SISSA - SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI
TRIESTE
Strada Costiera 11

ISAS - INTERNATIONAL SCHOOL FOR ADVANCED STUDIES

STABILITY AND SECULAR HEATING OF GALACTIC DISCS

Thesis submitted to the
International School for Advanced Studies, Trieste, Italy
– Astrophysics Sector –
in partial fulfilment of the requirements for the degree of

Doctor Philosophiae

Candidate:
Alessandro B. Romeo

Supervisor:
Prof. Dennis W. Sciama

Academic Year 1989/90

TRIESTE

STABILITY AND SECULAR HEATING OF GALACTIC DISCS

Thesis submitted to the
International School for Advanced Studies, Trieste, Italy
– Astrophysics Sector –
in partial fulfilment of the requirements for the degree of

Doctor Philosophiae

Candidate:
Alessandro B. Romeo

Supervisor:
Prof. Dennis W. Sciama

Academic Year 1989/90

In memory of my little brother

LUCKY

and of the last nights spent together with him

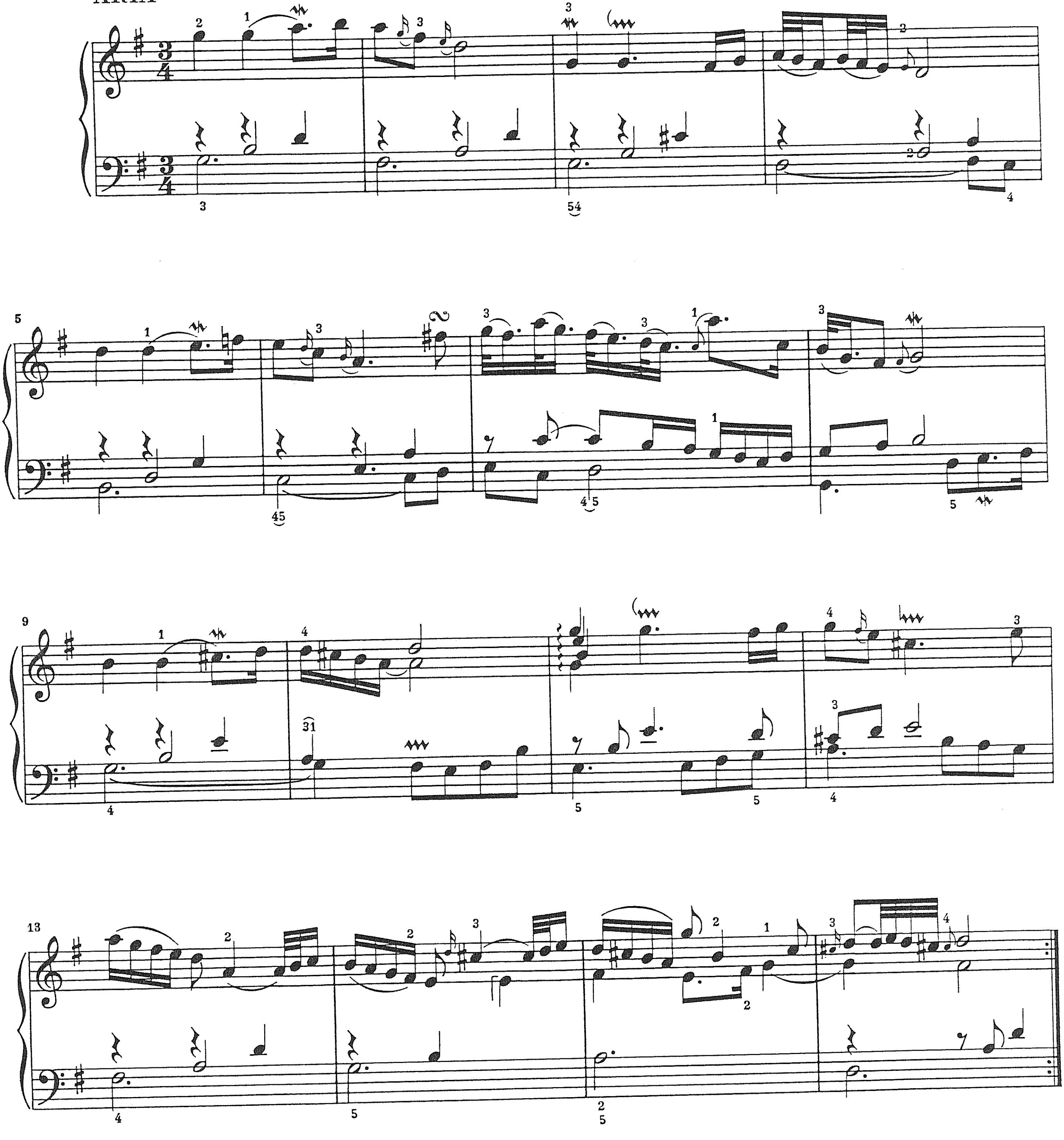
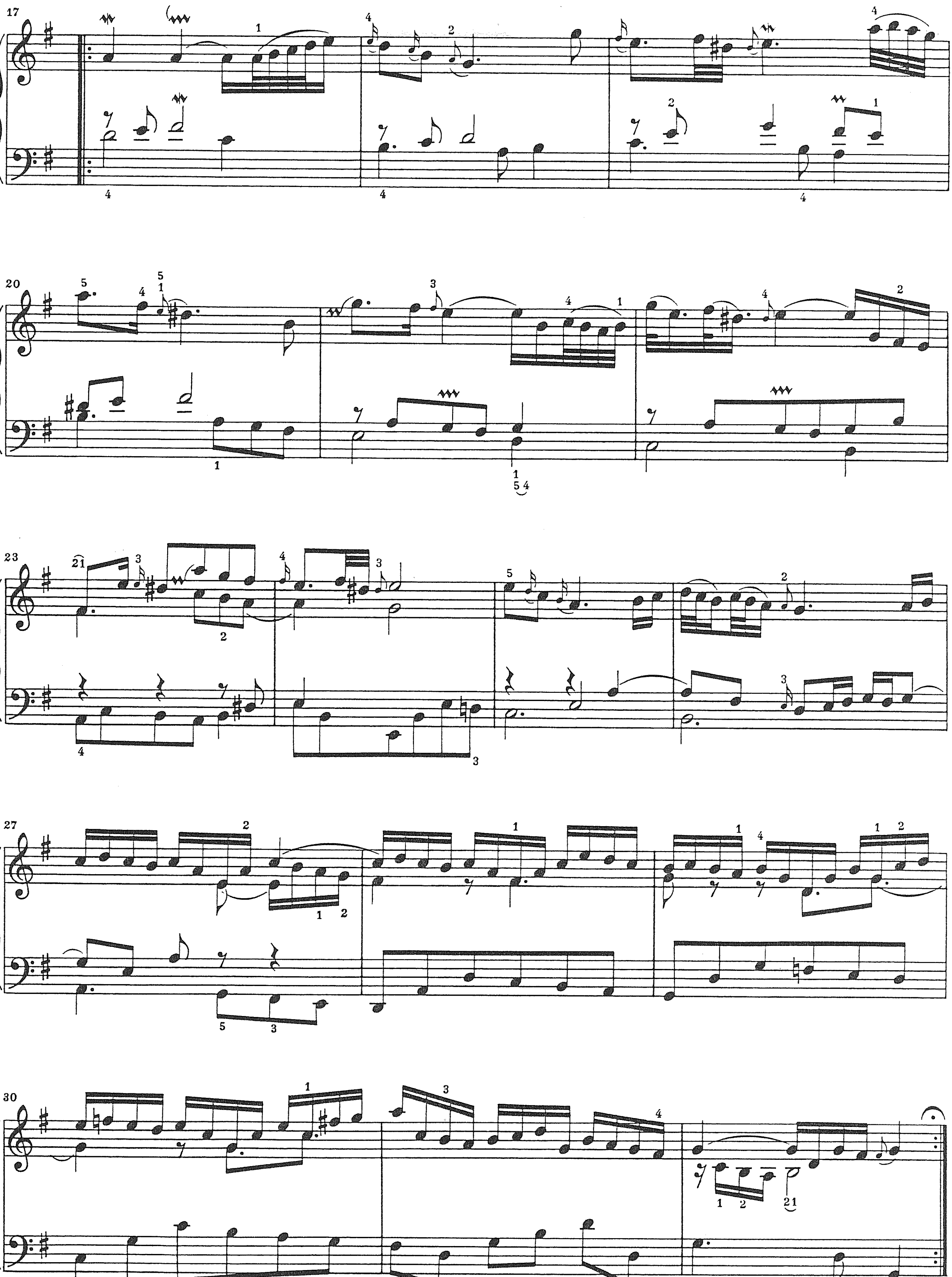
There are several ways to approach God:
one of them is to study the universe and the nature with our intellect;
another is to perceive Its presence and to feel It inside us,
loving with simplicity what life offers us
and listening to the music of Its words.

The author

ARIA

mit verschiedenen Veränderungen für Cembalo mit 2 Manualen
(Goldberg-Variationen)

BWV 988

ARIA

Acknowledgements

It is a pleasure to acknowledge Professor Dennis W. Sciama for having suggested the subject of this thesis, for valuable suggestions and fatherly encouragement during the work. In addition, Professors G. Bertin and C. C. Lin are acknowledged for interesting discussions; Professors Michel Crézé, Jens Knude, Bernard Pagel, Roland Wielen for their warm hospitality and useful discussions; and Professor Mauri Valtonen for having offered me a Post-Doctoral position at the Turku University Observatory in Finland. In my Institution there are two "LaTeX-wizards", Mauro Orlandini and Pasquale Pavone, and one "guru-TeX", Professor Massimo Calvani, to whom I feel indebted. In addition, Max deserves special thanks for his support during the period of my military service. Professor John Miller, Phil Cuddeford and Mark Dubal deserve all my gratitude for language suggestions and their friendship. I wish also to express my gratitude to all my friends, and in particular to Manolis Plionis, Milan Mijić, Vincenzo Antonuccio, José Acosta-Pulido, Sebastiano Sonego and Professor Federico Ferrini, for their hearty encouragement and help; and to all the SISSA staff for having borne me for five years!

I shall never be able to express all my gratitude to my parents for their incomparable help and moral support, so useful in such a hard period; to my little brother "Lucky", died three years ago, who spent all the last days and nights of his life together with me, while I was writing the first part of this thesis; and to the "Stella Polare che risplende in cielo anche di giorno" Astrid for her support and celestial inspiration.

Contents

Abstract

The secular evolution of galactic discs, of which the increase of the stellar velocity dispersion with age is the most striking expression from a kinematical point of view, is closely related to their stability properties because of the collective nature of such systems. In this context, however, the crucial role of collective effects is often underestimated or not properly taken into account.

We propose a global collective heating mechanism leading to a self-regulation process of the kind suggested by the spiral structure theory, when both the linear effects of wave-wave interactions and the quasi-linear effects of wave-particle interactions at the relevant resonances are taken into account. The cold interstellar gas is expected to play a crucial role in ensuring self-regulation together with the internal excitation and feedback mechanisms invoked for the maintenance of global spiral modes. As a result, the planar and vertical components of the stellar velocity dispersion are expected to have a different age-dependence. Some observational evidences in support of this qualitative prediction are also discussed.

Quantitative predictions can only be made provided a deep understanding of both local and global self-regulation mechanisms acting in galactic discs is attained. This is not an easy task at all, and in turn it requires a scrupulous investigation into their stability properties. Our contribution is thus aimed at clarifying the role of certain effects, namely those related to the presence of the cold interstellar gas and to the finite thickness of galactic discs, which are generally neglected for mathematical convenience.

Most theoretical investigations into the spiral structure of galaxies are based on one-component models, because only low-velocity dispersion stars seem to play a fundamental role. However, it has long been recognized that in some cases also the contribution of the cold interstellar gas can be important because of its low turbulent velocity dispersion, although it represents a small fraction of the total mass in normal spiral galaxies. Our analysis is devoted to such cases.

We first perform a local linear stability analysis. It is found that in some regimes of astrophysical interest the role of the cold interstellar gas can even be dominant at short wavelengths.

The results obtained in this context are used to investigate global spiral modes in regimes which are expected to be associated with normal spiral structure. We use two-component equilibrium models which incorporate the essential features of the cold interstellar gas, as suggested by some recent observational surveys. Appreciable modifications to the structure of the modes, with respect to the corresponding one-component cases, are present only when a peaked distribution of molecular hydrogen is simulated. However, even in the cases where no qualitative modifications are present, the basic states which support these modes are characterized by relatively high stellar planar velocity dispersions, i.e. by values of the local stability parameter Q larger than unity. Finally, some qualitative predictions concerning the expected structure of global spiral modes in peculiar gas-dominated regimes (where a more complicated global analysis is required) are made.

The crucial role that the cold interstellar gas can play in the dynamics and structure of early normal spiral galaxies has been shown in Chapter 7, where finite-thickness effects have not been taken into account. In view of the importance that such effects might have in the self-regulation mechanisms which are expected to operate in galactic discs and to be at the basis of their secular heating, we have tried to evaluate them. This can be done only after that their vertical structure at equilibrium has carefully been investigated.

An asymptotic analysis has thus been carried out to study the thickness-scales relevant to both the equilibrium and stability of two-component galactic discs in regimes of astrophysical interest. Two parametrizations have been introduced and examined in view of their relevance to the stability analysis which we shall perform in Chapter 9.

The results obtained in Chapter 8 as regards the vertical structure at equilibrium of two-component galactic discs are used to investigate their local linear stability properties. Under reasonable assumptions finite-thickness corrections to the local dispersion relation can be expressed in terms of two reduction factors lowering the response of the two components or, equivalently, their equilibrium surface densities. Different ansatz for such reduction factors, justified by extending the analysis performed for one-component purely stellar discs, are compared by studying the corresponding two-fluid marginal stability curves in standard star-dominated and peculiar gas-dominated regimes. It is found that the stabilizing role of finite-thickness effects can partially counterbalance the destabilizing role of the cold interstellar gas in linear regimes.

Chapter 1

Introduction

1.1 Motivations and Overview of the Problem

It is a well-known fact that the kinematical properties of nearby disc stars are systematically related to their spectral class, and in particular that the components of the stellar velocity dispersion show a tendency to increase with increasing spectral type. This has been interpreted in terms of a corresponding increase with age. Such a strong correlation between kinematical and physical properties of disc stars is indeed at the basis of their subdivision into archetypal population groups, which are generally referred to as the spiral-arm population, the young disc population, the intermediate disc population and the oldest disc population. It should be borne in mind, however, that globally the disc-component population exhibits much more drastically different features with respect to the spheroidal-component population, which reflect their different cosmological origin.

From a theoretical point of view, many efforts have been addressed to explain the observed increase of the components of the stellar velocity dispersion with age. The most diffused and currently accepted class of explanations is based on the existence of relaxation mechanisms which lead to a secular heating of galactic discs. The basic physical source responsible for such a heating process is still under debate.

In this context it should be noted that observations do not put yet any stringent constraint on the age-velocity dispersion relation, although the opposite is often claimed by observers. This is due to the large experimental uncertainties and to the presence of statistical biases ingrained in the sample selection, which often cannot properly be estimated.

For this reason only a few theoretical models have definitely been ruled out, while a lot of speculations involving hypothetical massive halo perturbers have been made mostly to find upper and lower bounds to the mass of such objects. Because of the large number of free parameters involved, these theories, although they are appealing due to their connection with the problem of dark matter in the universe, have a low level of predictability.

In all these approaches the restriction to nearly integrable situations is tacitly assumed. For strong departures from the integrability condition the relaxation is governed by the effects of Lyapunov (orbital) instability.

Even though it is often left out, the secular evolution of galactic discs, of which the increase of the stellar velocity dispersion with age is the most striking expression from a kinematical point of view, is *closely* related to their stability properties. This fact stresses the *crucial* role that collective effects play in stellar systems, or more in general in systems whose dynamics is governed by long-range interactions. Such a role is often underestimated in stellar dynamics, while perhaps too much emphasis is given to relaxation processes involving binary encounters alone.

It is indeed a well-known fact that in (electromagnetic) plasmas the rate of relaxation towards the equilibrium state can considerably be enhanced by collective effects. When such collisionless relaxation mechanisms occur, more effective heating processes become operative leading to a rapid but usually incomplete randomization of particle velocities. It is just after such a collisionless collective phase that binary encounters become effective, and lead to a slow evolution of the partially relaxed system towards the final state of thermodynamical equilibrium.

In virtue of the dynamical similarity between ordinary (electromagnetic) plasmas and "gravitational plasmas", the same phenomenon is expected to occur in stellar systems as well and to be competitive, if not *dominant*, with respect to other more commonly invoked relaxation mechanisms. However, the analysis required to describe quantitatively the relevant heating process would generally be much more complicated because of the natural inhomogeneity of stellar systems, which in particular makes the usual *local* quasi-linear approach no more suitable.

To avoid the difficulties connected with a *global* mode analysis, externally imposed and thus non-self-sustained perturbations of spiral form have generally been considered together with a local treatment in the action-angle canonical representation. Collective effects are thus not properly taken into account in this simplified approach. An effective horizontal heating of galactic discs is then produced provided these spiral waves are assumed to be recurrent transient large-scale phenomena. Despite the formal elegance of the action-angle canonical representation and the relative simplicity inherent in a local analysis, two defects characterize this approach:

- The interpretation of the theoretical predictions in terms of observable phenomena might not be straightforward. This lack of physical intuition somewhat lowers the predictability level of the theory.
- The most drastic consequence that arises from neglecting the self-consistency of the perturbations lies in the fact that internal (to the system) excitation and feedback mechanisms, *crucial* for the maintenance of global spiral modes, are not taken into account. Hence, most of the physics is missed.

For this reason it is worth formulating a *global* quasi-linear theory of spiral structure, in which the role of resonances is properly taken into account. To the above-mentioned difficulties one has to add also those deriving from the consideration of the cold interstellar gas, whose damping role in non-linear regimes cannot be disregarded as it contributes together with non-linear effects to saturate otherwise exponentially growing spiral overstabilities. Although the importance that such *self-regulation* mechanisms have in connection with the secular evolution of galactic discs has long been recognized, no quantitative theory free from the above-mentioned defects has been developed yet. This thesis is just devoted to lay the foundations for such an attempt.

1.2 Review of Observations

In this section we shall discuss the main observational surveys which have recently been performed to determine the age-velocity dispersion relation for disc stars, and it will be shown that some of them are mutually inconsistent. This fact, which essentially is due to the unavoidable use of biased samples of stars, shows that observations do not put yet any stringent constraint on the age-dependence of the components of the stellar velocity dispersion, although the opposite is often claimed by observers. Other factors which contribute to such indetermination are the large experimental uncertainties, which are difficult to estimate properly, and the indirect estimates of stellar age. Moreover, some sample-selection criteria imply a contamination by spheroidal-component stars, which have a different cosmological origin and hence should not be included. However, they are so spectroscopically distinctive that generally it is not difficult to exclude them from the analysis.

The following discussion does not pretend to be exhaustive at all. A more detailed report and comparison with other observational surveys can be found in the references cited.

In the 1970s three important observational surveys were performed by Byl (1974), Mayor (1974) and Wielen (1974).

In particular, the analysis performed by Wielen (1974) (see also Wielen 1977) is based on about 1000 stars contained in the Gliese (1969) catalogue of stars within 20 pc of the sun, for which trigonometric parallaxes accurate to $\pm 10\%$ and accurate radial velocities and proper motions are known. This sample can be plotted directly in an H-R diagram, and hence it can be divided into unambiguous age groups by choosing stars found in definite color intervals along the main sequence or near the positions of the subgiants or giant branches of clusters of known age. For each main-sequence group the average age is assumed to be about half the main-sequence lifetime of the proper stellar type (i.e., a constant star-formation rate is assumed), and to the giants are assigned the ages of the clusters along whose giant branches they most closely lie. The sample includes a large number of McCormick K and M dwarfs with known Ca II emission-line intensities, for which mean ages can be derived statistically from their relative abundances by assuming a constant star-formation rate over the lifetime of the Galaxy. These estimates can be checked by using observed average emission-line strengths in clusters of known ages. The two sets of ages turn out to be in good agreement. As a final result of this analysis, it is found that the age-velocity dispersion relation follows a ($\frac{1}{3}$–$\frac{1}{2}$)-power-law.

The estimates of the age-dependence of the components of the stellar velocity dispersion derived by Wielen (1974, 1977) have been questioned by very recent observational surveys (Carlberg et al. 1985; Knude, Schnedler Nielsen and Winther 1987; Strömgren 1987; see also: Palouš and Piskunov 1985; Lyngå and Palouš 1987; Shevelev, Marsakov and Suchkov 1989; Gómez et al. 1990; Grivnev and Fridman 1990; Knude 1990a), which however are also mutually inconsistent.

Carlberg et al. (1985) combined the Twarog (1980) sample (suitably reduced to about 250 F stars within 100 pc of the sun), for which ages and photometric distances can be determined, with astrometric data to obtain tangential velocities of a set of stars with a large age-range. The stellar age was estimated by means of a new set of stellar evolutionary sequences and isochrones incorporating substantial improvements in the input stellar physics. The resulting age-velocity dispersion relation rises fairly steeply for stars less than 6 Gyr old, thereafter becoming nearly constant with age.

Knude, Schnedler Nielsen and Winther (1987) considered the sample of stars obtained from the intersection of a photometric catalogue of A and F stars at the North Galactic Pole (Knude 1990b; cf. Teerikorpi 1990) with the AGK 3 catalogue of proper motions (Dieckvoss et al. 1975). Due to the high latitude of these stars ($b > 70°$), an accurate estimate of the plane-parallel velocities was obtained from proper motions and distances alone (i.e., without considering radial velocities). Complete subsamples of about 550 unevolved and slightly evolved F stars of solar composition roughly within 200 pc of the sun were used to study the variation of the velocity dispersions σ_U and σ_V with age. Both dispersions are found to follow power-laws very closely, but the two laws have significantly different powers, $.53 \simeq \frac{1}{2}$ and $.27 \simeq \frac{1}{4}$, respectively. The total planar velocity dispersion is found to obey roughly a $\frac{1}{2}$-power-law. The most immediate consequence of this result would be a considerable change of the shape of the velocity ellipsoid with age. More precisely, the axial ratio would change from 1 to its equilibrium value of about .5 during a period lasting 5 Gyr. The observed relaxation time seems to be much longer than that suggested by studies of early-type stars, which is of the order of the epicyclic period.

