that the warps may provide clues to the mass distribution in the halo and disc, but similar optimistic claims have long been made by spiral structure theorists, and there is no strong reason to believe that bending waves are more likely to be useful probes of disc structure than density waves.

Even if galactic warps arise from misalignment of the disc and halo, a number of issues remain to be resolved. It is likely that halos are not only flattened but triaxial, so that the halo not only warps the disc but distorts circular streamlines into ellipses. Unfortunately, the kinematic signature of elliptical streamlines is similar to that of inclined circular streamlines; nevertheless, it is important to investigate whether triaxial halos are compatible with kinematic observations of our own and other disc galaxies. A second issue, already mentioned by Toomre (1983) and Dekel & Shlosman (1983), is that dynamical friction may re-align the disc with the halo in times short compared to the Hubble time.

Accretion discs in binary star systems are also believed to exhibit warps, although here the evidence is less direct. The 35-day modulation of the X-ray flux from Hercules X-1 is believed to arise from occultation of the X-rays by the rim of a warped accretion disc, and the 164-day precession of the jets in SS433 is likely to arise from precession of a warped disc that feeds the jets. The most popular models for these systems require that (1) the spin angular momentum of the companion star that feeds the disc is misaligned with the orbital angular momentum of the binary system; (2) the disc is 'slaved', that is, the viscosity in the disc is so high that the disc orientation rigidly follows the orientation of the precessing companion star (this requires that the viscous diffusion time through the disc is shorter than the precession time). The required viscosity is very high (*e.g.* Katz 1980), which has the advantage that it should be possible to model the discs easily with relatively unsophisticated hydrodynamic codes to see if the model works in detail. I am personally sceptical that slaved discs are the correct explanation for the behaviour of these systems; I hope that some of the theorists at this meeting who have worked on other types of disc system will turn their attention to this problem and seek a radically different solution.

I am grateful to Alar Toomre, both for countless enlightening discussions of disc dynamics over the years and for detailed and constructive comments on this manuscript. I would also like to thank the Berkeley Astronomy Department for their hospitality while this talk was being written up.

References

- Brush, S. G., Everitt, C. W. F. & Garber, E., 1983. *Maxwell on Saturn's Rings*, 44, 135, The MIT Press, Cambridge, MA.
- Dekel, A. & Shlosman, I., 1983. In *Internal Kinematics and Dynamics of Galaxies*, IAU Symposium 100, p. 187, ed. Athanassoula, E., Reidel, Dordrecht.
- Goldreich, P. & Lynden-Bell, D., 1965. *Mon. Not. R. Astron. Soc.*, 130, 125.
- Hohl, F., 1971. *Astrophys. J.*, 168, 343.
- Hunter, C. & Toomre, A., 1969. *Astrophys. J.*, 155, 747.
- Jeffreys, H., 1924. *The Earth*, p. 55, Cambridge University Press, Cambridge.
- Katz, J. I., 1980. *Astrophys. J. Lett.*, 20, 135.
- Lin, C. C. & Shu, F. S., 1964. *Astrophys. J.*, 140, 646.
- Lüst, V. R., 1952. *Z. Naturforsch.*, 7a, 87.
- Lynden-Bell, D. & Pringle, J. E., 1974. *Mon. Not. R. Astron. Soc.*, 168, 603.
- Ostriker, J. P. & Peebles, P. J. E., 1973. *Astrophys. J.*, 186, 467.
- Safronov, V. S., 1960. *Ann. d'Astrophys.*, 23, 979.
- Sellwood, J. A. & Athanassoula, E., 1986. *Mon. Not. R. Astron. Soc.*, 221, 195.
- Toomre, A., 1964. *Astrophys. J.*, 139, 1217.

- Toomre, A., 1981. In *Structure and Evolution of Normal Galaxies*, p. 111, eds. Fall, S. M. & Lynden-Bell, D., Cambridge University Press, Cambridge.
- Toomre, A., 1983. In *Internal Kinematics and Dynamics of Galaxies*, IAU Symposium 100, p. 177, ed. Athanassoula, E. Reidel, Dordrecht.