Strömgren (1987) considered a sample of about 2300 A5 to G Population I stars within 100 pc of the sun, belonging to the Olsen and Perry (1984) photometric catalogue and for which a reliable determination of radial velocities was possible. Since for all these stars adequate photometric distances and proper motions were available, galactic velocity components relative to the sun of satisfactory accuracy were derived. It is found that while the plane-parallel components of the velocity dispersion σ_U and σ_V increase markedly throughout the range 3–9 Gyr, their ratio $\sigma_V/\sigma_U \simeq .6$ showing no appreciable variation, the perpendicular component of the velocity dispersion σ_W stops at a nearly constant value for stellar ages larger than 5 Gyr.

The apparent mutual inconsistency of the results of these observational surveys shows that we are still far away from a satisfactory knowledge of the age-velocity dispersion relation. In this context a more careful estimate of selection (and probably also contamination) effects might indeed raise the confidence level of observations.

Some indirect constraints on the age-dependence of the components of the stellar velocity dispersion can be obtained by constructing consistent kinematical and chemical (e.g., Vader and de Jong 1981; Lacey and Fall 1983; see also Lacey and Fall 1985) or dynamical models of the Galactic disc (e.g., Bienaymé, Robin and Crézé 1987). See also the comparative analysis performed by Neese and Yoss (1988), and the interesting approaches developed by te Lintel Hekkert and Dejonghe (1989) and Antonuccio (1990).

1.3 Plan of the Thesis

This thesis is divided into two parts.

Part I is devoted to a review of the results already known in the literature and of the methods employed to tackle the problem of the secular heating of galactic discs. This review does not pretend to be exhaustive, mainly because the interest in this specific problem arose at the beginning of the 1950s and since then a conspicuous number of papers have been written to shed light on it. However, among these only a few contributions have been so important as to represent real turning points in the understanding of this problem. It is just to such efforts that the review is mainly devoted.

It should be borne in mind, anyway, that even more fundamental results have been found in other branches of physics, which could suitably be extended to stellar dynamics and in particular applied to the specific problem of the secular heating of galactic discs. Often the importance of such suggestions is underestimated in favour of more standard viewpoints. The first part of this thesis is indeed devoted also to clarify the *crucial* role that some not properly appreciated effects (i.e., collective effects in nearly integrable systems and the effects of Lyapunov instability in non-integrable systems) have in the secular evolution of stellar systems.

Part II expresses my own point of view about the effectiveness of the heating mechanisms described in Chapter 5 of Part I in situations of astrophysical interest. In particular, it will be stressed the fact that no such approaches take collective effects properly into account, which instead are expected to drive the dynamical evolution of galactic discs.

A *global collective* heating mechanism is then proposed in analogy with the quasi-linear theory of plasma waves, which predicts the occurrence of the so-called turbulent heating whenever an initial overstability is saturated or damped by non-linear effects. Before tackling such a complicated global non-linear analysis, in which the cold interstellar gas plays a damping role, simpler *self-regulation* mechanisms in the linear regime will be considered, in which such a cold component has instead a destabilizing role.

Finally, some results are given concerning the effects that the finite thickness of galactic discs has on their stability properties, to which the above-mentioned self-regulation mechanisms are *intimately* related. The originality of this analysis on finite-thickness corrections to the local dispersion relation lies in the fact that the stellar and the gaseous components are self-consistently taken into account in a more rigorous way than in previous works.

References

Antonuccio-Delogu V. (1990), *Preprint, NORDITA A. 61*, Copenhagen, Denmark, submitted to *M.N.R.A.S.*.

Bienaymé O., Robin A. C. and Crézé M. (1987), *A. Ap.* **180**, 94.

Byl J. (1974), *M.N.R.A.S.* **169**, 157.

Carlberg R. G., Dawson P. C., Hsu T. and VandenBerg D. A. (1985), *Ap. J.* **294**, 674.

Dieckvoss W., Heckmann O., Kox H., Günther A. and Brosterhus E. (1975), *Star Catalogue of Positions and Proper Motions North of* $-2.5°$ *Declination*, Bergedorf, Hamburg.

Gliese W. (1969), *Veröffentl. Astron. Rechen-Inst. 22*, Heidelberg, FRG.

Gómez A. E., Delhaye J., Grenier S., Jaschek C., Arenou F. and Jaschek M. (1990), *A. Ap.* **236**, 95.

Grivnev E. M. and Fridman A. M. (1990), *Sov. Astron.* **34**, 10.

Knude J. (1990a), *A. Ap.* **230**, 16.

Knude J. (1990b), "$uvby\,\beta$ Photometry of All A5–G0 Stars Brighter than $B = 11.5$ Above $b = 70°$", in preparation.

Knude J., Schnedler Nielsen H. and Winther M. (1987), *A. Ap.* **179**, 115.

Lacey C. G. and Fall S. M. (1983), *M.N.R.A.S.* **204**, 791.

Lacey C. G. and Fall S. M. (1985), in *IAU Symp.* **106**, Reidel, Dordrecht, van Woerden H. et al. eds., p. 597 (Groningen, The Netherlands, 1983).

Lyngå G. and Palouš J. (1987), *A. Ap.* **188**, 35.

Mayor M. (1974), *A. Ap.* **32**, 321.

Mihalas D. and Binney J. (1981), *Galactic Astronomy*, Freeman, San Francisco.

Neese C. L. and Yoss K. M. (1988), *A. J.* **95**, 463.

Olsen E. H. and Perry C. L. (1984), *A. Ap. Suppl.* **56**, 229.

Palouš J. and Piskunov A. E. (1985), *A. Ap.* **143**, 102.

Shevelev Yu. G., Marsakov V. A. and Suchkov A. A. (1989), *Sov. Astron.* **33**, 160.

Strömgren B. (1987), in *The Galaxy*, Reidel, Dordrecht, Gilmore G. and Carswell B. eds., p. 229 (Cambridge, UK, 1986).

Teerikorpi P. (1990), *A. Ap.* **235**, 362.

te Lintel Hekkert P. and Dejonghe H. (1989), in *Dynamics of Astrophysical Discs*, Cambridge University Press, Cambridge, Sellwood J. A. ed., p. 139 (Manchester, UK, 1988).

Twarog B. A. (1980), *Ap. J. Suppl.* **44**, 1.

Vader J. P. and de Jong T. (1981), *A. Ap.* **100**, 124.

Wielen R. (1974), in *Highlights Astron.* **3**, Reidel, Dordrecht, Contopoulos G. ed., p. 395 (Sidney, Australia, 1973).

Wielen R. (1977), *A. Ap.* **60**, 263.

Part I

GENERAL BACKGROUND AND CRITICAL REVIEW OF PREVIOUS ANALYSES

Chapter 2

Relaxation Processes in Dynamical Systems: Classical Estimates and Their Validity*

2.1 Introduction

It is a well-known fact that many-particle systems tend asymptotically to a state of thermodynamical equilibrium characterized by a Maxwellian distribution function, provided some general assumptions on the nature of the collision processes between particles are fulfilled (see, e.g., Huang 1963). The relaxation towards this equilibrium state is governed by a characteristic timescale which generally is referred to as the relaxation time of the system. In this particular context ordinary binary collisions give the dominant contribution to the relaxation process leading to such a randomization of particle velocities.

In many situations of physical interest this binary relaxation time turns out to be extremely long compared to the dynamical timescale or even to any other "observable" characteristic time. Systems whose dynamics is governed by long-range interactions, such as plasmas and stellar systems, can indeed exhibit such a peculiar feature. If this is the case, other relaxation processes toward approximate equilibrium states (stationary states[1]) are pos-

*In this introductory chapter a number of concepts and results will be given without a satisfactory discussion, that can instead be found in Chapters 3 and 4 which are intended to be a continuation of Section 2.2.

[1]The term "equilibrium" is properly referred to the thermodynamical state which is ultimately attained in the process of randomization of particle velocities. In any other time-

sible on the intermediate timescales of interest. These kinds of relaxation mechanism can actually be much more effective than ordinary two-body encounters—which need not be "physical collisions" in the case of long-range interactions—because they arise from the collective nature of long-range interactions.

We shall now stress the main differences and similarities between binary and collective relaxation processes:

- Ordinary binary encounters can be viewed as short-wavelength fluctuations of the interaction field.[2] They are essentially random, and produce small effects which accumulate slowly in time. Each encounter can lead in two directions, increasing or decreasing the energy of one of the particles, so that the cumulative effect is a random walk of each particle in velocity space, which gradually takes the whole system towards thermal equilibrium.

- Collective encounters are long-wavelength fluctuations of the interaction field.[2] They are completely analogous to binary encounters, except that one particle collides simultaneously with many particles collected together by some coherent process such as a wave. Observe that in this context the impact parameter can never be taken much smaller than the characteristic wavelength of the perturbation, and the collected bunch of particles moves with the group velocity of the wave rather than with the typical particle velocity. Again the process is random—generally it is not a random walk, as in the previous case, because memory effects cannot be disregarded—but usually much stronger, and rapidly takes the whole system towards a stationary state.

In this chapter we shall analyze in some detail the assumptions which are at the basis of classical estimates of the relaxation time in many-particle dynamical systems, with particular reference to plasmas and stellar systems. Often some of these assumptions are tacitly taken for granted, or even worse the results obtained in this context are claimed to be more general than they really are. A crucial point that must indeed be stressed right now is that all classical estimates of the relaxation time and related quantities apply to integrable systems alone. In non-integrable systems, in fact, these

independent situation the term "stationary" is commonly used. In the following we shall drop this distinction whenever no ambiguity arises, bearing in mind, however, that such a distinction does indeed exist.

[2]The fluctuations of the interaction field that a particle experiences during its motion among the other particles of the system can be Fourier-analyzed (under the assumption of local homogeneity) into wave components with different wave-vector $\boldsymbol{k}$. The distinction between binary and collective encounters is based, indeed, on the different typical wavelengths involved in the two cases, which are to be compared to the mean inter-particle distance.

calculations lose their validity because of the existence of extremely rapid phase-mixing mechanisms which make the orbits very sensitive to the initial conditions and to perturbations (Lyapunov instability).

General reference is made to Chandrasekhar (1960), who studied extensively in the 1940s the role of binary encounters in the relaxation of stellar systems; Gurzadyan and Savvidy (1986), Pfenniger (1986), who stressed the crucial point mentioned above; and to books on dynamical systems and ordinary differential equations where rigorous formulations of the ergodic theory and of the Lyapunov stability are given (e.g., Arnold and Wihstutz 1986; Arnold 1978, 1988, 1989; Arnold and Avez 1968; Bergé, Pomeau and Vidal 1986; Contopoulos 1966, 1973, 1985; Galiullin 1984; Gallavotti 1986; Grassberger 1985; Lichtenberg and Lieberman 1983; Lindblad 1983; Moser 1973; Schuster 1988; Starzhinskii 1980; Voronov 1985; Wightman 1985). See also the review papers by Aizawa, Murakami and Kohyama (1984), Aizawa et al. (1989), Akhromeyeva et al. (1989), Eckmann and Ruelle (1985), Escande (1985), Gaeta (1990), Hietarinta (1987), Konishi (1989), Martens (1984), Penrose (1979), Suzuki (1984), Vivaldi (1984). General reference is made also to the books on plasma physics and stellar dynamics listed in Section 3.1.

2.2 The Global Relaxation Time of Integrable Systems: Collisional vs Collisionless Processes

From the discussion made in the previous section it appears that systems whose dynamics is governed by long-range interactions are characterized by two relaxation times: the binary relaxation time $\tau_{\rm bin}$ associated with the relaxation towards the final thermodynamical equilibrium state, and the collective relaxation time $\tau_{\rm coll}$ associated with the relaxation towards an intermediate stationary state. It is implicitly assumed that the systems under consideration are close to a situation of integrability (and quasi-stationariness), because otherwise other relaxation mechanisms occurring on shorter timescales are to be considered. A discussion of the effects of such an orbital instability is deferred to the next section.

Another important fact to bear in mind is the characterization of the collision processes in which binary encounters and collective effects are involved. Due to the long-range nature of the interactions involved, a test particle during its motion interacts simultaneously with all the other field particles of the system, so that the question arises whether a similar distinction between binary and collective encounters does make sense. This question can be answered by observing that the main effect of the collective property of these systems consists in the presence of large-scale self-consistent mean fields (only in the gravitational case) and in the possibility of exciting self-sustained oscillation modes.[3] In this respect, differences in the large-scale dynamics between plasmas and stellar systems arise from the fact that in the former screening effects are present which make them locally neutral.

The role of collective effects can thus be singled out in a first approximation by formally dividing the potential into two parts: a part consisting in the mean field itself plus possible self-sustained perturbations of relatively long wavelengths, both produced by the "smoothed-out" distribution of matter, and a part which takes into account the fluctuations of the interaction field of relatively short and intermediate wavelengths arising from the "discrete" distribution of matter. With such a decomposition the bulk of collective effects is contained in the first part, but a non-negligible contribution (with respect to binary encounters) is given also by the intermediate-wavelength fluctuations, in which particle correlations are taken into account. The following characterization hence follows: collective effects are mainly associated with collisionless relaxation processes, even though they contribute also to collisional relaxation processes, whereas binary encounters are relevant to collisional relaxation processes alone.

It turns out that binary and collective encounters give comparable contributions to collisional relaxation processes in several situations of physical interest, as will be discussed later on in this section. The non-Markovian character of collective encounters, however, makes them difficult to treat. In this context it should be mentioned that several authors have tried to evaluate such a contribution by calculating the relaxation time or the dynamical friction in idealized models of stellar systems (e.g., Julian and Toomre 1966; Julian 1967; Thorne 1968; Kalnajs 1972). Although they claim to have taken collective effects fully into account, actually their analyses are restricted to neutral fluctuations alone, and thus disregard the most essential contribution which comes indeed from overstabilities of the system.

We shall now list the basic assumptions underlying classical estimates of the relaxation time of systems governed by long-range interactions (Chandrasekhar 1960, for the gravitational case). Some of them will be discussed in some detail, together with their physical and mathematical implications, in Chapter 3. The system is assumed to be homogeneous and in a steady state; and only instantaneous, distant, mutually independent, binary encounters in the impulsive approximation[4] (straight-line orbits) are considered.

[3]The self-consistency property concerning the mean fields and the self-sustenance property concerning the oscillation modes express the fact that these large-scale phenomena are produced by the distribution of matter in the system (Poisson equation), and thus are not externally imposed.

[4]The analysis performed by Chandrasekhar (1960) does not rely on this assumption.

As all these working assumptions lead to an overidealized model of relaxation process, only the order of magnitude of the relaxation time so derived is thought to be meaningful. For this reason we prefer to give a rough expression, as derived by Hénon (1973) in the case of stellar systems, rather than the extremely precise formulae—different possible definitions of relaxation time can in fact be given—found in Chandrasekhar (1960):

$$\tau_{\text{bin}} = \frac{v^3}{8\pi G^2 m\rho \ln N}, \tag{2.1}$$

where N is the total number of stars in the system of mass m, mean volume density ρ and typical random velocity v. Here the virial theorem[5] for stellar systems has been used to express the Coulomb logarithm $\ln(b_{\text{max}}/b_{\text{min}})$ in terms of the total number of particles in the system, b being the impact parameter. Note that the relaxation time of large stellar systems, as elliptical and spiral galaxies, largely exceeds even the age of the universe (Hubble time).

An unphysical feature deriving from the use of drastic assumptions such as the restriction to instantaneous binary encounters and the impulsive approximation consists in the fact that logarithmic divergences occur at small and large impact parameters, which are formally removed by introducing the short-range and long-range cutoffs b_{min} and b_{max}, respectively (b_{max} is different in the two cases of plasmas and stellar systems; see Section 3.3).

Several attempts have been made to relax some of the classical assumptions mentioned previously (e.g., Woolley 1954; Hénon 1958; King 1958; Lee 1968; Ostriker and Davidsen 1968; Danilov and Beshenov 1987; see also Horedt 1984) and to take collective effects into account (see the references cited in the previous paragraph). The order of magnitude of the relaxation time or related quantities turns out to be preserved, except in some "pathological" cases (e.g., Kalnajs 1972) which can be explained in the light of more sophisticated approaches (see the references listed in Footnote [7] of Section 3.3).

The use of the collisionless Boltzmann equation (Vlasov equation) for investigating galactic structure and dynamics has been justified by simple estimates of the ratio of the relaxation time by particle encounters τ_{bin} to the typical orbital time in the mean field, i.e. the crossing time T_{cross}. It is found

[5] Bear in mind that the virial theorem implies a relaxation in configuration space, which generally is attained on a dynamical timescale $T_{\text{cross}} \ll \tau_{\text{bin}}$, the latter being associated with the relaxation in velocity space (see the discussion concerning the validity of the collisionless Boltzmann equation in strictly disc systems). Large stellar systems, for instance, have a stationary shape, but have not yet reached a state characterized by a complete randomization of particle velocities.

that (see, e.g., Hénon 1973; King 1967)

$$\frac{\tau_{\rm bin}}{T_{\rm cross}} \sim \frac{N}{\ln N}, \tag{2.2}$$

so that for large N the effects of particle encounters can be neglected on a dynamical timescale.

Galactic discs, even though they may be considered very thin, are still 3-dimensional systems, and it is to 3-dimensional systems that these estimates of relaxation time apply. However, galactic discs are often approximated as strictly disc systems, in which stars are still assumed to interact by inverse-square forces but are constrained to move in a plane. Therefore, it is of some interest to consider the problem of relaxation time for strictly disc systems. It is found that (Rybicki 1972)

$$\frac{\tau_{\rm bin}}{T_{\rm cross}} \sim \frac{v}{2V}, \tag{2.3}$$

where V is the typical total particle velocity and v the typical random particle velocity relative to a local frame of rest,[6] so that the relaxation time is at most of the same order of magnitude as the crossing time, independently of the number of particles. Therefore, the collisionless Boltzmann equation can never provide an adequate description of strictly disc systems, however large.

It can be shown also that, in contrast to the 3-dimensional case, the relaxation is substantially due to close encounters, and that the cumulative effect of long-range encounters is of no more than the same order of magnitude. Thus, the assumption of independence of encounters does not present any difficulty in this 2-dimensional case, since close encounters are indeed expected to occur independently. In this regard it is interesting to note that the long-range divergence occurring in the 3-dimensional case when simple derivations are used does not occur in the 2-dimensional case, so that there is no need to introduce a long-range cutoff. All these differences can be traced back to the different statistical weighting of impact parameters in the two cases.

These arguments concerning strictly disc systems of course do not apply to actual galaxies, which are 3-dimensional systems: the validity of the Vlasov equation is rather well established in this case (see, however, Section 2.3). However, strictly disc-system approximations are commonly used in analytical and numerical treatments of disc galaxies, and hence it is necessary to judge these approximations in the light of the preceding results.

[6]This result can be expressed in a compact form relating the relaxation time $\tau_{\rm bin}$ to the epicyclic frequency κ and to the *local stability parameter* Q of infinitesimally thin, one-component, self-gravitating disc systems in differential rotation (for the definition of Q see Toomre 1964): $\kappa\tau_{\rm bin} \sim \frac{1}{2}Q$.

For analytical treatments there is no such problem of the relaxation time at all. The use of Vlasov theory is first established in view of the finite thickness of the disc, and then it is simply a question whether a 2-dimensional form of the Vlasov equation is a good approximation to the 3-dimensional form. Although this question is not trivial, at least it can be answered within the framework of Vlasov theory.

The situation concerning numerical simulations, on the other hand, is not so straightforward. 2-dimensional models of disc galaxies are often used, and the relaxation results presented here apply. Therefore, 2-dimensional N-body codes, if performed in a sufficiently precise way, do not provide faithful simulations of the Vlasov equation and thus do not apply to actual disc galaxies. Fortunately, numerical simulations are themselves subject to further approximations which tend to reduce the severity of this difficulty.

For a more detailed discussion reference is made to Rybicki (1972), Sellwood (1987), Zotov and Morozov (1987), White (1988), Schroeder and Comins (1989), Hernquist and Barnes (1990).

2.3 The Orbital Relaxation Time of Non-Integrable Systems: The Effects of Lyapunov Instability

Many numerical tests have been performed during the last three decades concerning the applicability of the classical expression for the relaxation time $\tau_{\rm bin}$ given in Section 2.2. Some of them (e.g., Miller 1964), although they were based on 3-dimensional models of stellar systems and thus were free from the criticism arisen by Rybicki (1972), showed a general and fast exponential divergence of systems starting very close to each other in the $6N$-dimensional phase space. Even when no close encounters occurred, such systems diverged exponentially at a rate much larger than that estimated by Chandrasekhar (1960).

A rough explanation of this peculiar behaviour was put forward by Miller (1966) in terms of "polarization effects in the difference medium (system)". Put in another way, he gave an original formulation of the well-known problem of Lyapunov instability in dynamical systems, but a number of wrong conclusions were drawn—e.g., regular orbits do not show an exponential but rather a linear divergence (see below). The roughness of this formulation lies indeed in the fact that it cannot discriminate integrable from non-integrable systems. See Sokolov and Kholshevnikov (1986) for some interesting considerations concerning the integrability of the N-body problem. The aim of the forthcoming discussion is just to explain these concepts in some more detail and to describe the effects of Lyapunov instability on the relaxation of dynamical systems, with particular reference to stellar systems.

In the framework of the ergodic theory dynamical systems are divided into two classes:

- Integrable systems, for which the number of isolating integrals of motion is equal to the number of degrees of freedom and the phase-space trajectories lie on N-dimensional tori.
- Non-integrable systems, whose classification is given by increasing the degree of their statistical properties: dynamical systems with divided phase space (i.e., containing both motion on N-dimensional tori and chaotic motions); ergodic systems; systems with weak and n-fold mixing; K-systems; and finally Bernoulli systems, which are a subclass of K-systems. More precisely, the classification criterion is the rate at which an initial cell of phase space tends to cover uniformly the energy hypersurface. In mixing systems an initial cell complicates its shape in such a way (i.e., preserving its volume) as to cover uniformly the energy hypersurface asymptotically. In this sense, a mixing system in a non-equilibrium state tends asymptotically to equilibrium. K-mixing systems, which possess maximally strong statistical properties, tend to such microcanonical equilibrium state at an exponential rate, the relaxation time being proportional to the Kolmogorov entropy. One of their main properties is, in fact, the decay of phase-space trajectories into beams of exponentially approaching and expanding trajectories (transversal fibers).

Several attempts have been made to relate the exponential divergence observed in the above-mentioned numerical experiments to the peculiar behaviour characterizing strongly non-integrable systems. From the point of view of the ergodic theory, this can be attained by reducing the problem of a self-gravitating N-body system to the investigation of the behaviour of a geodesic flow on a Riemannian manifold, making use of the Maupertuis principle. It is found that the negativity of the 2-dimensional curvature of this manifold is a sufficient condition for an exponential deviation of the geodesics, and the minimum of its absolute value defines an orbital relaxation time (e.g., Gurzadyan and Savvidy 1986; Gurzadyan and Kocharyan 1987a,b, 1988). Although this geometrical method for investigating the stochasticity of dynamical systems is attractive from a formal point of view, another method has been found to be more predictive from a numerical point of view. It is based on the calculation of the so-called Lyapunov characteristic exponents χ_i, which will now be discussed in some detail.[7]

[7]The Melnikov method can also be employed to study the onset of stochasticity when

An important property of non-integrable systems is to contain a definite fraction of irregular orbits, also qualified as stochastic, semi-ergodic, etc., exhibiting an exponential sensitivity to the initial conditions and to perturbations (as, e.g., the granularity of the system), which thus are rapidly amplified. In contrast, regular orbits are only linearly sensitive.[8] This intrinsic sensitivity is measured, indeed, by the Lyapunov exponents.

After a certain time the largest one, if positive, will dominate the divergence, and is therefore the best physically observable one. It determines an (individual) orbital relaxation time $\tau_{\rm orb} \sim \chi_{\rm max}^{-1}$,[9] which is not a global relaxation time of the system (see also below; for a more detailed discussion see Pfenniger 1986). The other χ_i can be computed by various numerical techniques. For an autonomous (i.e., time-independent) Hamiltonian system with n degrees of freedom there exist $2n$ Lyapunov exponents, two of which vanish for each isolating integral of motion[10] and the others appear in pairs $(-\chi_i, \chi_i)$. Each isolating integral, therefore, makes the motion robust in two directions of phase space, which are characterized by a simple linear divergence. As a consequence, actions characterize regular orbits, while the positive Lyapunov exponents characterize irregular orbits. The sum of these positive exponents turns out to be just the (specific) Kolmogorov entropy (see also the discussion made in the previous paragraph), which otherwise vanishes for regular orbits, so that only non-integrable systems evolve irreversibly.

We now turn to analyze the most direct physical implications of Lyapunov instability. The extreme sensitivity of irregular orbits to the initial conditions and to perturbations makes estimates of the binary relaxation time clearly meaningless. This difficulty also occurs when collective effects are taken into account, because the rapid phase-mixing mechanisms associated with this orbital instability can damp self-sustained oscillations of the system on timescales much shorter than the dynamical timescale, which in turn is generally comparable to the inverse of the growth rates of these oscillations—i.e., no coherent process such as a self-sustained wave can be maintained over the relevant timescales. A global relaxation time is thus no more meaningful because, apart from the impossibility of defining a binary or a collective relaxation time, one has to consider also the fact that

small perturbations are imposed on an integrable system (see, e.g., Gerhard 1985).

[8]The exponential or the linear character of this sensitivity is restricted to the linear regime. In a non-linear regime, in fact, saturation effects occur due to the presence of damping terms neglected in the linear treatment.

[9]If $\chi_{\rm max}$ is close to zero (nearly regular orbits), then a more detailed analysis, as for instance that performed by Chandrasekhar (1960), is required since the relaxation time turns out to be directly dependent on the characteristics of the system.

[10]The number of isolating integrals of motion is connected with the symmetry properties of the system, and therefore depends on the form of the (self-consistent) potential (see, e.g., Freeman 1975; Woltjer 1967).

the Lyapunov relaxation time $\tau_{\rm orb}$ can be very different from orbit to orbit since different values of $\chi_{\rm max}$ are involved. On the other hand, binary encounters and collective effects are the most effective relaxation mechanisms in integrable systems, or in those regions of nearly integrable systems where regular orbits are dominant. Moreover, classical estimates of the relaxation time still hold in non-integrable systems, provided it is defined in such a way as to refer to the exchange of isolating integrals alone.

The same considerations apply as regards the formal validity of the stochastic equations which will be investigated in Section 3.2. In non-integrable systems the diffusion in velocity space produced by Lyapunov instability cannot in fact be disregarded (see also below). In non-integrable systems the non-uniform coverage of phase space by irregular orbits makes the use of the Vlasov equation and the applicability of the Jeans theorem questionable as well (cf. Binney 1982).

The question thus naturally arises how often the departure from integrability of observed stellar systems can be neglected. But since all kinds of system exist, from systems far from integrability as small open clusters up to nearly integrable systems as spherical globular clusters, no general rule can be given. Analytical methods usually apply to nearly integrable problems, so that the successful models are strongly biased toward "nicely symmetric" situations. Attention must be paid, however, not to extrapolate superficially the results so obtained to real stellar systems. From the KAM (Kolmogorov, Arnold, Moser) theorem it follows, in fact, that asymmetries can generally destroy the principal isolating integrals of motion, since stochasticity invades phase space in a complicated manner as a perturbation grows, sometimes abruptly (Arnold diffusion). Note, however, that the Nekhoroshev theorem on the Arnold diffusion shows that under quite mild assumptions this is a very slow phenomenon (see, e.g., Benettin, Galgani and Giorgilli 1985a,b, 1987; Galgani 1985, 1988; Benettin 1986, 1988; Benettin and Gallavotti 1986; Giorgilli 1988; Mistriotis 1989).

So far we have more or less tacitly assumed to be in time-independent or at least in weakly time-dependent situations. Indeed, the "autonomous" assumption is not restrictive at all. We can, in fact, always transform a time-dependent Hamiltonian system into an autonomous Hamiltonian system by extending its phase space in such a way as to include the time-coordinate. The considerations made in the previous paragraph can thus be applied even to stellar systems in a collapse phase.

We shall now discuss some of their physical implications in relation to the theory of violent relaxation (Lynden-Bell 1967; see also Section 4.2). In some cases it may happen that the collapsing system is integrable, so that some non-classical individual stellar integrals are conserved. This situation would not be radically different from a steady-state system. But what makes the concept of violent relaxation nevertheless mostly correct is that integrable systems are very rare, so that a spherical collapse is expected to produce a large fraction of stochastic orbits. As a consequence of their exponential sensitivity to perturbations, strong phase-mixing[11] mechanisms become operative and lead to an efficient relaxation of the system. The violence of the relaxation is thus the consequence of the strongly non-integrable situations considered.

For a more detailed discussion reference is made to Pfenniger (1986).

References

Aizawa Y., Murakami C. and Kohyama T. (1984), *Prog. Theor. Phys. Suppl.* **79**, 96.

Aizawa Y., Kikuchi Y., Harayama T., Yamamoto K., Ota M. and Tanaka K. (1989), *Prog. Theor. Phys. Suppl.* **98**, 36.

Akhromeyeva T. S., Kurdyumov S. P., Malinetskii G. G. and Samarskii A. A. (1989), *Phys. Rep.* **176**, 189.

Arnold L. and Wihstutz V. (1986), in *Lyapunov Exponents*, Springer-Verlag, Berlin, Arnold L. and Wihstutz V. eds., p. 1 (Bremen, FRG, 1984).

Arnold V. I. (1978), *Ordinary Differential Equations*, MIT Press, Cambridge.

Arnold V. I. (1988), *Geometrical Methods in the Theory of Ordinary Differential Equations*, Springer-Verlag, New York.

Arnold V. I. (1989), *Mathematical Methods of Classical Mechanics*, Springer-Verlag, New York.

Arnold V. I. and Avez A. (1968), *Ergodic Problems of Classical Mechanics*, Benjamin, New York.

Benettin G. (1986), in *Stochastic Processes in Classical and Quantum Systems*, Springer-Verlag, Berlin, Albeverio S. et al. eds., p. 39 (Ascona, Switzerland, 1985).

Benettin G. (1988), in *Nonlinear Evolution and Chaotic Phenomena*, Plenum Press, New York, Gallavotti G. and Zweifel P. F. eds., p. 121 (Noto, Italy, 1987).

Benettin G. and Gallavotti G. (1986), *J. Stat. Phys.* **44**, 293.

Benettin G., Galgani L. and Giorgilli A. (1985a), *Cel. Mech.* **37**, 1.

Benettin G., Galgani L. and Giorgilli A. (1985b), in *Advances in Nonlinear Dynamics and Stochastic Processes*, World Scientific, Singapore, Livi R. and Politi A. eds., p. 1 (Arcetri, Italy, 1985).

Benettin G., Galgani L. and Giorgilli A. (1987), *Commun. Math. Phys.* **113**, 87.

Bergé P., Pomeau Y. and Vidal C. (1986), *Order within Chaos*, Wiley, New York, Hermann, Paris.

[11] This extremely rapid phase mixing is not to be confused with the ordinary phase mixing occurring in quasi-stationary nearly integrable systems, which instead proceeds only linearly in time (see, e.g., Freeman 1975).

Binney J. (1982), *M.N.R.A.S.* **201**, 15.

Chandrasekhar S. (1960), *Principles of Stellar Dynamics*, Dover, New York.

Contopoulos G. (1966), in *Space Mathematics*, Part 1, Am. Math. Soc., Providence, Rosser J. B. ed., p. 169 (Ithaca, New York USA, 1963).

Contopoulos G. (1973), in *Dynamical Structure and Evolution of Stellar Systems*, Geneva Observatory, Sauverny, Martinet L. and Mayor M. eds., p. 52 (Saas-Fee, Switzerland, 1973).

Contopoulos G. (1985), in *Chaos in Astrophysics*, Reidel, Dordrecht, Buchler J. R. et al. eds., p. 259 (Palm Coast, Florida USA, 1984).

Danilov V. M. and Beshenov G. V. (1987), *Sov. Astron.* **31**, 27.

Eckmann J.-P. and Ruelle D. (1985), *Rev. Mod. Phys.* **57**, 617.

Escande D. F. (1985), *Phys. Rep.* **121**, 165.

Freeman K. C. (1975), in *Galaxies and the Universe*, The University of Chicago Press, Chicago, Sandage A. et al. eds., p. 409.

Gaeta G. (1990), *Phys. Rep.* **189**, 1.

Galgani L. (1985), in *Chaos in Astrophysics*, Reidel, Dordrecht, Buchler J. R. et al. eds., p. 245 (Palm Coast, Florida USA, 1984).

Galgani L. (1988), in *Nonlinear Evolution and Chaotic Phenomena*, Plenum Press, New York, Gallavotti G. and Zweifel P. F. eds., p. 147 (Noto, Italy, 1987).

Galiullin A. S. (1984), *Inverse Problems of Dynamics*, Mir, Moscow.

Gallavotti G. (1986), *Meccanica Elementare*, Boringhieri, Torino.

Gerhard O. E. (1985), *A. Ap.* **151**, 279.

Giorgilli A. (1988), in *Nonlinear Evolution and Chaotic Phenomena*, Plenum Press, New York, Gallavotti G. and Zweifel P. F. eds., p. 161 (Noto, Italy, 1987).

Grassberger P. (1985), in *Chaos in Astrophysics*, Reidel, Dordrecht, Buchler J. R. et al. eds., p. 193 (Palm Coast, Florida USA, 1984).

Gurzadyan V. G. and Kocharyan A. A. (1987a), *Ap. Space Sci.* **133**, 253.

Gurzadyan V. G. and Kocharyan A. A. (1987b), *Ap. Space Sci.* **135**, 307.

Gurzadyan V. G. and Kocharyan A. A. (1988), *A. Ap.* **205**, 93.

Gurzadyan V. G. and Savvidy G. K. (1986), *A. Ap.* **160**, 203.

Hénon M. (1958), *Ann. Astrophysique* **21**, 186.

Hénon M. (1973), in *Dynamical Structure and Evolution of Stellar Systems*, Geneva Observatory, Sauverny, Martinet L. and Mayor M. eds., p. 183 (Saas-Fee, Switzerland, 1973).

Hernquist L. and Barnes J. E. (1990), *Ap. J.* **349**, 562.

Hietarinta J. (1987), *Phys. Rep.* **147**, 87.

Horedt G. P. (1984), *A. Ap.* **133**, 412.

Huang K. (1987), *Statistical Mechanics*, Wiley, New York.

Julian W. H. (1967), *Ap. J.* **148**, 175.

Julian W. H. and Toomre A. (1966), *Ap. J.* **146**, 810.

Kalnajs A. J. (1972), in *IAU Coll.* **10**, Reidel, Dordrecht, Lecar M. ed., p. 13 (Cambridge, UK, 1970).

King I. R. (1958), *A. J.* **63**, 109.

King I. R. (1967), in *Relativity Theory and Astrophysics*, Vol. 2, Am. Math. Soc., Providence, Ehlers J. ed., p. 116 (Ithaca, New York USA, 1965).

Konishi T. (1989), *Prog. Theor. Phys. Suppl.* **98**, 19.

Lee E. P. (1968), *Ap. J.* **151**, 687.

Lichtenberg A. J. and Lieberman M. A. (1983), *Regular and Stochastic Motion*, Springer-Verlag, New York.

Lindblad G. (1983), *Non-Equilibrium Entropy and Irreversibility*, Reidel, Dordrecht.

Lynden-Bell D. (1967), *M.N.R.A.S.* **136**, 101.

Martens P. C. H. (1984), *Phys. Rep.* **115**, 315.

Miller R. H. (1964), *Ap. J.* **140**, 250.

Miller R. H. (1966), *Ap. J.* **146**, 831.

Mistriotis A. D. (1989), *Physica Scripta* **39**, 33.

Moser J. (1973), *Stable and Random Motions in Dynamical Systems*, Princeton University Press and University of Tokyo Press, Princeton.

Nicolis G. (1986), *Rep. Prog. Phys.* **49**, 873.

Ostriker J. P. and Davidsen A. F. (1968), *Ap. J.* **151**, 679.

Penrose O. (1979), *Rep. Prog. Phys.* **42**, 1937.

Pfenniger D. (1986), *A. Ap.* **165**, 74.

Rybicki G. B. (1972), in *IAU Coll.* **10**, Reidel, Dordrecht, Lecar M. ed., p. 22 (Cambridge, UK, 1970).

Schroeder M. C. and Comins N. F. (1989), *Ap. J.* **346**, 108.

Schuster H. G. (1988), *Deterministic Chaos*, VCH, Weinheim.

Sellwood J. A. (1987), *Ann. Rev. A. Ap.* **25**, 151.

Sokolov L. L. and Kholshevnikov K. V. (1986), *Sov. Astron. Lett.* **12**, 235.

Starzhinskii V. M. (1980), *Applied Methods in the Theory of Nonlinear Oscillations*, Mir, Moscow.

Suzuki M. (1984), *Prog. Theor. Phys. Suppl.* **79**, 125.

Thorne R. M. (1968), *Ap. J.* **151**, 671.

Toomre A. (1964), *Ap. J.* **139**, 1217.

Vivaldi F. (1984), *Rev. Mod. Phys.* **56**, 737.

Voronov A. A. (1985), *Basic Principles of Automatic Control Theory*, Mir, Moscow.

White R. L. (1988), *Ap. J.* **330**, 26.

Wightman A. S. (1985), in *Regular and Chaotic Motions in Dynamic Systems*, Plenum Press, New York, Velo G. and Wightman A. S. eds., p. 1 (Erice, Italy, 1983).

Woltjer L. (1967), in *Relativity Theory and Astrophysics*, Vol. 2, Am. Math. Soc., Providence, Ehlers J. ed., p. 1 (Ithaca, New York USA, 1965).

Woolley R. v. d. R. (1954), *M.N.R.A.S.* **114**, 514.

Zotov V. M. and Morozov A. G. (1987), *Sov. Astron. Lett.* **13**, 136.

Chapter 3

Collisional Relaxation Processes: The Fokker-Planck Approach and Alternative Descriptions

3.1 Introduction

The Fokker-Planck equation has widely been applied to the study of plasmas and stellar systems for describing the evolution of the one-particle distribution function when collisional effects are taken into account. Its derivation is based on a number of assumptions which in general are not clearly specified or are tacitly taken for granted.

In this chapter we shall inquire into the validity of this equation by analyzing the underlying assumptions in some detail. The following discussion does not pretend to be exhaustive, because most of the mathematical concepts and techniques inherent in this description are subtle and cannot thus properly be expressed and discussed in this context. Some other more general and/or correct, but less predictive, approaches will also be described. The three basic approaches discussed in this chapter are schematically compared in Table 3.1.

Reference is made to books on probability theory and stochastic processes where rigorous derivations of the Fokker-Planck equation (cf. Kolmogorov forward equation) are given (e.g., Cox and Miller 1965; Feller 1970, 1971; Friedman 1975, 1976; Gardiner 1985; Kac and Logan 1987; Montroll and West 1987; Nelson 1967; van Kampen 1981; Ventsel 1983; Wax 1954). See also the review papers by Haken (1975), Kandrup (1980), Li (1986), Padmanabhan (1990), Putterman and Roberts (1988), Spohn (1980). General reference is made also to books on plasma physics (e.g., Boyd and Sanderson 1969; Hinton 1983; Ichimaru 1980; Krall and Trivelpiece 1973; Schmidt 1979; Sivukhin 1966) and stellar dynamics (e.g., Binney and Tremaine 1987; Saslaw 1985; Spitzer 1987).

The evolution of the one-particle distribution function $f(\boldsymbol{x}, \boldsymbol{v}; t)$ in the 6-dimensional phase space μ is described by the Boltzmann equation

$$\frac{df}{dt} \equiv \frac{\partial f}{\partial t} + [f, H] = \left(\frac{\partial f}{\partial t}\right)_{\mathrm{enc}}, \tag{3.1}$$

where H is the Hamiltonian of the system, self-consistently related to f via the Poisson equation, the symbol $[\ldots, \ldots]$ denotes the Poisson brackets, and the term $(\partial f/\partial t)_{\mathrm{enc}}$ represents the contribution of particle encounters to the time-variation of f. In general this is an integro-differential equation, which can be reduced to a differential equation only by making certain assumptions on the collision processes.

For instance, if in the relevant timescale collisional effects can be neglected —as in high-temperature plasmas and large stellar systems, where the relaxation time largely exceeds the dynamical timescale—we recover the Vlasov equation

$$\frac{\partial f}{\partial t} + [f, H] = 0\,. \tag{3.2}$$

It states that f is a conserved quantity along the particle orbit. For steady-state systems this implies that f is a function of the isolating integrals of motion alone (Jeans theorem; see, e.g., Chandrasekhar 1960; Lynden-Bell 1962; for some controversial points see Section 2.3).

The collisionless Boltzmann equation should not be confused with the Liouville equation

$$\frac{\partial f^{(N)}}{\partial t} + [f^{(N)}, H] = 0\,, \tag{3.3}$$

which instead describes the evolution of the N-particle distribution function $f^{(N)}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N; \boldsymbol{v}_1, \ldots, \boldsymbol{v}_N; t)$ in the $6N$-dimensional phase space Γ without any assumption on the collisional nature of the system.

3.2 Markovian Stochastic Approaches: The Concept of Dynamical Friction

When the effect of encounters is taken into account, the only way to reduce the Boltzmann equation to a differential equation, i.e. to make it operate locally in time, is to require that the system has no memory in the collision processes (ergodic assumption), so that a test particle suffers random displacements in velocity space generated by the fluctuating part of the interaction field in a manner that can be described in terms of a random walk. This is equivalent to state that the increments of velocities are regarded as stochastically independent in disjoint time intervals.

In systems whose dynamics is governed by long-range interactions, such as indeed plasmas and stellar systems, this assumption may not be well justified because correlations among particles cannot be disregarded a priori. Collective effects, in fact, always play an important or even dominant role in these two kinds of system—in the gravitational case even more than in the electromagnetic case because no Debye shielding length (i.e., no local neutrality) exists. We then keep this as a working assumption, bearing in mind that the resulting evolution equation neglects collisional collective effects. A different approach which takes them fully into account will be discussed in Section 3.3.

The standard approach consists in deriving a diffusion process in velocity space (see, e.g., Chandrasekhar 1943a,b,c,d). The evolution of the distribution function $f(\boldsymbol{x}, \boldsymbol{v}; t)$ is then written in the form of a Fokker-Planck-type equation:

$$\frac{\partial f}{\partial t} + [f, H] = \boldsymbol{\nabla}_v \cdot (q \boldsymbol{\nabla}_v f + \eta f \boldsymbol{v}) , \tag{3.4}$$

where we recall that H is the Hamiltonian related to the smoothed-out distribution of matter, $q = q(\boldsymbol{v})$ is the diffusion coefficient and $\eta = \eta(\boldsymbol{v})$ the coefficient of dynamical friction appearing in the Langevin equation[1]. These two coefficients are related by the condition that a given Maxwellian distribution function f_M remains invariant in time, i.e. $(\partial f_M / \partial t)_{enc} \equiv 0$, so that η turns out to be connected with the reciprocal of the relaxation time of the system.

More standard forms of the Fokker-Planck equation are the following:

$$\frac{\partial f}{\partial t} + [f, H] = -\frac{\partial}{\partial v_i} \left(\frac{\langle \Delta v_i \rangle}{\Delta t} f \right) + \frac{1}{2} \frac{\partial^2}{\partial v_i \partial v_j} \left(\frac{\langle \Delta v_i \Delta v_j \rangle}{\Delta t} f \right) \tag{3.6}$$

(see, e.g., Hénon 1973), where the averages are taken with respect to a transition-probability distribution of gaussian type; and

$$\frac{\partial f}{\partial t} + [f, H] = -\frac{\partial}{\partial v_i} (A_i f) + \frac{1}{2} \frac{\partial^2}{\partial v_i \partial v_j} (D_{ij} f) \tag{3.7}$$

[1]The Langevin equation represents an attempt to rewrite the equations of motion in many-particle systems in such a way as to split the contribution of the smoothed-out distribution of matter from the effect of the fluctuating part of the interaction field (of the perturbers):

$$\dot{\boldsymbol{v}} = \boldsymbol{a} + \boldsymbol{R} - \eta \boldsymbol{v} , \tag{3.5}$$

where $\boldsymbol{a}$ is the systematic acceleration produced by the former, $\boldsymbol{R}$ and $-\eta \boldsymbol{v}$ the stochastic acceleration and the dynamical friction due to the action of the latter, respectively.

found in most plasma-physics textbooks, where A_i and D_{ij} are referred to as the dynamical-friction vector and the diffusion tensor, respectively.[2]

Sometimes it is more useful, especially in the case of stellar systems, to use the action-angle variables as the proper canonical coordinates in virtue of their physical meaning (the actions correspond to adiabatic invariants). The orbit-averaged Fokker-Planck equation retains exactly the same form with the velocity components v_i replaced by the actions J_i. See Inagaki and Lynden-Bell (1990) for an interesting discussion concerning the derivation of approximate solutions of this evolution equation making use of a generalized variational principle for collisional stellar dynamics.

Now it should be borne in mind that the sample paths of every diffusion process are continuous (with probability one), so that the random velocity of the test object varies continuously in the course of time. Because of this fact, the Fokker-Planck approach seems not to be suitable to describe systems governed by the electromagnetic or the gravitational interaction, since a close encounter of a test particle with a field particle is able to produce a large change of velocity within a small time interval, clearly contradicting the notion of continuity.

This difficulty arises because the ergodic assumption is indeed more general than the choice of a diffusion process. In other words, the phrase "stochastically independent events in disjoint time intervals" is not equivalent to the property "diffusion", since there exist infinitely many Markov processes which share only the first property but not the second (these concepts will be explained in some more detail in the forthcoming discussion). An oversimplified hierarchical scheme for the stochastic processes discussed in this section is shown in Figure 3.1.

If the stochastic variations in velocity space can approximately be described as a Markov process on the whole, the question arises whether it is uniquely determined by the properties of the fluctuating part of the interaction field, and whether the jump phenomena mentioned above can be explained by an analysis of its sample paths alone without employing additional assumptions. Several attempts have been made to answer this question, and more in general to formulate a statistical theory in the framework of stellar dynamics, but we are still far away from a satisfactory understanding (e.g., Chandrasekhar 1941, 1943a,b,c,d, 1944a,b; Chandrasekhar and von Neumann 1942, 1943; see also Chandrasekhar 1960 for a review; Camm 1963; Lee 1968; Tscharnuter 1972).

Bearing this fact in mind, we shall now briefly discuss some basic ideas which lead to the derivation of another, still not completely satisfactory, stochastic differential equation for plasmas and stellar systems. Let us first

[2] This definition is different from that given by Chandrasekhar, apart from the obvious generalizations inherent in this last equation.

reformulate more definitely the conditions of a random walk: the total increment of velocity within the time interval $(0, t)$, where t is much larger than the characteristic time T during which an elementary fluctuation of the random interaction field takes place, can be written as a sum of a great number of independent[3] random variables representing the small (at least on the average) displacements in velocity space after the small amount of time T has passed.

Now the crucial point is the determination of the distribution law of this sum. This problem, however, is solved exhaustively by the so-called extended central limit theorems of probability theory, which were essentially established by Lévi and Khintchine in the 1930s. In general, a convergence to the normal (i.e., gaussian) distribution and hence a diffusion is expected, but the probability distribution of the random gravitational field is shown to be asymptotically the Holtsmark distribution, whose characteristic function (Fourier transform) is $h(\rho) = \exp(-a|\rho|^{3/2})$. (However, even this distribution contains some unphysical features; see, e.g., Chandrasekhar 1941, 1943a, 1960; Feller 1971; Petrovskaya 1986; see Antonuccio-Delogu and Atrio-Barandela 1990 for the derivation of a modified Holtsmark law describing stochastic field fluctuations in galaxies and clusters.)

The Holtsmark distribution is a symmetric stable distribution[4], and belongs therefore to its own domain of attraction. This leads necessarily to a distribution law for the total increment of velocity within the time interval $(0, t)$ with characteristic function $\varphi(\omega) = \exp(-\sigma t|\omega|^{3/2})$, where $\sigma = a\sqrt{T}$.[5] φ is the Fourier transform of the transition function belonging to the Markov process which is called the stable process with characteristic exponent 3/2. From general theorems on Markov processes it follows that its sample paths are right-continuous (i.e., jump phenomena occur). It can also be shown that

[3]The ergodic property characterizing random walks "no memory of the initial state after a macroscopically small time interval" or equivalently "stochastically independent events in disjoint time intervals" is stronger than the Markovian property "future development dependent only on the present state, and not on the past history of the process or on the manner in which the present state was reached".

[4]Stable distributions play an important role in the theory of stochastic processes as a natural generalization of the normal distribution. The importance of the normal distribution $\mathcal{N}$ is largely due to the central limit theorem. Let $X_1, \ldots, X_n$ be mutually independent variables with a common distribution $\mathcal{F}$ having zero expectation and unit variance, and define $S_n \equiv X_1 + \cdots + X_n$. The central limit theorem states that the distribution of $S_n n^{-1/2}$ tends asymptotically to $\mathcal{N}$. For distributions without variance similar limit theorems (extended central limit theorems) can be formulated, but the norming constants must be chosen differently. The interesting point is that all stable distributions and no others occur as such limits.

[5]The use made by Chandrasekhar (1941) of a gaussian distribution is justified by the fact that the modified Holtsmark distribution, that he derived for avoiding some unphysical divergences at small distances, has finite variance and hence falls into the cases considered by the central limit theorem. The corresponding evolution equation is thus of diffusion type.

the mean number of jumps increases to infinity as their heights converge to zero, and conversely. This is a very important property because, if one identifies these jumps as the results of far and close encounters respectively, the importance of far encounters is emphasized on the one hand, but spontaneous large changes in velocity due to close encounters are also possible on the other hand.

So far dynamical friction acting in a purely systematic manner has been ignored. Taking it into account, a stochastic equation for the stable Markov process with characteristic exponent 3/2, analogous to the diffusion equation derived by Chandrasekhar, can be written down (Tscharnuter 1972):

$$\frac{\partial f}{\partial t} + [f, H] = -(-\nabla_v^2)^{3/4}(\sigma f) + \nabla_v \cdot (\eta f \boldsymbol{v}) , \tag{3.8}$$

where σ appears to play a role similar to the diffusion coefficient q in the Chandrasekhar diffusion equation. The $\frac{3}{4}$-power of the Laplace operator ∇_v^2 is uniquely defined in the sense of its spectral representation: this elliptic pseudo-differential operator (see, e.g., Hörmander 1976, 1983, 1985; Taylor 1981) acts on a given function f in such a way that the Fourier transform of $(-\nabla_v^2)^{3/4} f(\boldsymbol{v})$ is the function $|\boldsymbol{\omega}|^{3/2}\hat{f}(\boldsymbol{\omega})$, where $\hat{f}$ denotes the Fourier transform of f. The correct derivation of this term in the previous evolution equation is not simple and requires sophisticated functional analysis techniques (semi-group theory).

Since it seems impossible to solve the Tscharnuter stochastic equation in the whole 6-dimensional phase space analytically as well as numerically, its investigation relies on the assumption of spatial homogeneity, i.e. $[f, H] \equiv 0$, which is a quite drastic assumption for stellar systems. Bear in mind, however, that this assumption is already inherent in the derivation of the Holtsmark distribution, and is used also to calculate explicitly the coefficients of diffusion and dynamical friction in the Fokker-Planck equation. In contrast to the Fokker-Planck equation, it can be shown that a Maxwellian distribution function f_{M} is not an invariant distribution of the given Markov process, i.e. $(\partial f_{\mathrm{M}}/\partial t)_{\mathrm{enc}} \not\equiv 0$. This fact causes troubles since the relation between σ and η cannot directly be established, and might have important physical implications.

Now the question arises which of the two Markovian stochastic approaches described here is more correct from a physical point of view. Both of them, in fact, seem to give rise to unphysical features: the Fokker-Planck approach predicts continuous sample paths, whereas the Tscharnuter approach, although it is characterized by right-continuous sample paths (jump phenomena), predicts too high probabilities for high-field values (the Holtsmark distribution has infinite variance). In other words, they seem to overestimate the effect of distant and close encounters, respectively.

The unphysical features present in the two cases are intimately related and seem unavoidable. In fact, any finite-variance distribution for the interaction field gives rise to a normal distribution for the total increment of velocity (central limit theorem), which is at the basis of the Fokker-Planck approach for diffusion processes. Any attempt to regularize the Holtsmark distribution for avoiding high-field divergences falls therefore into this case. It is difficult to judge whether these are real unphysical features because, for instance, the notion of continuity (with probability one) of the sample paths is not intuitive at all (it implies the existence of a stochastic process with continuous sample paths equivalent to that physically observed). Apart from these difficulties, which however should be borne in mind, the Fokker-Planck equation has widely been used in view of its higher level of predictability.

3.3 General Statistical Approach: The Concept of Dynamical Friction Revised

As mentioned in Section 3.2, an unphysical feature inherent in the ergodic assumption, used to reduce the Boltzmann equation to a differential equation, lies in the fact that this approach does not take collisional collective effects into account, which cannot be disregarded a priori in systems whose dynamics is governed by long-range interactions. A different description which overcomes this difficulty and is not restricted by the assumption of spatial homogeneity, inherent in the two previous approaches, will now be discussed.

Statistical correlations among particles arise both from the initial probability distribution and the dynamics. It seems plausible that in most cases the disorganized motions of the particles will disrupt groups which were initially nearby, quickly erasing the original correlations. The resulting correlations will then be determined by the dynamics and the single-particle distribution function alone.

In order to obtain a closed theory in which the one-particle distribution function is the only variable, Gilbert (1968) (see also Gilbert 1972) made the basic assumption that the probability distributions have evolved from initially uncorrelated states—although it is possible to imagine also quite different situations. The strategy he adopted in its theory on collisional collective processes in stellar systems consists in a decoupling of the BBGKY (Bogoliubov, Born, Green, Kirkwood, Yvon) hierarchy[6] accomplished by a

[6]The BBGKY hierarchy of equations is obtained by integrating the Liouville equation over the phase space of all but s particles ($1 \leq s \leq N-1$). It turns out that the evolution equation for $f^{(s)}$ involves $f^{(s+1)}$ as well, so that these $N-1$ equations are all coupled. The

perturbation series expansion in powers of $1/N$, the inverse of the total number of stars in the system, with the aid of certain combinations of the distribution functions called the correlation functions $g^{(s)}$ (they represent multiparticle correlations).

A system of two coupled evolution equations of integro-differential type for $f^{(1)}$ and $g^{(2)}$ is thus obtained, which in principle may simultaneously be solved. The simplest situation occurs when the system is in equilibrium with respect to purely collective motions, and the only time-dependence is through the slow, secular effects of stellar encounters. In that case a kinetic equation for $f^{(1)}$ alone can formally be derived (the corresponding equation of plasma physics is the Balescu-Lenard equation).

The two coupled evolution equations for $f^{(1)}$ and $g^{(2)}$ derived for stellar systems are similar but not identical to the corresponding plasma equations. The differences come about because the latter are based upon a perturbation series expansion in powers of the inverse of the number of electrons contained in a Debye sphere (see, e.g., Rostoker and Rosenbluth 1960), which is independent of the total number of electrons in the system, usually taken as infinite. On the other hand, the role of the Debye screening length for a stellar system is played by the linear dimensions of the system itself. The number of stars in a Debye sphere is thus equal to the total number of stars, so that N has a dual meaning.

Another point which should be stressed in this context is the fact that, while it is reasonable to assume spatial homogeneity in plasmas (not subject to strong external fields), the same is not true in self-gravitating systems, because the absence of screening effects makes them naturally inhomogeneous on large scales. The mathematical counterpart of this different physical feature is expressed by the fact that an explicit elimination of $g^{(2)}$ in terms of $f^{(1)}$ cannot be achieved in the gravitational case, but it is still possible to construct a formal solution (determining $g^{(2)}$ as a functional of $f^{(1)}$) and to interpret it in terms of the underlying physical processes.

The physical content of this formal solution can more easily be understood in terms of the auxiliary concept of gravitational polarization (for the plasma analogue of this effect see, e.g., Balescu 1960). It represents the response of the system to the gravitational field of a selected star moving in a specified orbit. In calculating this response one ignores collisional effects entirely and treats the field of a selected star as a small externally applied perturbation. The polarization is the change in the single-particle distribution function that this perturbation induces.

The final result of this analysis is that collisional effects in stellar systems, i.e. dynamical effects of order $1/N$ (this ordering holds provided the system is in equilibrium with respect to purely collective motions), may be divided

closure of the system is given by the Liouville equation itself, as it involves $f^{(N)}$ alone.

into two distinct phenomena:

- The gravitational force exerted on each star by the polarization (wake) it induces, which may be termed polarization drag (it represents a more precise formulation of the concept of dynamical friction[7], first introduced by Chandrasekhar 1943b). It is expected to retard the motion of a test star, the deceleration being directly related to its velocity. Moreover, since the polarization induced by a given star is proportional to its mass, we expect heavy stars to be slowed more effectively than light stars. It may be worth observing also that, since the characteristic distance over which the gravitational polarization extends is of the same order as the linear dimensions of the stellar system, a given star cannot be thought of as being affected only by stars in its immediate neighbourhood.

- The effect upon each star of the random fluctuating field resulting from the superposition of the fields of the other stars, each modified by its own polarization. These stars are to be considered to move in unperturbed orbits and not to respond to the influence of the test star under consideration. It may be termed statistical acceleration. The statistical acceleration acting on a star increases on the average its energy and, because of the identity between the inertial and the gravitational mass, affects all stars in the same way.

A nice description of these two effects can be found in Hénon (1973) as well. A similar decomposition exists also for the Fokker-Planck collisional term calculated under the assumptions of spatial homogeneity and binary encounters. There, however, the polarization term is incompletely calculated and the statistical term consists of a superposition of "bare" (i.e., not modified by polarization effects) inter-particle forces.

To stress the contribution of these two effects to the collisional relaxation of stellar systems, the resulting kinetic equation for the evolution of

[7]The important role played by the dynamical friction especially in situations of astrophysical interest has stimulated many numerical (e.g., White 1976; Keenan 1979; Faulkner and Coleman 1984; Palmer and Papaloizou 1985; Byrd, Saarinen and Valtonen 1986; Bontekoe and van Albada 1987; Zaritsky and White 1988; Hernquist and Weinberg 1989; Valtaoja 1990) and analytical works confined to nearly integrable systems (e.g., Binney 1977; Palmer and Papaloizou 1982, 1985; Mulder 1983; Palmer 1983; Faulkner and Coleman 1984; Tremaine and Weinberg 1984; Kashlinsky 1986; Weinberg 1986, 1989; Narasimhan and Ballabh 1988; Bekenstein and Zamir 1990) in addition to those listed in Section 2.2. In particular, most of them stress the crucial contribution of the resonances where precisely in an integrable system slightly perturbed irregular orbits first appear. For a more exhaustive and detailed discussion reference is made to Alladin and Narasimhan (1982), Manorama (1986), Tremaine (1981).

the single-particle distribution function can be written in the following form:

$$\frac{\partial f}{\partial t} + [f, H] \equiv \frac{\partial f}{\partial t} + \boldsymbol{v} \cdot \boldsymbol{\nabla}_x f + \left(1 - \frac{1}{N}\right) \boldsymbol{a} \cdot \boldsymbol{\nabla}_v f = \left(\frac{\partial f}{\partial t}\right)_{\text{Pol}} + \left(\frac{\partial f}{\partial t}\right)_{\text{Stat}}, \quad (3.9)$$

where the factor $(1 - 1/N)$ reflects the fact that the test star feels the average gravitational acceleration due to the other $N - 1$ stars, and not the total average gravitational acceleration. The competition between polarization and statistical effects is expected to lead to a relative concentration of heavier stars in the central regions of stellar systems and lighter stars in their outer parts, i.e. an approach to equipartition (see, e.g., Hénon 1973; Lynden-Bell 1973).

The theory sketched here takes thoroughly into account the effects of collective interactions and spatial inhomogeneity which are absent from more elementary treatments. As a consequence, no long-range divergence appears as it does in the Fokker-Planck approach (when the coefficients of diffusion and dynamical friction are evaluated according to the standard treatment; see, e.g., Braginskii 1965). It is interesting to note that, although the equations of plasma physics and stellar dynamics are very similar, the mechanism for the elimination of this divergence is different in the two cases. In plasmas the repulsive inter-particle force results in Debye-shielding which cuts the force off, eliminating the divergence. In stellar systems the attractive interstellar force results in anti-shielding or amplification of the bare gravitational force of a star. This tends to make the divergence worse, and it is only the limited spatial extent of the system that finally removes it.

There is, however, in the present theory a divergence at small distances, arising because the perturbation series expansion in powers of $1/N$ is non-uniformly convergent. The physics behind this is quite simple, since this divergence is precisely equivalent to that occurring in elementary Fokker-Planck treatments, where the impulsive approximation (straight-line orbits) fails at small impact parameters (see, e.g., Braginskii 1965). The corresponding failure in the Gilbert approach can easily be expressed in terms of $f^{(1)}$ and $g^{(2)}$. The suppression of this unphysical divergence can be achieved by retaining all terms in $g^{(2)}$ appearing in the two original exact coupled equations. Simpler approaches, in which the true inverse-square force is replaced by an effective fictitious force, can also be used.

References

Alladin S. M. and Narasimhan K. S. V. S. (1982), *Phys. Rep.* **92**, 339.

Antonuccio-Delogu V. and Atrio-Barandela F. (1990), "Stochastic Field Fluctuations in Galaxies and Clusters", submitted to *Ap. J. Lett.*.

Balescu R. (1960), *Phys. Fluids* **3**, 52.

Bekenstein J. D. and Zamir R. (1990), *Ap. J.* **359**, 427.

Binney J. (1977), *M.N.R.A.S.* **181**, 735.

Binney J. and Tremaine S. (1987), *Galactic Dynamics*, Princeton University Press, Princeton.

Bontekoe Tj. R. and van Albada T. S. (1987), *M.N.R.A.S.* **224**, 349.

Boyd T. J. M. and Sanderson J. J. (1969), *Plasma Dynamics*, Nelson, London.

Braginskii S. I. (1965), in *Reviews of Plasma Physics* **1**, Consultants Bureau, New York, Leontovich M. A. ed., p. 205.

Byrd G. G., Saarinen S. and Valtonen M. J. (1986), *M.N.R.A.S.* **220**, 619.

Camm G. L. (1963), *M.N.R.A.S.* **126**, 283.

Chandrasekhar S. (1941), *Ap. J.* **94**, 511.

Chandrasekhar S. (1943a), *Rev. Mod. Phys.* **15**, 1.

Chandrasekhar S. (1943b), *Ap. J.* **97**, 255.

Chandrasekhar S. (1943c), *Ap. J.* **97**, 263.

Chandrasekhar S. (1943d), *Ap. J.* **98**, 54.

Chandrasekhar S. (1944a), *Ap. J.* **99**, 25.

Chandrasekhar S. (1944b), *Ap. J.* **99**, 47.

Chandrasekhar S. (1960), *Principles of Stellar Dynamics*, Dover, New York.

Chandrasekhar S. and von Neumann J. (1942), *Ap. J.* **95**, 489.

Chandrasekhar S. and von Neumann J. (1943), *Ap. J.* **97**, 1.

Cox D. R. and Miller H. D. (1965), *The Theory of Stochastic Processes*, Methuen, London.

Faulkner D. J. and Coleman C. S. (1984), *M.N.R.A.S.* **206**, 121.

Feller W. (1970), *An Introduction to Probability Theory and Its Applications*, Vol. 1, Wiley, New York.

Feller W. (1971), *An Introduction to Probability Theory and Its Applications*, Vol. 2, Wiley, New York.

Friedman A. (1975), *Stochastic Differential Equations and Applications*, Vol. 1, Academic Press, New York.

Friedman A. (1976), *Stochastic Differential Equations and Applications*, Vol. 2, Academic Press, New York.

Gardiner C. W. (1985), *Handbook of Stochastic Methods*, Springer-Verlag, Berlin.

Gilbert I. H. (1968), *Ap. J.* **152**, 1043.

Gilbert I. H. (1972), in *IAU Coll.* **10**, Reidel, Dordrecht, Lecar M. ed., p. 5 (Cambridge, UK, 1970).

Haken H. (1975), *Rev. Mod. Phys.* **47**, 67.

Hénon M. (1973), in *Dynamical Structure and Evolution of Stellar Systems*, Geneva Observatory, Sauverny, Martinet L. and Mayor M. eds., p. 183 (Saas-Fee, Switzerland, 1973).

Hernquist L. and Weinberg M. D. (1989), *M.N.R.A.S.* **238**, 407.

Hinton F. L. (1983), in *Basic Plasma Physics*, Vol. 1, North-Holland, Amsterdam, Galeev A. A. and Sudan R. N. eds., p. 147.

Hörmander L. (1976), *Linear Partial Differential Operators*, Springer-Verlag, Berlin.

Hörmander L. (1983), *The Analysis of Linear Partial Differential Operators*, Vols. 1-2, Springer-Verlag, Berlin.

Hörmander L. (1985), *The Analysis of Linear Partial Differential Operators*, Vols. 3-4, Springer-Verlag, Berlin.

Ichimaru S. (1980), *Basic Principles of Plasma Physics*, Benjamin-Cummings, London.

Inagaki S. and Lynden-Bell D. (1990), *M.N.R.A.S.* **244**, 254.

Kac M. and Logan J. (1987), in *Fluctuation Phenomena*, North-Holland, Amsterdam, Montroll E. W. and Lebowitz J. L. eds., p. 1.

Kandrup H. E. (1980), *Phys. Rep.* **63**, 1.

Kashlinsky A. (1986), *Ap. J.* **306**, 374.

Keenan D. W. (1979), *A. Ap.* **71**, 245.

Krall N. A. and Trivelpiece A. W. (1973), *Principles of Plasma Physics*, McGraw-Hill, New York.

Lee E. P. (1968), *Ap. J.* **151**, 687.

Li K.-H. (1986), *Phys. Rep.* **134**, 1.

Lynden-Bell D. (1962), *M.N.R.A.S.* **124**, 1.

Lynden-Bell D. (1973), in *Dynamical Structure and Evolution of Stellar Systems*, Geneva Observatory, Sauverny, Martinet L. and Mayor M. eds., p. 147 (Saas-Fee, Switzerland, 1973).

Manorama S. (1986), *Ph. D. Thesis*, SISSA, Trieste, Italy.

Montroll E. W. and West B. J. (1987), in *Fluctuation Phenomena*, North-Holland, Amsterdam, Montroll E. W. and Lebowitz J. L. eds., p. 61.

Mulder W. A. (1983), *A. Ap.* **117**, 9.

Narasimhan K. S. V. S. and Ballabh G. M. (1988), *Ap. Space Sci.* **149**, 115.

Nelson E. (1967), *Dynamical Theories of Brownian Motion*, Princeton University Press, Princeton.

Padmanabhan T. (1990), *Phys. Rep.* **188**, 285.

Palmer P. L. (1983), *M.N.R.A.S.* **204**, 507.

Palmer P. L. and Papaloizou J. (1982), *M.N.R.A.S.* **199**, 869.

Palmer P. L. and Papaloizou J. (1985), *M.N.R.A.S.* **215**, 691.

Petrovskaya I. V. (1986), *Sov. Astron. Lett.* **12**, 237.

Putterman S. J. and Roberts P. H. (1988), *Phys. Rep.* **168**, 209.

Rostoker N. and Rosenbluth M. N. (1960), *Phys. Fluids* **3**, 1.

Saslaw W. C. (1985), *Gravitational Physics of Stellar and Galactic Systems*, Cambridge University Press, Cambridge.

Schmidt G. (1979), *Physics of High Temperature Plasmas*, Academic Press, New York.

Sivukhin D. V. (1966), in *Reviews of Plasma Physics* 4, Consultants Bureau, New York, Leontovich M. A. ed., p. 93.

Spitzer L. Jr. (1987), *Dynamical Evolution of Globular Clusters*, Princeton University Press, Princeton.

Spohn H. (1980), *Rev. Mod. Phys.* **52**, 569.

Taylor M. E. (1981), *Pseudodifferential Operators*, Princeton University Press, Princeton.

Tremaine S. (1981), in *The Structure and Evolution of Normal Galaxies*, Cambridge University Press, Cambridge, Fall S. M. and Lynden-Bell D. eds., p. 67 (Cambridge, UK, 1980).

Tremaine S. and Weinberg M. D. (1984), *M.N.R.A.S.* **209**, 729.

Tscharnuter W. (1972), in *IAU Coll.* **10**, Reidel, Dordrecht, Lecar M. ed., p. 18 (Cambridge, UK, 1970).

Valtaoja L. (1990), *A. Ap.* **228**, 37.

van Kampen N. G. (1981), *Stochastic Processes in Physics and Chemistry*, North-Holland, Amsterdam.

Ventsel A. D. (1983), *Teoria dei Processi Stocastici*, Editori Riuniti, Roma.

Wax N. (1954), *Selected Papers on Noise and Stochastic Processes*, Dover, New York.

Weinberg M. D. (1986), *Ap. J.* **300**, 93.

Weinberg M. D. (1989), *M.N.R.A.S.* **239**, 549.

White S. D. M. (1976), *M.N.R.A.S.* **174**, 19.

Zaritsky D. and White S. D. M. (1988), *M.N.R.A.S.* **235**, 289.

Figure Captions

Figure 3.1. Oversimplified hierarchical scheme for the stochastic processes discussed in Section 3.2.

Table Captions

Table 3.1. Schematic comparison between the three basic approaches discussed in this chapter.

Figure 3.1

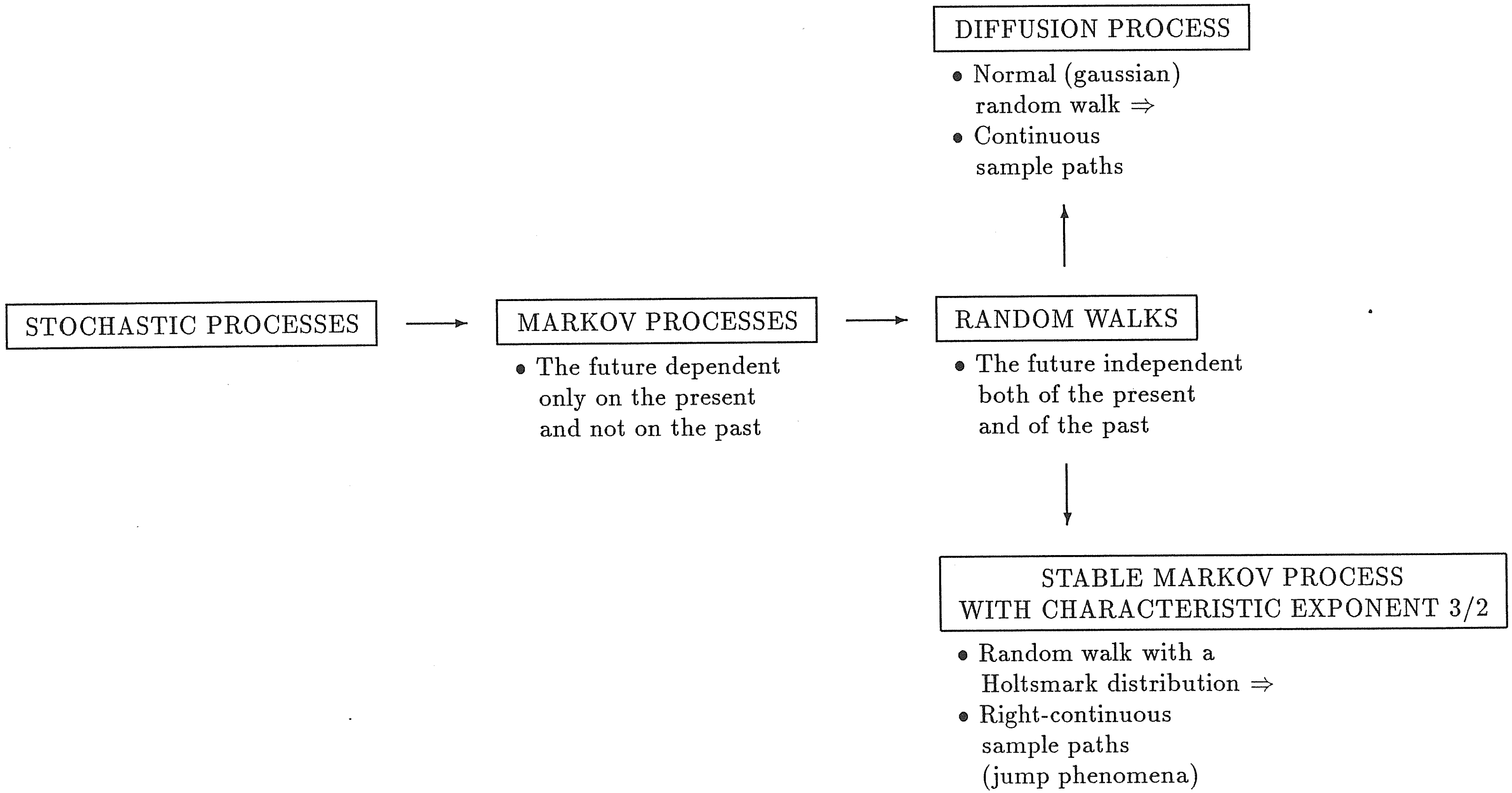

STOCHASTIC PROCESSES
MARKOV PROCESSES
• The future dependent only on the present and not on the past
RANDOM WALKS
• The future independent both of the present and of the past
DIFFUSION PROCESS
• Normal (gaussian) random walk ⇒
• Continuous sample paths
STABLE MARKOV PROCESS WITH CHARACTERISTIC EXPONENT 3/2
• Random walk with a Holtsmark distribution ⇒
• Right-continuous sample paths (jump phenomena)

Table 3.1

APPROACH	FOKKER-PLANCK	TSCHARNUTER	GILBERT
Type of evolution equation	Differential equation	Differential equation	Integro-differential equation
Basic assumptions	Ergodic assumption	Ergodic assumption	Initially uncorrelated states
			Equilibrium with respect to purely collective motions
Consequent assumptions	Binary encounters	Binary encounters	
Type of stochastic process	Diffusion process	Stable Markov process with characteristic exponent 3/2	Non-Markovian process of general type
Implicit assumptions		Spatial homogeneity	
Method	Chapman-Kolmogorov equation + Langevin equation	Semi-group theory + Langevin equation	Decoupling of the BBGKY hierarchy
Assumptions for calculating the coefficients explicitly	Instantaneous distant encounters in the impulsive approximation		
	Spatial homogeneity		
Comments	Starting differential equation $\rightarrow$ integro-differential equation $\rightarrow$ differential equation using $f \sim f_{\mathrm{M}}$	The relation between the two coefficients cannot directly be established; thus, no explicit calculation can be carried out	The resulting kinetic equation can only formally (i.e., not explicitly) be derived
Divergences inherent in the theory		Short-range divergence	Short-range divergence
Divergences related to the explicit calculation of the coefficients	Short-range divergence		
	Long-range divergence		

Chapter 4

Collisionless Relaxation Processes: The Role of Collective Effects in Electromagnetic and Gravitational Plasmas*

4.1 Introduction

In the previous chapters we have often stressed the fact that collective effects always play a crucial role in systems whose dynamics is governed by long-range interactions, such as plasmas and stellar systems. Large-scale organized motions and coherent processes such as self-sustained waves in isolated systems (e.g., differential rotation and spiral density waves in disc galaxies) are indeed expressions of the collective nature of such interactions. From a mathematical point of view, one of the main implications of this lies in the fact that a local analysis is often no more suitable for describing electromagnetic and gravitational plasmas, and a global analysis is thus required (i.e., boundary conditions are to be taken into account).

In the gravitational case a further mathematical complication arises from the fact that stellar systems are naturally inhomogeneous, because the gravitational force is always attractive—there is, in fact, only one gravitational charge—and screening effects are thus absent. The presence of these large-scale inhomogeneities generally requires the use of certain asymptotic per-

*The term "gravitational plasmas" is often used to indicate stellar systems in virtue of the fact that they are in several respects dynamically similar to ordinary (electromagnetic) plasmas (see, e.g., Bertin 1980; Lin and Bertin 1981).

turbation methods, whose validity depends upon the value of some local parameters characterizing the equilibrium state of the system.

In this chapter we shall restrict only to one particular role played by collective effects in electromagnetic and gravitational plasmas, namely the enhancement of relaxation processes by collective effects, a complete discussion of the general topic being extremely wide (see, e.g., Fridman and Polyachenko 1984) and not directly related to the argument of the thesis. To be more specific, we shall analyze collisionless relaxation processes alone (i.e., we shall adopt the Vlasov description), since the role of collective interactions on collisional relaxation processes has already been discussed in Section 3.3. In the framework of the Gilbert (1968) approach, this corresponds to drop the assumption that the system is in equilibrium with respect to purely collective motions and to neglect particle correlations at all. Two limits of of the weakly non-linear theory of plasma waves, which gives the correct framework for studying such collective relaxation processes, will finally be discussed.

General reference is made to Kulsrud (1972), Padmanabhan (1990), Sagdeev (1966), and to the books on plasma physics and stellar dynamics listed in Section 3.1. Other more specific references will be given later on.

4.2 Enhancement of Relaxation Processes by Collective Effects

It is a well-known fact that in a plasma the rate of relaxation towards the equilibrium state can be enhanced by collective effects. In virtue of the analogy between electromagnetic and gravitational plasmas, the question naturally arises whether a similar enhancement may be observed in stellar systems as well.

From an observational point of view, the existence of rapid relaxation mechanisms can be inferred from the fact that galaxies seem well relaxed, as they exhibit well-developed velocity distributions, and from the consideration that in such systems ordinary two-body collision processes operate on timescales largely exceeding even the Hubble time (Zwicky paradox).

An important step to explain this phenomenon was made by Lynden-Bell (1967), who formulated the theory of collisionless violent relaxation (see also Section 2.3). This theory, although it has a great heuristic advantage and has stimulated a lot of interest (e.g., Saslaw 1968, 1969, 1970; Cuperman, Goldstein and Lecar 1969; Goldstein, Cuperman and Lecar 1969; Shu 1969, 1978, 1987; Tremaine, Hénon and Lynden-Bell 1986, errata corrige 1987; Dejonghe 1987; Kandrup 1987; Madsen 1987; Tanekusa 1987; Aarseth, Lin and Papaloizou 1988; Mathur 1988; Wiechen, Ziegler and Schindler 1988; Aly 1989;

Ziegler and Wiechen 1989; Mineau, Feix and Rouet 1990; Stahl, Ziegler and Wiechen 1990), nevertheless cannot avoid certain difficulties. One of these lies in the fact that it describes essentially the non-equilibrium phase of evolution (collapse) of stellar systems and not their quasi-equilibrium phase. Among many further attempts to understand the relaxation of collisionless stellar systems it is worth mentioning those of Severne and Luwel (1980), Luciani and Pellat (1987a,b), Kandrup (1988), where the contribution of fluctuations of the mean self-consistent field to the relaxation process has been considered (see also Section 4.4).

Before considering the gravitational case in more detail, it is better to review such phenomena in plasmas, since they have long been studied in this context. Because of the reasons mentioned in the previous section, the extension of these results to stellar systems is not straightforward.

Examples of rapid relaxation in plasmas and the corresponding relaxation times are the following:

- The confinement of a plasma in a mirror machine: $\tau_{\rm coll} \sim (n\lambda_{\rm D}^3)^{-1}\tau_{\rm bin} \sim (10^{-6}\text{–}10^{-4})\tau_{\rm bin}$. A similar situation in which no exponential relaxation occurs is provided by the particles of the van Allen belt trapped (mirror effect) in the Earth dipole magnetic field.
- The two-stream instability: $\tau_{\rm coll} \sim (n_{\rm beam}\lambda_{\rm D}^3)^{-1}\tau_{\rm bin}$.

Note the different orders of magnitude involved in the two cases of binary and collective relaxation processes. The Debye shielding length $\lambda_{\rm D}$ is defined as

$$\lambda_{\rm D} \equiv \sqrt{\frac{KT}{4\pi n e^2}} = \frac{v_{\rm th}}{\omega_{\rm p}}, \tag{4.1}$$

where n is the number density of the relevant particles contributing to the relaxation (ions in the first example, electrons in the second example), $\omega_{\rm p}$ their plasma frequency and $v_{\rm th}$ their one-dimensional thermal velocity.

These examples are only two of the possible cases in which enhancement of relaxation can occur in a plasma, but in some sense they are typical and the following comments about them are relevant:

- These instabilities are in some sense weak. Strong instabilities tend to destroy the equilibrium. These weak instabilities work instead on a smaller scale, leading to a rapid relaxation towards a situation in which the equilibrium is no more unstable.
- These instabilities can occur in a time-dependent situation, in which case they usually lead to a relaxation at a rate comparable to the growth rate of the instability, or in a steady state, in which case they lead to

a marginally stable equilibrium with a relaxation rate in balance with whatever external forces tend to disturb it.

- The process of growth of the waves and relaxation towards the equilibrium are intimately related. They can generally be interpreted as a maser in which the unstable equilibrium corresponds to an over-population of the emitting states for the wave. The induced emission process is then the scattering process, as well as the process which makes the wave grow.

- The wave scattering is a collective process in which the particles collide with bunches of particles.

It turns out that these phenomena can no longer be described in the framework of a linear theory of plasma waves. There are two limits in which the weakly non-linear theory is tractable.

In the first case, where there are only a few waves of finite amplitudes, it is possible to treat each wave individually. This is called the theory of weak coherent waves.

The second case concerns the situation in which so many waves are present that a statistical approach can be employed to find those features of the time-evolution of the plasma state which do not depend on the details of the initial phase of the waves (random-phase approximation; see, e.g., Pines and Schrieffer 1962). Other approaches resembling the van der Pol method used in non-linear mechanics (see, e.g., Starzhinskii 1980; Sanders and Verhulst 1985) can also be employed. This is called the theory of weak turbulence, in which three basic interactions are taken into account:

- The wave-particle interaction (quasi-linear effect), studied in the framework of the quasi-linear theory, which is particularly strong near the resonance $\omega = \boldsymbol{k} \cdot \boldsymbol{v}$ when no external magnetic field is applied.[1]

- The non-linear wave-wave interaction (second-order effect), also known as the resonant-mode coupling, characterized by the resonance condition $\omega_1 + \omega_2 + \omega_3 = 0$, $\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3 = \boldsymbol{0}$.[2]

[1]The wave-particle resonance condition in the absence of an external magnetic field is expressed by the requirement that the velocity component of the particle along the direction of propagation of the wave should be equal to its phase velocity (see the formula given in the text). In the more general case in which an external magnetic field is also present two resonance conditions are possible: the Cherenkov resonance condition $\omega - l\omega_c = k_{\|}v_{\|}$, and the cyclotron resonance condition $\omega - l\omega_c = k_{\|}v_{\perp}$, ($l \in \mathbb{Z}$) where reference is made to the direction of the magnetic field and ω_c is the corresponding cyclotron frequency.

[2]The wave-wave resonance condition when more than three waves are involved in the scattering process (higher-order effects) is $\sum_{i=1}^{i^0} \omega_i = 0$, $\sum_{i=1}^{i^0} \boldsymbol{k}_i = \boldsymbol{0}$ ($i^0 \geq 3$).

- The wave-particle-wave interaction (third-order effect), also known as the non-linear wave-particle interaction, characterized by the resonance condition $\omega_1 \pm \omega_2 = (\boldsymbol{k}_1 \pm \boldsymbol{k}_2) \cdot \boldsymbol{v}$.

Both theories proceed essentially out of an iteration of the Vlasov equation, and fail when the amplitudes become so large that either the perturbation theory fails to converge, or the particle orbits become so distorted by the wave fields that the equilibrium distribution function can no longer be used to calculate accurately the linear wave properties of the plasma. The most usual example of such a distortion of orbits occurs when the particles become trapped in the troughs of plasma waves.

4.3 The Quasi-Linear Theory of Plasma Waves

The quantitative theory which describes rapid relaxation processes of the kind discussed in Section 4.2 is the quasi-linear theory, first proposed by Vedenov, Velikhov and Sagdeev (1961, 1962), Romanov and Filippov (1961), Drummond and Pines (1962). For general reference and different formulations see, in addition to the references cited in Section 4.1, also Akhiezer et al. (1975), Biskamp (1973), Bychenkov, Silin and Uryupin (1988), Coppi, Rosenbluth and Sudan (1969), Dewar (1970), Drummond (1965), Drummond and Pines (1964), Drummond and Ross (1973), Frieman and Rutherford (1964), Frieman, Bodner and Rutherford (1963), Galeev and Sagdeev (1979, 1983), Goldman (1984), Kadomtsev (1965), Krall and Trivelpiece (1973), Lifshitz and Pitaevskii (1981), Pines and Schrieffer (1962), Rosenbluth, Coppi and Sudan (1969), Sagdeev and Galeev (1969), Schmidt (1979), Sitenko (1982), Tsytovich (1970, 1972, 1989), Vedenov (1967), Vedenov and Ryutov (1975), Whitham (1965), Yasseen and Vaclavik (1983).

We shall now sketch out the basic ideas underlying this theory. When studying small (linear) oscillations in a plasma the distribution function is taken to be split into two terms: a non-oscillating part (the initial distribution function) and a small correction to it which oscillates with the frequency of the plasma waves. The non-oscillating part is then assumed not to be connected at all with the oscillations. Actually, however, either the damping or the growth of plasma waves affects the unperturbed distribution function, and this in turn generally alters the stability properties of the plasma. This effect increases with increasing amplitude of the oscillations. When the amplitude of the oscillations increases, the basic property of linear oscillations, i.e. the independence of the propagation of oscillations with different wavevector and frequency (superposition principle), tends also to be violated since processes involving the interactions between different waves begin to play an ever more important role.

The simplest among the non-linear processes which cannot be treated without taking into account the effect of plasma oscillations on the non-oscillating part of the distribution function, while the violation of the superposition principle is still neglected, is indeed the quasi-linear relaxation. In this process only the distribution of the resonant particles, whose number is assumed to be much smaller than the total number of particles (i.e., only sharp wave packets in $\boldsymbol{k}$-space are considered), is affected in a non-negligible manner by these weakly non-linear waves (quasi-linear diffusion). Such particles, in fact, are involved in strong interactions with the plasma oscillations, which lead to damping (Landau damping) or amplification (inverse Landau damping) phenomena depending on the monotonicity properties of their velocity distribution. The non-resonant particles do not exchange energy with the waves on the average, so that their distribution is almost insensitive to the effect of the oscillations (adiabatic quasi-linear diffusion).

Having stressed the main ideas which are at the basis of the quasi-linear theory, we now turn to discuss their implications in some more detail also from a quantitative point of view. As mentioned at the beginning of this section, different (almost equivalent) formulations can be given. In what follows we shall try to extract the common essential features of these approaches, avoiding any particular reference to specific physical situations.

Taking into account the fact that two considerably different timescales are involved, one governing the relaxation towards the equilibrium state and the other associated with the plasma oscillations, we separate the distribution function into a slowly varying part f_0 and a rapidly varying part f_1:

$$f(\boldsymbol{x},\boldsymbol{v};t) = f_0(\boldsymbol{x},\boldsymbol{v};t) + f_1(\boldsymbol{x},\boldsymbol{v};t)\,, \qquad \left|\frac{\partial \ln f_0}{\partial t}\right| \ll \left|\frac{\partial \ln f_1}{\partial t}\right|\,. \tag{4.2}$$

The distribution function f is then taken to satisfy the system of the (coupled) Vlasov and Maxwell equations (self-consistent description). By singling out the two contributions and performing a Fourier expansion of the perturbations, it can be shown that under the basic assumptions discussed previously (quasi-linear approximation) the time-evolution of f_0 is described by the quasi-linear diffusion equation

$$\frac{\partial f_0}{\partial t} = \frac{1}{2}\frac{\partial}{\partial v_i}\left(D_{ij}\frac{\partial f_0}{\partial v_j}\right)\,, \tag{4.3}$$

where the diffusion tensor $D_{ij} = D_{ij}(\boldsymbol{v},t)$ is of second order in the perturbations, being a linear functional of the energy density of the waves in the turbulent plasma.[3] It is the main task of the quasi-linear theory to express

[3] Bear in mind that to zeroth order of expansion f_0 satisfies the stationary Vlasov equation $[f_0, H_0] = 0$.

it in terms of the wave spectrum.

This evolution equation has exactly the same form as the Fokker-Planck equation with dynamical-friction vector $A_i = \frac{1}{2}(\partial D_{ij}/\partial v_j)$. It should be noted, however, that the quadratic form $D_{ij}v_iv_j$ is not necessarily positive definite (for non-resonant particles) as in the case of collisional relaxation processes, and this stresses the fact that in collisionless relaxation processes stochastic deceleration mechanisms can take place, leading to a "reversed diffusion" in velocity space. Processes in which particles are scattered by plasma waves (wave-particle interaction[4]) are thus formally similar, but not identical, to ordinary particle-particle scattering processes. The derivation of an analogous "diffusion" equation in a more general context starting from the system of the Vlasov and Poisson equations can be found in Wollman (1985), together with an extremely interesting discussion of some conceptual difficulties inherent in such a description (see Antonuccio-Delogu 1990 for an interesting application of this formalism to the angular-momentum transfer through non-axisymmetric gravitational instabilities between the halo and disc components of spiral galaxies).

From a quantum-mechanical point of view, the resonance condition for this interaction expresses the conservation of energy and momentum in the elementary process involving the emission or the absorption of a plasmon with energy $\hbar\omega$ and momentum $\hbar\boldsymbol{k}$ by a particle moving with velocity $\boldsymbol{v}$. Thus it is not surprising that the wave-particle interaction conserves the total energy and momentum of the waves and particles, rather than the energy and momentum of the waves alone.

The number of plasmons tends to be conserved and satisfies a continuity equation with a source term in the $(\boldsymbol{x}, \boldsymbol{k})$ phase space:

$$\frac{\partial N_k}{\partial t} + [N_k, \omega_k] = 2\gamma_k N_k \,, \quad \text{where} \quad N_k \equiv \frac{\mathcal{E}_k}{\omega_k} \,, \tag{4.4}$$

valid for inhomogeneous equilibrium states, provided the wavelengths are sufficiently short and the frequencies sufficiently large. N_k is the plasmon density (wave-action density) in phase space, $\mathcal{E}_k$ the energy density of the waves in phase space, ω_k and γ_k their frequency and growth (or damping) rate, respectively. These two quantities are related to the wave-vector $\boldsymbol{k}$ and to the quasi-equilibrium distribution function f_0 by the same dispersion relation holding in the linear regime:

$$\mathcal{D}(\omega, \gamma; \boldsymbol{k}; f_0) = 0 \,. \tag{4.5}$$

The quasi-linear diffusion equation for f_0, the continuity equation for the number of plasmons, and the dispersion relation for ω and γ represent the complete set of equations of the quasi-linear theory.

[4] Since this interaction involves resonant particles, it cannot be considered within the framework of an equivalent fluid theory.

In the forthcoming discussion we shall assume that no external magnetic field is applied. In the more general case in which an external magnetic field is present the results presented below are still roughly valid, even though a more detailed description is required.

Particle diffusion resulting from the scattering by plasma waves leads to the establishment of an asymptotic stationary state, which is characterized by a fixed distribution of resonant particles and some definite spectral level. More precisely, either the oscillations are damped or a plateau is formed on the distribution function (i.e., f_0 is constant in the resonance interval $\Delta(\omega_k/k)$ along the direction of wave propagation). It may be worth noting that if the phase-space volume occupied by resonant particles is rather large, the formation of a plateau in that volume becomes impossible as it would require too much energy. In that case either the oscillations are damped, or the spectrum becomes one-dimensional while along the direction of wave propagation a plateau is formed on the distribution function.

As a consequence of the quasi-linear relaxation process in which an initial overstability[5] is finally saturated (i.e., $\gamma(t) \to 0$ as $t \to +\infty$) and a plateau is formed on the distribution function the non-resonant particles of the system undergo an effective collisionless stochastic heating, which is called the turbulent heating, whereas the resonant particles are cooled down. When an initial overstability is instead finally damped (i.e., $\gamma(t) < 0$ as $t \to +\infty$) or in the case of transient waves (i.e., $\gamma(t) < 0$) the situation turns out to be reversed. Part of the ordered motion associated with the waves is thus converted into random motion of the particles. This is just an example of anomalous transport phenomena occurring in a plasma which, as a consequence of an overstability, passes from a laminar to a turbulent state.

Other weaker heating processes due to wave-particle interactions can be described in the framework of the more general weakly non-linear theory. Such non-linear wave-particle interactions are responsible for the damping and amplification phenomena occurring in the waves which lead to these heating processes, and are thus referred to as the non-linear Landau damping and the non-linear inverse Landau damping, respectively.

We shall now inquire into the validity of the quasi-linear theory, by explaining in what physical situations it becomes inapplicable. To the order in the amplitude of the waves to which the quasi-linear theory is valid no interaction between the waves themselves occurs, but the second-order effect of the interaction between waves and particles is included. As one considers

---

[5]Given a perturbation whose time-dependence is of the form $f_1 \sim e^{-i\omega_{\mathbb{C}} t}$ with $\omega_{\mathbb{C}} = \omega + i\gamma$, the following terminology is used:

- Instability: $\gamma > 0$, $\omega = 0$.
- Overstability: $\gamma > 0$, $\omega \neq 0$.

the situation in which the amplitude is larger, it is expected that even this interaction is not well represented by the theory. Since the theory basically assumes the orbits of the particles to be modified by small amounts from their unperturbed orbits, it can be guessed that a limit on the theory will occur when the amplitude of the waves is large enough to trap the particles. Therefore, if the bounce period of a particle trapped in a wave is shorter than the time this particle spends in the wave packet, it could reasonably be expected the theory to be inaccurate.

So far we have considered the situation in which so many waves of finite amplitudes are present that a statistical approach can be employed to formulate a quasi-linear theory of plasma waves. However, in some important physical situations only a few waves are involved, so that a different treatment is required (theory of weak coherent waves). Using the Drummond and Pines (1962) approach, which is not based on statistical assumptions, it can be shown that the time-evolution of the slowly varying distribution function f_0 is indeed described by a quasi-linear diffusion equation exactly of the same form as that derived in the case of many waves (to which this approach refers). The same result can be obtained in the framework of other quasi-linear approaches which are not based on the random-phase approximation. In this context (wave-particle interaction) only the calculation of the diffusion tensor cannot be carried out along the same line (an integration over the wave-vector $\boldsymbol{k}$ space is replaced by a finite sum over $\boldsymbol{k}_i$).

Because of this fact, bearing in mind that we are interested in the wave-particle interaction alone, we shall extend the meaning of the term "quasi-linear theory" to include also the case in which a few waves are considered.

4.4 Attempts to Achieve a Satisfactory Formulation of a Quasi-Linear Theory in the Gravitational Case

This section is mostly devoted to explain the main difficulties which are to be tackled, from a general point of view, for extending the quasi-linear theory to stellar systems as well, and to mention the suggestions of various authors to achieve a satisfactory formulation of such a theory, which is not available yet. My own contribution and proposals in the particular framework of the spiral structure theory will be stressed in the second part of this thesis, as they are intended to be a contribution to the understanding of the secular heating of galactic discs.

In plasma physics one usually deals with systems many orders of magnitude larger than the scale of the waves contributing to collective processes, namely the Debye shielding length. This makes the analysis comparatively simple, since the standard Fourier expansion technique can be used. Stated in another way, in plasmas the requirement of performing a global mode analysis can often be bypassed because the assumption of spatial homogeneity is reasonably satisfied in many cases of physical interest, so that a local analysis can suitably be used.

Unfortunately this is not the case for self-gravitating systems, whose size is not so drastically different from the wavelength scale. This makes it necessary to treat the waves as eigenmodes of the system. There is, however, even in the case of stellar systems a cunning trick for eliminating, from a formal point of view, the "unpleasant" effect of large-scale inhomogeneities. It simply consists in using the action-angle variables $\{(J_i, w_j);\ i,j = 1,2,3\}$ as the proper canonical coordinates. In this representation, in fact, the equilibrium quantities of integrable systems turn out to depend only on J_i, which in this context (and also in a more general context; cf. the Hamilton equations) play thus the role of the velocity components v_i (see, e.g., Kalnajs 1971; Galgani 1985). Taking into account the physical meaning of these variables (the actions J_i correspond to adiabatic invariants) and the "strengthened" Jeans theorem, it can be shown that even in quasi-equilibrium situations a similar result holds apart from the fact that in this case an explicit dependence on time is allowed (see, e.g., Binney and Lacey 1988). In particular, this is true for the slowly varying part of the distribution function $f_0 = f_0(J_i, t)$ and for the self-consistent Hamiltonian $H_0 = H_0(J_i, t)$, so that $[f_0, H_0] = 0$.

The most direct physical implication of the dependence of f_0 and H_0 on the actions alone consists in the possibility of adopting the standard Fourier representation[6] in the angle space for the perturbations. While on the one hand there is the advantage of using this local treatment avoiding the difficulties connected with the solution of a global-mode equation, on the other hand this canonical representation has the drawback of not being simply related to directly observable quantities as indeed $\boldsymbol{x}$ and $\boldsymbol{v}$. It is just for this reason that basically important theories of stellar systems, such as for instance the spiral structure theory, are expressed in terms of the usual canonical coordinates $(\boldsymbol{x}, \boldsymbol{v})$: the formal elegance and the compactness deriving from the use of the action-angle variables are indeed sacrificed in favour of a higher level of predictability and a simpler interpretation in terms of observable phenomena.

Because of the above-mentioned difficulties, only local formulations of

[6]By "standard Fourier expansion" we mean that the perturbation can be expressed as a series of wave components of the form $f_k = \hat{f}_k e^{ikx}$. If the system is inhomogeneous along the x-direction, the previous expression is to be replaced by the more general dependence $f_k = \hat{f}_k(x) \exp\left[i \int^x k(x')\, dx'\right]$.

the quasi-linear theory for gravitational plasmas have been given so far, and mostly in the framework of the spiral structure theory (e.g., Marochnik and Suchkov 1969; see also Marochnik 1970; Dekker 1976, who extended the work of Lynden-Bell and Kalnajs 1972; see also Contopoulos 1974). These works, however, neither have shed new light on the problem of spiral structure —apart from some local results which can easily be extrapolated from the plasma analogue of spiral waves (i.e., the Bernstein waves)—nor have tried to incorporate the fundamental role played by excitation mechanisms at the corotation resonance shown in global linear treatments (for a more exhaustive discussion see Section 6.3 of Part II). In the framework of the theory of weak turbulence the non-linear effects of resonant spiral waves have also been investigated (e.g., Contopoulos 1972; Churilov and Shukhman 1981, 1982; Tagger et al. 1987; Sygnet et al. 1988).

A local formulation of considerably different type related to the theory of violent relaxation (Lynden-Bell 1967) has been proposed by Severne and Luwel (1980) along a line similar to that followed by Kadomtsev and Pogutse (1970) in the context of weak homogeneous plasma turbulence. Other interesting treatments linking the collisionless and the collisional viewpoints into a unified description have been developed by Luciani and Pellat (1987a,b) and Kandrup (1988) in analogy with the approaches adopted by Klimontovich (1967) and Thompson (1964), respectively, in the context of plasma fluctuations.

Extremely interesting discussions of related subjects can be found in Saslaw (1985).

References

Aarseth S. J., Lin D. N. C. and Papaloizou J. (1988), *Ap. J.* **324**, 288.

Akhiezer A. I., Akhiezer I. A., Polovin R. V., Sitenko A. G. and Stepanov K. N. (1975), *Plasma Electrodynamics*, Vol. 2, Pergamon Press, Oxford.

Aly J. J. (1989), *M.N.R.A.S.* **241**, 15.

Antonuccio-Delogu V. (1990), "Angular Momentum Transfer in Young Spiral Disks", submitted to *M.N.R.A.S.*.

Bertin G. (1980), *Phys. Rep.* **61**, 1.

Binney J. and Lacey C. G. (1988), *M.N.R.A.S.* **230**, 597.

Biskamp D. (1973), in *Turbulence and Non-Linear Effects in Plasmas*, UKAEA-SRC, Abingdon, Keen B. E. and Laing E. W. eds., p. 93 (Abingdon, UK, 1973).

Bychenkov V. Yu., Silin V. P. and Uryupin S. A. (1988), *Phys. Rep.* **164**, 119.

Churilov S. M. and Shukhman I. G. (1981), *Sov. Astron.* 25, 147.

Churilov S. M. and Shukhman I. G. (1982), *Sov. Astron.* 26, 660.

Contopoulos G. (1972), in *IAU Coll.* **10**, Reidel, Dordrecht, Lecar M. ed., p. 169 (Cambridge, UK, 1970).

Contopoulos G. (1974), in *IAU Symp.* **58**, Reidel, Dordrecht, Shakeshaft J. R. ed., p. 413 (Canberra, Australia, 1973).

Coppi B., Rosenbluth M. N. and Sudan R. N. (1969), *Ann. Phys.* **55**, 207.

Cuperman S., Goldstein S. and Lecar M. (1969), *M.N.R.A.S.* **146**, 161.

Dejonghe H. (1987), *Ap. J.* **320**, 477.

Dekker E. (1976), *Phys. Rep.* **24**, 315.

Dewar R. L. (1970), *Phys. Fluids* **13**, 2710.

Drummond W. E. (1965), in *Plasma Physics*, IAEA, Vienna, p. 527 (Trieste, Italy, 1964).

Drummond W. E. and Pines D. (1962), *Nucl. Fusion Suppl.*, Part 3, 1049.

Drummond W. E. and Pines D. (1964), *Ann. Phys.* **28**, 478.

Drummond W. E. and Ross D. W. (1973), in *Turbulence and Non-Linear Effects in Plasmas*, UKAEA-SRC, Abingdon, Keen B. E. and Laing E. W. eds., p. 1 (Abingdon, UK, 1973).

Fridman A. M. and Polyachenko V. L. (1984), *Physics of Gravitating Systems*, 2 Vols., Springer-Verlag, New York.

Frieman E. and Rutherford P. (1964), *Ann. Phys.* **28**, 134.

Frieman E., Bodner S. and Rutherford P. (1963), *Phys. Fluids* **6**, 1298.

Galeev A. A. and Sagdeev R. Z. (1979), in *Reviews of Plasma Physics* **7**, Consultants Bureau, New York, Leontovich M. A. ed., p. 1.

Galeev A. A. and Sagdeev R. Z. (1983), in *Basic Plasma Physics*, Vol. 1, North-Holland, Amsterdam, Galeev A. A. and Sudan R. N. eds., p. 677.

Galgani L. (1985), in *Chaos in Astrophysics*, Reidel, Dordrecht, Buchler J. R. et al. eds., p. 245 (Palm Coast, Florida USA, 1984).

Gilbert I. H. (1968), *Ap. J.* **152**, 1043.

Goldman M. V. (1984), *Rev. Mod. Phys.* **56**, 709.

Goldstein S., Cuperman S. and Lecar M. (1969), *M.N.R.A.S.* **143**, 209.

Kadomtsev B. B. (1965), *Plasma Turbulence*, Academic Press, London.

Kadomtsev B. B. and Pogutse O. P. (1970), *Phys. Rev. Lett.* **25**, 1155.

Kalnajs A. J. (1971), *Ap. J.* **166**, 275.

Kandrup H. E. (1987), *M.N.R.A.S.* **225**, 995.

Kandrup H. E. (1988), *M.N.R.A.S.* **235**, 1157.

Klimontovich Yu. L. (1967), *The Statistical Theory of Non-Equilibrium Processes in a Plasma*, Pergamon Press, Oxford.

Krall N. A. and Trivelpiece A. W. (1973), *Principles of Plasma Physics*, McGraw-Hill, New York.

Kulsrud R. M. (1972), in *IAU Coll.* **10**, Reidel, Dordrecht, Lecar M. ed., p. 337 (Cambridge, UK, 1970).

Lifshitz E. M. and Pitaevskii L. P. (1981), *Physical Kinetics*, Pergamon Press, Oxford.

Lin C. C. and Bertin G. (1981), in *Plasma Astrophysics*, *ESA SP*-**161**, Paris, Guyenne T. D. and Levy G. eds., p. 191 (Varenna, Italy, 1981).

Luciani J. F. and Pellat R. (1987a), *J. Physique* **48**, 591.

Luciani J. F. and Pellat R. (1987b), *Ap. J.* **317**, 241.

Lynden-Bell D. (1967), *M.N.R.A.S.* **136**, 101.

Lynden-Bell D. and Kalnajs A. J. (1972), *M.N.R.A.S.* **157**, 1.

Madsen J. (1987), *Ap. J.* **316**, 497.

Marochnik L. S. (1970), in *IAU Symp.* **38**, Reidel, Dordrecht, Becker W. and Contopoulos G. eds., p. 317 (Basel, Switzerland, 1969).

Marochnik L. S. and Suchkov A. A. (1969), *Ap. Space Sci.* **4**, 317.

Mathur S. D. (1988), *M.N.R.A.S.* **231**, 367.

Mineau P., Feix M. R. and Rouet J. L. (1990), *A. Ap.* **228**, 344.

Padmanabhan T. (1990), *Phys. Rep.* **188**, 285.

Pines D. and Schrieffer J. R. (1962), *Phys. Rev* **125**, 804.

Romanov Yu. A. and Filippov G. F. (1961), *Sov. Phys. JETP* **13**, 87.

Rosenbluth M. N., Coppi B. and Sudan R. N. (1969), *Ann. Phys.* **55**, 248.

Sagdeev R. Z. (1966), in *Reviews of Plasma Physics* **4**, Consultants Bureau, New York, Leontovich M. A. ed., p. 23.

Sagdeev R. Z. and Galeev A.A. (1969), *Nonlinear Plasma Theory*, Benjamin, New York, O'Neil T. M. and Book D. L. eds..

Sanders J. A. and Verhulst F. (1985), *Averaging Methods in Nonlinear Dynamical Systems*, Springer-Verlag, New York.

Saslaw W. C. (1968), *M.N.R.A.S.* **141**, 1.

Saslaw W. C. (1969), *M.N.R.A.S.* **143**, 437.

Saslaw W. C. (1970), *M.N.R.A.S.* **150**, 299.

Saslaw W. C. (1985), *Gravitational Physics of Stellar and Galactic Systems*, Cambridge University Press, Cambridge.

Schmidt G. (1979), *Physics of High Temperature Plasmas*, Academic Press, New York.

Severne G. and Luwel M. (1980), *Ap. Space Sci.* **72**, 293.

Shu F. H. (1969), *Ap. J.* **158**, 505.

Shu F. H. (1978), *Ap. J.* **225**, 83.

Shu F. H. (1987), *Ap. J.* **316**, 502.

Sitenko A. G. (1982), *Fluctuations and Non-Linear Wave Interactions in Plasmas*, Pergamon Press, Oxford.

Stahl B., Ziegler H. J. and Wiechen H. (1990), *M.N.R.A.S.* **244**, 260.

Starzhinskii V. M. (1980), *Applied Methods in the Theory of Nonlinear Oscillations*, Mir, Moscow.

Sygnet J. F., Tagger M., Athanassoula E. and Pellat R. (1988), *M.N.R.A.S.* **232**, 733.

Tagger M., Sygnet J. F., Athanassoula E. and Pellat R. (1987), *Ap. J. Lett.* 318, L43.

Tanekusa J. (1987), *Publ. Astron. Soc. Japan* **39**, 425.

Thompson W. B. (1964), *An Introduction to Plasma Physics*, Pergamon Press, Oxford.

Tremaine S., Hénon M. and Lynden-Bell D. (1986), *M.N.R.A.S.* **219**, 285; errata corrige (1987), **227**, 543.

Tsytovich V. N. (1970), *Nonlinear Effects in Plasma*, Plenum Press, New York.

Tsytovich V. N. (1972), *An Introduction to the Theory of Plasma Turbulence*, Pergamon Press, Oxford.

Tsytovich V. N. (1989), *Phys. Rep.* **178**, 261.

Vedenov A. A. (1967), in *Reviews of Plasma Physics* **3**, Consultants Bureau, New York, Leontovich M. A. ed., p. 229.

Vedenov A. A. and Ryutov D. D. (1975), in *Reviews of Plasma Physics* **6**, Consultants Bureau, New York, Leontovich M. A. ed., p. 1.

Vedenov A. A., Velikhov E. P. and Sagdeev R. Z. (1961), *Nucl. Fusion* **1**, 82.

Vedenov A. A., Velikhov E. P. and Sagdeev R. Z. (1962), *Nucl. Fusion Suppl.*, Part 2, 465.

Whitham G. B. (1965), *J. Fluid Mech.* **22**, 273.

Wiechen H, Ziegler H. J. and Schindler K. (1988), *M.N.R.A.S.* **232**, 623.

Wollman S. (1985), *Transport Theory and Stat. Phys.* **14**, 567.

Yasseen F. and Vaclavik J. (1983), *Phys. Fluids* **26**, 468.

Ziegler H. J. and Wiechen H. (1989), *M.N.R.A.S.* **238**, 1261.

Chapter 5

Heating Mechanisms in Galactic Discs

5.1 Introduction

In Section 1.2 we have seen that the components of the velocity dispersion of disc stars in a solar neighbourhood show a tendency to increase with increasing spectral type, and this has been interpreted in terms of a corresponding increase of the components of the stellar velocity dispersion with age. Because of the observational difficulties connected with the determination of the stellar velocity dispersion, we do not know yet for certain whether such a systematic behaviour is restricted only to a small solar neighbourhood or instead is a general feature of galactic discs.[1] We shall see, however, that from a theoretical point of view there is no problem to account for the same phenomenon on larger scales or even in other similar stellar systems, provided the sun is not thought of as belonging to a privileged region of the Galaxy.

Two classes of explanations have been invoked. The most diffused and currently accepted of them is based on the existence of relaxation mechanisms leading to a secular heating of galactic discs. The other regards the observed velocity dispersions as native properties.

In this chapter we shall review the relaxation mechanisms which are thought to contribute more effectively to the increase of the components of the stellar velocity dispersion with age. Other less important relaxation mechanisms will be mentioned for the sake of completeness in the forthcoming discussion. We shall also describe in less detail the basic ideas underlying

[1]Recall that the spheroidal component has a different cosmological origin than the disc component, so that the same age-velocity dispersion relation is not expected to hold.

the second class of explanations.

The velocity dispersion of disc stars can be affected by the following mechanisms[2] (see Fuchs and Wielen 1987; Wielen and Fuchs 1985; see also: Binney and Tremaine 1987; Mihalas and Binney 1981):

- The stochastic heating (random increase of the stellar velocity dispersion) caused by local irregularities in the galactic gravitational field due to the existence of massive perturbers, as giant molecular clouds (GMCs) and hypothetical massive halo objects, or by large-scale phenomena as transient spiral waves[3]. Ordinary binary encounters between stars, in fact, are known to be completely inefficient (see Section 2.2).
- The deflections (random changes in the direction of the stellar velocity) caused by the same phenomena responsible for the stochastic heating. Their overall importance lies in the fact that they can transfer energy (and energy changes) between the motions of a star perpendicular and parallel to the galactic plane, so that deflections may be of primary importance to determine the axial ratios of the velocity ellipsoid even though the heating effect of the same irregularities is nearly negligible. This fact occurs when the velocity dispersion of the massive perturbers is much smaller than that of the test stars. In this case, in fact, the relaxation time for deflections $\tau_{\mathrm{bin}_\mathrm{D}}$ turns out to be much smaller than the relaxation time for equipartition of energy $\tau_{\mathrm{bin}_\mathrm{E}}$ (see Chandrasekhar 1960; see also Woltjer 1967).
- The adiabatic heating or cooling (adiabatic changes in the stellar velocity dispersion) produced by slow changes in the regular gravitational field of galactic discs. Its effect is probably stronger perpendicularly to the galactic plane, because the disc is nearly self-gravitating in this direction. Two typical examples of adiabatic changes in galactic discs are the adiabatic cooling due to stochastic heating and the adiabatic heating due to infall of gas from haloes.

While the adiabatic heating and cooling discussed above primarily affect the perpendicular motion of stars, deflections can partially transfer this energy change to their parallel motions.

Only the first two relaxation mechanisms will be considered in this chapter, the third one acting on much longer timescales.

[2]Recall that the restriction to nearly integrable situations is assumed. For strong departures from the integrability condition other relaxation mechanisms become operative due to the effects of Lyapunov instability (see Section 2.3).

[3]Transient spiral waves are not self-sustained. In the second part of this thesis we shall propose the interaction of disc stars with self-sustained spiral waves as the dominant heating mechanism in galactic discs.

5.2 The Role of Massive Perturbers

5.2.1 Giant Molecular Clouds and Complexes

The importance that massive perturbers might have in the dynamical evolution of disc stars has first been stressed by Spitzer and Schwarzschild (1951, 1953), who hypothesized the existence of low-velocity dispersion ($c_g \sim 10$ km s^{-1}) massive ($M_g \sim 10^5$–10^6 $M_\odot$) gas clouds and complexes to account for the observed velocity dispersion of stars of different spectral class.

In the first paper they performed a numerical integration of the Fokker-Planck equation, taking the low gas velocity dispersion into account but disregarding the effects of galactic rotation and the vertical motion of stars. They noted that the so-obtained age-dependence of the stellar velocity dispersion, $c(\hat{t}) \approx c_g(1+\hat{t})^{1/5}$ with time expressed in dimensionless units, could more simply be derived assuming that the stellar distribution function remains Maxwellian at all times (a Maxwellian distribution function was taken as the initial condition).

Taking this fact into account, in the second paper they performed analytical calculations of the kind used for studying the secular effects of binary encounters (see Chandrasekhar 1960; see also Hénon 1973), considering the epicyclic motion of stars in the galactic plane but still disregarding their vertical motion and also the low turbulent velocities of the interstellar clouds. They obtained a different power-law for the evolution of the stellar velocity dispersion: $c(\hat{t}) \approx c_0(1+\hat{t})^{1/3}$, where the same notations as before have been used.

The existence of giant molecular clouds and complexes has definitely been shown in the 1970s. Since then a lot of observational and theoretical works have been devoted to investigate their physical and kinematical properties, as well as to study their role in the context of galactic disc stability and evolution (for a more detailed discussion see Section 7.1 of Part II and references cited therein). It is quite surprising that the mass of GMCs has been estimated to be of the same order as that theoretically predicted by Spitzer and Schwarzschild (1951, 1953), even though their turbulent velocities are thought to be smaller ($\sim$ 4–8 km s^{-1}).

As regards the problem of the stochastic heating of galactic discs, in which we are mostly interested, the contribution of Lacey (1984a) (see also Lacey 1984b, 1985) deserves particular attention, and thus will be discussed in detail. It can be considered a generalization of the Spitzer and Schwarzschild (1953) work, since it relies on the same assumptions and employs similar methods but takes the vertical motion of stars into account. This extension is expected to be important both because of the intrinsic interest in predicting the evolution of the vertical velocity dispersion of disc stars, and because their vertical epicyclic oscillations can take stars out of the layer of perturbing GMCs so as to reduce the scattering rate.

The assumptions made by Lacey (1984a) are those commonly used to treat star-GMC encounters in such a degree of approximation as to make analytical calculations not particularly stiff, but at the same time to give a correct physical description retaining the essential features of the relaxation mechanism. They are the following:

1. The orbits of disc stars in the background galactic potential (assumed to be axi-symmetric and plane-symmetric) are described by the first-order epicyclic theory.

2. GMCs are long-lived, much more massive than disc stars and move in circular orbits.

3. GMCs are randomly distributed and act independently. Thus the possibility of the GMC distribution being organized on a large scale (i.e., into spiral arms) is neglected.

4. For a typical star-GMC encounter the effective interaction time is short compared to the epicyclic period, and the velocity difference between the Local Standard of Rest (LSR) at the star and at the GMC positions is negligible with respect to the peculiar velocity of the star.

5. The change of the stellar peculiar velocity is dominated by the effect of many distant weak encounters.

The first, the second and the fifth assumption are reasonably satisfied, whereas the third and the fourth assumption may be criticized, the former being the most drastic one. These working assumptions, however, allow us to make use of standard methods for deriving the diffusion and the dynamical-friction coefficients in binary-encounter processes (see Chandrasekhar 1960; see also Hénon 1973).

The evolution of the stellar velocity dispersion can be divided into two phases:

- A transient relaxation in which the shape of the velocity ellipsoid relaxes to a final steady state with $c_r : c_\theta : c_z = 1 : (\kappa/2\Omega) : (c_z/c_r)_s$, $\Omega = \Omega(r)$ and $\kappa = \kappa(r)$ being the angular velocity and the epicyclic frequency, respectively.[4] The existence of this phase depends on the non-vanishing of the dynamical-friction coefficient.

[4]Note that collisionless phase-mixing mechanisms (Lynden-Bell 1962; see also Freeman 1975) act on comparable timescales (see, e.g., Berry 1973; Byl and Ovenden 1973), and are thus competitive with respect to GMC-induced collisional phase-mixing mechanisms.

- A steady heating (absent for a solid-body rotation curve) in which the velocity dispersion increases steadily on a longer timescale, while its components maintain constant ratios depending only on the local value of $\kappa/2\Omega$: $c(\hat{t}) \approx c_0(1+\hat{t})^{1/4}$. It is worth noting that the Spitzer and Schwarzschild (1953) $\frac{1}{3}$-power-law is recovered in the unphysical limit in which the scale-height of disc stars is much smaller than the scale-height of GMCs.

The apparent discrepancy between these theoretical predictions and observational results (up to the year of publication of this paper) seemed to rule GMCs out of the role of most promising heating mechanism in galactic discs. It should be borne in mind, however, as stressed in Section 1.2 and less explicitly pointed out by Lacey (1984a), that observations do not put yet any stringent constraint on the age-dependence of the components of the stellar velocity dispersion and on the shape of the velocity ellipsoid because strong selection and contamination effects, inherent in the choice of otherwise claimed to be reliable samples, tend to bias such samples in a not simply estimable manner. A proof of this lies in the fact that even some recent observational surveys are mutually inconsistent.

The contribution of GMCs to the stochastic heating of galactic discs has been investigated also by several other authors both analytically (e.g., Fujimoto 1980: $\frac{1}{2}$; Kamahori and Fujimoto 1986a: $\frac{1}{3}$ for vanishing dynamical friction, otherwise saturation; Semenzato 1987: $\frac{1}{3}$; Binney and Lacey 1988: $\frac{1}{4}$, and $\frac{1}{3}$ in the limiting case of infinitesimally thin discs) and numerically (e.g., Icke 1982: $\frac{1}{3}$–$\frac{1}{2}$; Villumsen 1983, 1985a,b: $\frac{1}{4}$–$\frac{1}{2}$; Kamahori and Fujimoto 1987: $\frac{1}{3}$). Differences in the results can be ascribed to the different approaches and approximations employed in the various cases. However, they all predict observationally consistent power-laws for the age-dependence of the stellar velocity dispersion (as indicated at the side of each reference) except Kamahori and Fujimoto (1986a), where the observed saturation of its components is clearly due to a wrong treatment of the dynamical friction in the framework of the Langevin approach.

5.2.2 Hypothetical Massive Halo Objects

Stimulated by the ever more growing interest in the problem of dark matter in the universe, a number of authors have recently speculated upon the existence of massive ($\sim 10^6\ M_\odot$) halo objects, as massive black holes and dark clusters, as possible candidates for the heating mechanism invoked in galactic discs (e.g., Lacey 1984b; Lacey and Ostriker 1985; Ipser and Semenzato 1985; see also Ipser and Semenzato 1983; Kamahori and Fujimoto 1986b, 1987; Carr and Lacey 1987).

Dark clusters produce similar heating effects as massive black holes, but have the advantage of circumventing some of the problems inherent in the black-hole model. In particular, the dynamical friction is prevented from building up too much mass at the galactic centre if the clusters are disrupted by mutual collisions or tidal effects before the dynamical friction can become operative, and the problem that these halo objects may generate too much luminosity through accretion is avoided.

The methods employed in this case are substantially the same as those used in the case of GMCs. The calculations are anyway more complicated because the velocity dispersion of these hypothetical halo objects cannot of course be neglected as well as their vertical distribution. For suitable values of some free parameters the results are in agreement with observations, but the large number of these free parameters makes indeed the theory not highly predictive.

5.3 The Role of Transient Spiral Waves

A different point of view was introduced by Barbanis and Woltjer (1967), who stressed the importance that large-scale phenomena as spiral waves[5] might have in the secular evolution of the components of the stellar velocity dispersion parallel to the galactic disc. A heuristic argument was also presented to show that the same heating mechanism might account for the increase of the vertical component of the velocity dispersion with age as well.[6] Their analysis does not make reference to any specific formulation of the spiral structure theory. They only investigated, in fact, the effect of an imposed spiral potential of a particular form on the epicyclic motion of disc stars. In this sense the spiral waves they considered are not self-sustained.

Their suggestion that recurrent transient spiral waves might naturally heat galactic discs, even though no explicit time-dependence was derived, lies at the basis of further analytical (e.g., Byl 1974; Carlberg 1984, 1987; Carlberg and Sellwood 1985; Binney and Lacey 1988) and numerical investigations (e.g., Carlberg and Sellwood 1983; Sellwood and Carlberg 1984; see also Renz 1985; Carlberg and Freedman 1985; Sellwood and Lin 1989). In addition, the combined effect of spiral waves and GMCs has been studied by Carlberg (1987) and Jenkins and Binney (1990) employing Monte Carlo simulations for integrating the orbits and the orbit-averaged Fokker-Planck

[5] Small-scale irregular spiral features can be induced as a wake by massive perturbers in the galactic plane (Julian and Toomre 1966; see also: Thorne 1968; Saslaw 1985), and thus their influence on the secular evolution of the stellar velocity dispersion can be studied in that context by taking collective effects into account (Julian 1967).

[6] The suspicion that spiral-arm formation might be the dominant relaxation mechanism in galactic discs had already been expressed by Goldreich and Lynden-Bell (1965).

equation, respectively; and by Yasutomi and Fujimoto (1989) employing numerical simulations.

In this context it should be noted that the restriction to transient non-self-sustained spiral waves, while on the one hand it avoids the theoretical difficulties arising from the consideration of the Poisson equation and from all its physical implications (i.e., excitation mechanisms in self-sustained spiral waves, etc.), on the other hand at the same time lowers the level of predictability of the theory—which is, in fact, less constrained. Moreover, recall that some of the above-mentioned 2-dimensional numerical simulations might be biased by spurious collisional relaxation (see Section 2.2 and specifically White 1988; Schroeder and Comins 1990).

The theoretical framework of these recent investigations (except the confused attempt made by Byl 1974) is a simplified formulation of the quasi-linear theory in the action-angle canonical representation. The simplification corresponds indeed not to care about the self-consistency of the theory (hence, only the quasi-linear diffusion equation is considered), which instead is expected to have a crucial role (see Section 6.3 of Part II). Only spiral waves varying on a timescale comparable to the basic periods of oscillation in the disc are considered, because otherwise counter-reacting relaxation mechanisms would take adiabatically the system back to its initial unperturbed state without any appreciable dynamical effect.

If transient spiral waves of a particularly simple form recur at a constant rate in time, then the increase of the planar velocity dispersion with age for a coeval population of disc stars follows a $\frac{1}{2}$-power-law. However, as the velocity dispersion becomes as large as to make the size of the epicycles comparable to the wavelength of the spiral wave, the horizontal-heating rate follows a $\frac{1}{5}$-power-law.

On the other hand, the vertical heating associated with such transient spiral waves turns out to be completely inefficient, because their typical pattern speed is much smaller than the natural frequency of vertical oscillation so that only adiabatic effects are produced. Carlberg (1984) suggested that a way of overcoming this difficulty in the context of a similar scenario is to invoke the existence of bending waves, whose observational counterparts are the well-known warps in spiral galaxies. A more sceptic point of view has been expressed by Carlberg (1987).

5.4 More General Approaches

Since the basic physical mechanism responsible for the heating of galactic discs is not well-known at present, a phenomenological description of the heating process by the theory of orbital diffusion seems to be rather adequate. This line was first pursued by Wielen (1977), Wielen and Fuchs (1983, 1985), Fuchs and Wielen (1987). More precisely, in this approach the heating of galactic discs is basically described by a diffusion process in velocity space, in which the dynamical friction is not taken into account and the diffusion coefficient is empirically determined from the observed age-dependence of the components of the velocity dispersion of nearby stars. The advantage of such an empirical procedure lies in the fact that it avoids as far as possible uncertain assumptions on the basic physical source of the irregular part of the galactic gravitational field, apart from those inherent in the choice of a diffusion process among all the possible stochastic processes.

Wielen (1977) showed that a constant (i.e., time- and velocity-independent) diffusion coefficient, despite its extremely simple form, can explain fairly well both the age-dependence of the components of the stellar velocity dispersion (a $\frac{1}{2}$-power-law is obtained) and the axial ratios of the velocity ellipsoid. It turns out also that for a constant diffusion coefficient the Fokker-Planck equation admits self-similar solutions of Schwarzschild type (i.e., gaussian distribution functions with an anisotropic time-dependent velocity dispersion), provided the radial gradient of the (axi-symmetric) distribution function is neglected with respect to its vertical variation and the epicyclic approximation is used (Wielen and Fuchs 1983; cf. Renz 1985). This result is extremely important in view of the relevance that the Schwarzschild distribution function has on observational grounds. Although it is appealing due to its simplicity, a constant diffusion coefficient is not the only physically relevant one. Other more physically meaningful choices of isotropic (in velocity space) time-dependent diffusion coefficients can satisfactorily mimic the observed behaviour of the components of the stellar velocity dispersion as well (Wielen 1977; Fuchs and Wielen 1987).

A more detailed analysis (Wielen and Fuchs 1985) suggests that the stochastic heating is the main relaxation process in galactic discs, while other processes as the adiabatic cooling and the infall of gas from haloes are only of secondary importance from a dynamical point of view.

A similar approach has recently been undertaken by Binney and Lacey (1988), who transposed some of the Wielen results in the action-angle canonical representation, more elegant from a formal point of view but also less predictive when it is applied to real physical situations (see Section 4.4). As particular cases of heating mechanisms they considered the effects of GMCs and transient spiral waves, for which they derived the quasi-linear diffusion equation within the framework of the Fokker-Planck approach calculating the diffusion tensor by means of the Hamilton perturbation theory.

They showed that both these heating mechanisms are inconsistent with the "observed" $\frac{1}{2}$-power-law for the age-dependence of the components of the stellar velocity dispersion (bear in mind, however, the low confidence level of such observations; see Section 1.2 and cf. Subsection 5.2.1). This fact led them to the conclusion that other relaxation mechanisms, as the scattering of disc stars by massive halo objects, might play a major role in the stochastic heating of galactic discs.

5.5 Another Class of Explanations

As mentioned in the introductory Section 5.1, there is another class of explanations which interprets the observed increase of the stellar velocity dispersion with age in terms of native properties of disc stars. Tinsley and Larson (1978) suggested that the kinematics of stars older than 10^9 yr can be explained by a gradual decay of turbulent motions, as is predicted by certain extremely slow collapse models, and showed that the correlation between velocity dispersion and metallicity predicted by such models is in agreement with observations.

This effect cannot directly account for the rapid variation of the velocity dispersion with age observed even for stars younger than 10^9 yr, but they suggested that this could be explained if the velocity dispersion of younger stars reflects only the local turbulent motions in the gas, while the velocity dispersion of older stars reflects in addition larger-scale non-circular motions in the galactic gas layer. If the interstellar medium possesses a hierarchy of motions whose velocity dispersion increases with the size of the region considered, older stars, which have travelled farther since their formation, will experience gas motions over a larger space volume and thus will acquire larger velocity dispersions than younger stars.

This possibility was further investigated by Larson (1979). The relation between the gaseous velocity dispersion and the region size that is required if such interstellar motions are to explain the dependence of the stellar velocity dispersion c on age t can be estimated from the empirical relation $c \sim t^{1/2}$. If c is equal to the velocity dispersion of the gas in a region of size L in which the stars of age t have originated, then we obtain $c \sim L^{1/3}$ since $L \sim ct$.

The agreement between this power-law and the Kolmogorov spectrum for incompressible turbulence is suggestive, if perhaps only accidental. The Kolmogorov law depends on the assumption that energy is successively transferred into motions on ever smaller scales until it is entirely dissipated by viscosity. In general this is not expected if the motions are supersonic, as in the interstellar medium, since energy can then directly be dissipated on large scales by shock fronts. This leaves less energy for small-scale motions, and produces a steeper dependence of c on L.

Data assembled by Larson (1979) from a variety of sources show indeed that the velocity dispersion of young stars and of the cold interstellar gas increases systematically with the size of the region considered over a wide range of lengthscales, and this effect is sufficient to account for the observed age-dependence of the velocity dispersion of disc stars for ages up to about 10^9 yr. The observed dependence of the gas velocity dispersion on region size suggests the existence of a hierarchy of turbulent motions in which smaller-scale motions are produced by the turbulent decay of larger-scale motions.

References

Barbanis B. and Woltjer L. (1967), *Ap. J.* **150**, 461.

Berry C. L. (1973), *Ap. J.* **179**, 395.

Binney J. and Lacey C. G. (1988), *M.N.R.A.S.* **230**, 597.

Binney J. and Tremaine S. (1987), *Galactic Dynamics*, Princeton University Press, Princeton.

Byl J. (1974), *M.N.R.A.S.* **169**, 157.

Byl J. and Ovenden M. W. (1973), *M.N.R.A.S.* **164**, 289.

Carlberg R. G. (1984), in *Formation and Evolution of Galaxies and Large Structures in the Universe*, Reidel, Dordrecht, Audouze J. and Tran Thanh Van J. eds., p. 343 (La Plagne, France, 1983).

Carlberg R. G. (1987), *Ap. J.* **322**, 59.

Carlberg R. G. and Freedman W. L. (1985), *Ap. J.* **298**, 486.

Carlberg R. G. and Sellwood J. A. (1983), in *IAU Symp.* **100**, Reidel, Dordrecht, Athanassoula E. ed., p. 127 (Besançon, France, 1982).

Carlberg R. G. and Sellwood J. A. (1985), *Ap. J.* **292**, 79.

Carr B. J. and Lacey C. G. (1987), *Ap. J.* **316**, 23.

Chandrasekhar S. (1960), *Principles of Stellar Dynamics*, Dover, New York.

Freeman K. C. (1975), in *Galaxies and the Universe*, The University of Chicago Press, Chicago, Sandage A. et al. eds., p. 409.

Fuchs B. and Wielen R. (1987), in *The Galaxy*, Reidel, Dordrecht, Gilmore G. and Carswell B. eds., p. 375 (Cambridge, UK, 1986).

Fujimoto M. (1980), *Publ. Astron. Soc. Japan* **32**, 89.

Goldreich P. and Lynden-Bell D. (1965), *M.N.R.A.S.* **130**, 125.

Hénon M. (1973), in *Dynamical Structure and Evolution of Stellar Systems*, Geneva Observatory, Sauverny, Martinet L. and Mayor M. eds., p. 183 (Saas-Fee, Switzerland, 1973).

Icke V. (1982), *Ap. J.* **254**, 517.

Ipser J. R. and Semenzato R. (1983), *Ap. J.* **271**, 294.

Ipser J. R. and Semenzato R. (1985), *A. Ap.* **149**, 408.

Jenkins A. and Binney J. (1990), *M.N.R.A.S.* **245**, 305.

Julian W. H. (1967), *Ap. J.* **148**, 175.

Julian W. H. and Toomre A. (1966), *Ap. J.* **146**, 810.

Kamahori H. and Fujimoto M. (1986a), *Publ. Astron. Soc. Japan* **38**, 77.

Kamahori H. and Fujimoto M. (1986b), *Publ. Astron. Soc. Japan* **38**, 151.

Kamahori H. and Fujimoto M. (1987), *Publ. Astron. Soc. Japan* **39**, 201.

Lacey C. G. (1984a), *M.N.R.A.S.* **208**, 687.

Lacey C. G. (1984b), in *Formation and Evolution of Galaxies and Large Structures in the Universe*, Reidel, Dordrecht, Audouze J. and Tran Thanh Van J. eds., p. 351 (La Plagne, France, 1983).

Lacey C. G. (1985), in *IAU Symp.* **106**, Reidel, Dordrecht, van Woerden H. et al. eds., p. 493 (Groningen, The Netherlands, 1983).

Lacey C. G. and Ostriker J. P. (1985), *Ap. J.* **299**, 633.

Larson R. B. (1979), *M.N.R.A.S.* **186**, 479.

Lynden-Bell D. (1962), *M.N.R.A.S.* **124**, 23.

Mihalas D. and Binney J. (1981), *Galactic Astronomy*, Freeman, San Francisco.

Renz W. (1985), in *IAU Symp.* **106**, Reidel, Dordrecht, van Woerden H. et al. eds., p. 497 (Groningen, The Netherlands, 1983).

Saslaw W. C. (1985), *Gravitational Physics of Stellar and Galactic Systems*, Cambridge University Press, Cambridge.

Schroeder M. C. and Comins N. F. (1989), *Ap. J.* **346**, 108.

Sellwood J. A. and Carlberg R. G. (1984), *Ap. J.* **282**, 61.

Sellwood J. A. and Lin D. N. C. (1989), *M.N.R.A.S.* **240**, 991.

Semenzato R. (1987), *A. Ap.* **175**, 50.

Spitzer L. Jr. and Schwarzschild M. (1951), *Ap. J.* **114**, 385.

Spitzer L. Jr. and Schwarzschild M. (1953), *Ap. J.* **118**, 106.

Thorne R. M. (1968), *Ap. J.* **151**, 671.

Tinsley B. M. and Larson R. B. (1978), *Ap. J.* **221**, 554.

Villumsen J. V. (1983), *Ap. J.* **274**, 632.

Villumsen J. V. (1985a), *Ap. J.* **290**, 75.

Villumsen J. V. (1985b), in *IAU Symp.* **106**, Reidel, Dordrecht, van Woerden H. et al. eds., p. 491 (Groningen, The Netherlands, 1983).

White R. L. (1988), *Ap. J.* **330**, 26.

Wielen R. (1977), *A. Ap.* **60**, 263.

Wielen R. and Fuchs B. (1983), in *Kinematics, Dynamics and Structure of the Milky Way*, Reidel, Dordrecht, Shuter W. L. H. ed., p. 81 (Vancouver, Canada, 1982).

Wielen R. and Fuchs B. (1985), in *IAU Symp.* **106**, Reidel, Dordrecht, van Woerden H. et al. eds., p. 481 (Groningen, The Netherlands, 1983).

Woltjer L. (1967), in *Relativity Theory and Astrophysics*, Vol. 2, Am. Math. Soc., Providence, Ehlers J. ed., p. 1 (Ithaca, New York USA, 1965).

Yasutomi M. and Fujimoto M. (1989), *Publ. Astron. Soc. Japan* **41**, 823.

Part II

MY OWN CONTRIBUTION: PROPOSALS AND RESULTS

Chapter 6

Self-Regulation Mechanisms in Galactic Discs

Summary

The secular evolution of galactic discs, of which the increase of the stellar velocity dispersion with age is the most striking expression from a kinematical point of view, is closely related to their stability properties because of the collective nature of such systems. In this context, however, the crucial role of collective effects is often underestimated or not properly taken into account.

We propose a global collective heating mechanism leading to a self-regulation process of the kind suggested by the spiral structure theory, when both the linear effects of wave-wave interactions and the quasi-linear effects of wave-particle interactions at the relevant resonances are taken into account. The cold interstellar gas is expected to play a crucial role in ensuring self-regulation together with the internal excitation and feedback mechanisms invoked for the maintenance of global spiral modes. As a result, the planar and vertical components of the stellar velocity dispersion are expected to have a different age-dependence. Some observational evidences in support of this qualitative prediction are also discussed.

6.1 Introduction

A careful inspection of the heating mechanisms in galactic discs described in Chapter 5 of Part I shows that none of them takes collective effects properly into account. This is indeed a very severe restriction, because collective effects are known to play a *crucial* role in systems whose dynamics is governed by long-range interactions.

It may be objected that the heating mechanism by recurrent transient large-scale spiral waves proposed by Barbanis and Woltjer (1967) and further investigated by Carlberg and Sellwood (1985) (see also the references cited in Section 5.3 of Part I) represents an attempt to estimate the role that such effects have in driving the secular evolution of galactic discs. This is in part true, because the large-scale spiral structure observed in disc galaxies is a visible manifestation of them. The problem, however, lies in the fact that these simplified models do not retain an *essential* ingredient of the phenomenon, which derives from the self-sustenance property[1] of spiral waves. In fact, when a global linear analysis is performed taking the self-gravity of the perturbations into account, these modes are found to be maintained by internal excitation and feedback mechanisms, which a local treatment is not able to predict. In this context the role played by the resonances turns out to be *crucial*.

While on the one hand stellar dynamics tends to give too much emphasis to binary relaxation processes, on the other hand the spiral structure theory suggests the existence of global self-regulation mechanisms, due to collective effects, which lead to a secular increase of the stellar planar velocity dispersion in such a way as to saturate otherwise exponentially growing overstabilities. In simpler local self-regulation processes, which do not take internal excitation and feedback mechanisms into account, the stellar planar velocity dispersion is expected to settle at, or to exceed, a critical value which ensures a situation of local stability of the system at all wavelengths (Toomre 1964).

The importance of such self-regulation mechanisms in the dynamics and long-term evolution of spiral galaxies, in which the cold interstellar gas is expected to provide a welcome source of cooling, has also been suggested by various numerical simulations (see the references cited in Section 7.1). Bear in mind, however, that the interpretation of these results in terms of actual physical processes is not an easy task, and sometimes can also be misleading for other reasons than those discussed in Section 2.2 of Part I (see Lin and Bertin 1985, and cf. Dawson 1983 for the case of computer experiments in plasma physics). One of such difficulties derives from the fact that the requirement on the number of particles is highly demanding in order to get a proper simulation of processes involving resonances, where phase mixing and Landau damping are important.

These self-regulation mechanisms, together with the *crucial* role that the cold interstellar gas plays in them, will be discussed in more detail in this

[1]To neglect the self-gravity of the perturbations may reasonably be justified only in the case of small-scale (with respect to the typical wavelengths contributing to the instabilities of the system) spiral features when collisionless phase-mixing mechanisms are studied (Lynden-Bell 1962).

chapter. Our theoretical approach to the problem of the secular heating of galactic discs (Romeo 1987, 1989, 1990a,b) is perspectively outlined in Figure 6.1.

6.2 Local Self-Regulation Mechanisms

While on the one hand the formulation of a global quasi-linear theory of spiral structure is required to account for the secular heating of galactic discs at a quantitative level, on the other hand some physically relevant qualitative conclusions can even be drawn by considering simpler local approaches of the kind proposed by Bertin and Romeo (1988). The spiral structure theory, when the cold interstellar gas is taken into account, predicts in fact the existence of local self-regulation mechanisms which produce a *rapid* increase of the stellar radial velocity dispersion up to a quasi-stationary (*secularly* increasing) critical value that ensures a situation of local stability of the system at all wavelengths.

These local self-regulation processes are based on the following competing mechanisms:

- The stars of the active disc tend to heat up via gravitational instabilities, and possibly because of interaction with GMCs. The former, being a collective heating mechanism, is very sensitive to the local level of stability. Now obvious cooling mechanisms are in hand.

- The cold interstellar gas tends to cool down on a short timescale via turbulent dissipation due to inelastic GMC-GMC collisions. This cooling mechanism is not sensitive to the local level of stability. (Actually, a fraction of the cold component, which here is identified with the cold interstellar gas, consists of newly born stars and thus is not subject to cooling. This fraction slowly increases in time as a result of star-formation processes.) It also suffers heating via the same gravitational instabilities, but at a faster rate because of the stronger reaction of the thinner cold component.

- Cooling is a source of dynamical instabilities and thus generates heating, ensuring self-regulation.

A more specific discussion of the effects of such local self-regulation mechanisms on the long-term evolution of galactic discs is given in Subsection 7.2.3 in the framework of a two-fluid approach. Note in this context that a fluid description, obtained imposing a proper closure of the moments of the collisionless Boltzmann equation, is more convenient for investigating the stability properties of purely stellar discs analytically, and reproduces the main results almost completely except near the Lindblad resonances (see, e.g., Berman and Mark 1977, 1979; Sygnet, Pellat and Tagger 1987).

If the cold interstellar gas is not taken into account, the stellar radial velocity dispersion is expected to settle at, or to exceed, a smaller local critical value which remains *fixed* in time, so that no secular evolution occurs. In local self-regulation processes, in fact, internal excitation and feedback mechanisms contributing to the secular evolution of the stellar velocity dispersion are not taken into account. Moreover, of even more basic relevance, the cold interstellar gas prevents by turbulent dissipation an excessive heating of the active stellar disc, ensuring self-regulation. These facts show how *crucial* is the role of the cold interstellar gas in the physical picture of spiral galaxies.

6.3 Global Self-Regulation Mechanisms

As we have discussed in Section 4.4 of Part I, a number of attempts have been made to extend the quasi-linear theory of plasma waves to the gravitational case as well. The major difficulty which is to be tackled for achieving a satisfactory formulation of this theory consists in the necessity of using a global approach. The spiral structure theory, when internal excitation and feedback processes are taken into account, suggests in fact the existence of global self-regulation mechanisms which lead to a *secular* increase of the stellar radial velocity dispersion in such a way as to saturate otherwise exponentially growing overstabilities, in analogy with the phenomenon of turbulent heating predicted in the framework of the local quasi-linear theory of plasma waves. The following discussion is just devoted to explain these concepts in more detail, and to present the basic ideas underlying the formulation of a *global* quasi-linear theory of spiral structure.

Let us first consider the case of a self-gravitating purely stellar disc. As in the case of plasma waves the basis of this theory is the quasi-linear approximation. Taking into account the fact that two considerably different timescales are involved, the stellar distribution function can be split into two terms: a slowly varying part describing the collisionless collective relaxation process, and a small perturbation describing the self-sustained oscillations of the system. The perturbation distribution function is then treated as in the linear theory, whereas the long-term evolution of the quasi-equilibrium distribution function is determined by second-order terms in the perturbation representing wave-particle interactions.

To be more specific, perturbations of spiral form

$$f_1 = \tilde{f}_1(r, z; v_r, v_\theta, v_z)\, e^{i\,(m\theta - \omega t)} \tag{6.1}$$

are self-consistently imposed on an axi-symmetric and plane-symmetric basic state

$$f_0 = f_0(r, z; v_r, v_\theta, v_z; t) \tag{6.2}$$

in differential rotation $\Omega(r)$, with the requirement

$$\left|\frac{\partial \ln f_0}{\partial t}\right| \ll \left|\frac{\partial \ln f_1}{\partial t}\right| . \tag{6.3}$$

In situations of astrophysical interest the secular timescale is of order 10^9–10^{10} yr, whereas the dynamical timescale is of order 10^8 yr. In writing these expressions a system of cylindrical coordinates has been used to take into account the approximate symmetry properties of galactic discs. Furthermore, $k = k(r) \equiv -i\,(\partial \ln f_1/\partial r)$ is the complex radial wave-number of the perturbation, whose radial dependence takes the inhomogeneity of the system into account; m is the number of spiral arms; $\Omega_\mathrm{p} \equiv \Re(\omega)/m$ and $\gamma \equiv \Im(\omega)$ are the pattern frequency and the growth (or damping) rate of the spiral wave, respectively. In a local approach the radial dependence of the perturbations is partially explicited

$$f_1 = \hat{f}_1(r, z; v_r, v_\theta, v_z) \exp\left[i\left(\int^r k(r')\,dr' + m\theta - \omega t\right)\right] \tag{6.4}$$

(with this convention trailing waves are characterized by $k > 0$), and the further assumptions $|\partial \ln \hat{f}_1/\partial r| \ll |k|$, $|\partial k/\partial r| \ll k^2$ are made to obtain an "approximate" local dispersion relation in place of the "exact" global-mode equation.

In addition, the quasi-linear theory of spiral structure relies on the following working assumptions:

- The disc system is infinitesimally thin. As regards the perturbations, this assumption implies $|k|\langle z\rangle \ll 1$, where $\langle z\rangle$ is the thickness-scale of the system.

- The departures from circular orbits are small and satisfy the epicyclic approximation: $c_r/r\kappa \ll 1$, where c_r is the radial velocity dispersion and κ is the epicyclic frequency.

- The spiral waves are tightly wound: $m/|k|r \ll 1$. Consistently with this assumption, the radial gradient of f_0 is neglected with respect to that of f_1 and WKBJ asymptotic expansion techniques are employed.

- The winding and the epicyclic parameters are formally taken to be of the same order: $m/|k|r \sim c_r/r\kappa \ll 1$.

The same assumptions are also commonly invoked in the linear theory of spiral structure to make the system of the (coupled) Vlasov and Poisson equations more tractable. In a local approach, when finite-thickness effects are taken into account, the form of the dispersion relation remains the same provided the unperturbed surface density is multiplied by a suitable reduction factor. Furthermore, another reduction factor lowers the response of high-velocity dispersion stars (see Subsection 9.2.1).

A quasi-linear diffusion equation describing the secular evolution of the quasi-equilibrium distribution function can be derived along a line similar to that followed in the case of plasma waves. The wave-particle resonances which play a dominant role are the corotation resonance

$$\frac{m\,[\Omega_{\rm p} - \Omega(r_{\rm co})]}{\kappa(r_{\rm co})} \equiv 0\,, \tag{6.5}$$

the inner Lindblad resonance

$$\frac{m\,[\Omega_{\rm p} - \Omega(r_{\rm ILR})]}{\kappa(r_{\rm ILR})} \equiv -1\,, \tag{6.6}$$

and the outer Lindblad resonance

$$\frac{m\,[\Omega_{\rm p} - \Omega(r_{\rm OLR})]}{\kappa(r_{\rm OLR})} \equiv +1\,, \tag{6.7}$$

well-known in the linear theory. The quasi-linear diffusion equation is supplemented by two closure equations, which in a local approach reduce to the dispersion relation derived in the linear regime and to a continuity equation with a source term for the number of grexons.

At this stage a global approach is required just because the diffusion tensor is dominated by the effects of wave-particle resonances, which cannot properly be described in the framework of a local approach (breakdown of the concept of local dispersion relation). Note, however, that a local treatment can still be suitable far away from the relevant wave-particle resonances, where the so-called adiabatic quasi-linear diffusion occurs. But what makes the use of a global approach really *essential* is the fact that in the local linear approach propagating spiral waves $[m\,(\Omega_{\rm p} - \Omega) \neq 0]$ turn out to be neutral $(\gamma = 0)$, whereas the actual situation is not so simple, as explained below.

Unfortunately, an exhaustive discussion cannot straightforwardly be given in this context, because the subject is extremely specific. We shall only try to express the basic ideas underlying the internal excitation and feedback mechanisms which make the maintenance of global linear spiral modes possible. For a more detailed description of these mechanisms reference is made to the original papers by Mark (1974a,b, 1976a,b,c, 1977) and to the review papers by Bertin (1980), Lin and Bertin (1985), Lin and Lau (1979).

The starting point is the global-mode equation deduced by combining the linearized Vlasov (or fluid) and Poisson equations. This equation, which is of Schrödinger type but not identical to the usual quantum-mechanical time-independent wave equation, exhibits two turning points: a first-order turning point, the bulge radius r_{ce}, and a second-order turning point, the corotation radius r_{co}. Far away from these singular points and from the Lindblad resonances ($m = 2$ models do not generally exhibit the inner Lindblad resonance) the global-mode equation approximately reduces to the standard local dispersion relation. The solution of this wave equation can be found by methods similar to those employed in quantum mechanics when a WKBJ (Wentzel, Kramers, Brillouin, Jeffreys) approach is used: the global solution is obtained by performing an asymptotic matching of the local solutions at the turning points, and by imposing a radiation condition at infinity (which in the specific case is represented by the outer Lindblad resonance). In particular, a quantum condition for the wave-number of Bohr-Sommerfeld type is found.

As regards these local solutions, three kinds of spiral wave have been studied so far: short waves, long waves and finally open waves, which have intermediate properties with respect to the others (they are more open than long waves, but they propagate similarly to short waves). In each case the trailing and leading configurations are possible. Only trailing spiral waves can propagate between the corotation and outer Lindblad resonances in such a way as to satisfy the radiation condition.

The global linear treatment shows that at the two turning points of the global-mode equation wave-wave interactions occur, which are at the basis of internal excitation and feedback mechanisms necessary for the maintenance of global spiral modes, once the damping role of the Lindblad resonances is taken into account. The simplest among such mechanisms (WASER), which is shown to occur when lowest-order terms in the WKBJ expansion are retained, involves trailing spiral waves alone. When a long trailing wave propagating away from the bulge enters the corotation region, it undergoes an over-reflection process in which two short trailing waves are produced:

- The reflected wave is always amplified, because the energy flux associated with the wave changes sign when it passes through the corotation resonance. This wave propagates back towards the bulge, where it is turned into a long trailing wave by a feedback mechanism. During this cycle non-linear effects act in such a way as to damp the wave, whereas the amplification process occurs mainly inside the corotation region.

- The transmitted wave propagates out towards the outer Lindblad resonance, where it is absorbed due to Landau damping mechanisms.

Another possible mechanism (swing amplification), which is shown to occur when higher-order terms in the WKBJ expansion are retained, gives rise to an over-reflection process involving both trailing and leading open waves.

We expect the same kinds of internal excitation and feedback mechanism to occur also at the quasi-linear level, because only the evolution of the slowly varying part of the distribution function is determined by second-order terms in the perturbations, and this in turn affects only the form of the eigenvalue and not the form of the global-mode equation. Therefore, the relevant overstabilities would not be produced by non-monotonic features in the velocity distribution, as instead occurs in plasmas in the cases generally considered by the quasi-linear theory. The only complication which might invalidate the linear results at a quantitative level lies in the fact that in the linear approach a time-independent properly modified Schwarzschild distribution function is assumed, while it is not known a priori whether the form of the diffusion tensor allows self-similar solutions of this type for the quasi-linear diffusion equation (cf. Section 5.4 of Part I). We recall that the use of local Schwarzschild distribution functions is mainly invoked on observational grounds. The effect that a different choice of the stellar distribution function may have in the global stability properties of galactic discs is not known yet (see Lin and Bertin 1985; Romeo 1985).

Although this physical picture is already extremely complicated, it is not yet satisfactorily complete. The *dual* dynamical role that the cold interstellar gas plays in the stability of galactic discs cannot in fact be disregarded. On the one hand, in the *linear* regime it can significantly destabilize the system, and in some pathological situations may even excite more complicated wave channels and cycles (see Chapter 7). On the other hand, it can be shocked and thus contributes, together with *non-linear* effects, to saturate otherwise exponentially growing spiral overstabilities, inhibiting excessive heating in the active stellar disc and ensuring self-regulation.

6.4 Self-Regulation Mechanisms: A Unified View

6.4.1 Proposed Global Collective Heating Mechanism

The proposed *global collective* heating mechanism can thus be characterized as follows. The rapid phase of local self-regulation processes is expected to dominate the initial evolution of the stellar radial velocity dispersion up to its critical value for local stability. On comparable timescales phase-mixing processes occur competitively. The subsequent secular evolution of the stellar radial velocity dispersion is expected to be governed both by local and global self-regulation processes.

Note that there is no contradiction in expecting local and global self-regulation mechanisms to act simultaneously in the secular phase. Recall, in fact, that the global-mode linear analysis considers propagating spiral waves, and the propagation condition indeed coincides with the stability condition derived in the framework of the local linear analysis.

6.4.2 Qualitative Predictions and Observational Evidences

It is clear that the proposed global collective heating mechanism can only be effective in the galactic plane, because spiral waves propagate in it and their typical pattern speed is much smaller than the natural frequency of vertical oscillation so that only adiabatic effects are produced. The corresponding vertical heating is thus expected to be almost vanishing, and the consideration of finite-thickness effects cannot appreciably change the situation at all. Other global or local relaxation mechanisms, such as those associated with bending wave-star interactions or GMC-star encounters respectively, are surely more effective. The planar and vertical components of the stellar velocity dispersion are thus expected to have a *different* age-dependence.

In agreement with this qualitative prediction, a recent observational survey restricted to the solar neighbourhood (Strömgren 1987) suggests that the plane-parallel components of the stellar velocity dispersion increase markedly with age, their ratio showing no appreciable variation, whereas the vertical component soon stops at a nearly constant value. Note, however, that other recent observational surveys lead to radically different results, also from one another (see Section 1.2 of Part I).

In the galactic plane the proposed global collective heating mechanism is expected to be *dominant*, or at least competitive, with respect to the other local non-collective heating mechanisms so far invoked (see Chapter 5 of Part I).

There are indeed some observational evidences which seem to support and suggest this fact. Some normal spiral galaxies, whose most representative case is that of NGC 488, are characterized by relatively high stellar planar velocity dispersions (Kormendy 1985), whereas the stellar vertical velocity dispersions are heuristically expected to be comparatively low (see Romeo 1985 and also Section 7.4 for other implications in connection with their global stability properties). It seems reasonable to interpret such a strong "temperature" anisotropy as produced by radically different heating mechanisms, and to conclude that at least in such disc galaxies the planar heating mechanism is much more effective than the vertical one. This, indeed, is in agreement with our theoretical predictions.

In view of a future comparison between theoretical predictions and observational results, we now want to make a few considerations concerning some delicate points to which often is not paid a proper attention and which make such a comparison not straightforward.

From a theoretical point of view, a single equivalent stellar component is taken to be representative of the whole active stellar disc consisting of low-velocity dispersion stars (see Chapters 7–9). The consideration of more stellar populations would give rise to several complications due to their gravitational coupling via the Poisson equation, as required by the self-consistency condition. On the other hand, it should be noted that observations tend to overestimate the effect of high-velocity dispersion stars, which are not so dynamically relevant as regards their participation in spiral structure (see Lin and Bertin 1985; Romeo 1985).

From a theoretical point of view, a single gaseous component is taken to simulate HI regions of neutral atomic hydrogen and giant molecular clouds and complexes (see Chapters 7–9). On the other hand, observational surveys do often provide significantly different estimates as regards the molecular hydrogen (see Section 7.1).

References

Barbanis B. and Woltjer L. (1967), *Ap. J.* **150**, 461.

Berman R. H. and Mark J. W.-K. (1977), *Ap. J.* **216**, 257.

Berman R. H. and Mark J. W.-K. (1979), *Ap. J.* **231**, 388.

Bertin G. (1980), *Phys. Rep.* **61**, 1.

Bertin G. and Romeo A. B. (1988), *A. Ap.* **195**, 105.

Carlberg R. G. and Sellwood J. A. (1985), *Ap. J.* **292**, 79.

Dawson J. M. (1983), *Rev. Mod. Phys.* **55**, 403.

Kormendy J. (1985), in *IAU Symp.* **106**, Reidel, Dordrecht, van Woerden H. et al. eds., p. 541 (Groningen, The Netherlands, 1983).

Lin C. C. and Bertin G. (1985), in *IAU Symp.* **106**, Reidel, Dordrecht, van Woerden H. et al. eds., p. 513 (Groningen, The Netherlands, 1983).

Lin C. C. and Lau Y. Y. (1979), *Stud. Appl. Math.* **60**, 97.

Lynden-Bell D. (1962), *M.N.R.A.S.* **124**, 23.

Mark J. W.-K. (1974a), in *IAU Symp.* **58**, Reidel, Dordrecht, Shakeshaft J. R. ed., p. 417 (Canberra, Australia, 1973).

Mark J. W.-K. (1974b), *Ap. J.* **193**, 539.

Mark J. W.-K. (1976a), *Ap. J.* **203**, 81.

Mark J. W.-K. (1976b), *Ap. J.* **205**, 363.

Mark J. W.-K. (1976c), *Ap. J.* **206**, 418.

Mark J. W.-K. (1977), *Ap. J.* **212**, 645.

Romeo A. B. (1985), *Tesi di Laurea*, University of Pisa and Scuola Normale Superiore, Pisa, Italy.

Romeo A. B. (1987), *M. Phil. Thesis*, SISSA, Trieste, Italy.

Romeo A. B. (1989), *Preprint, SISSA Astro. 105*, Trieste, Italy (enlarged version of Romeo 1990a).

Romeo A. B. (1990a), "Stability and Secular Heating of Galactic Discs", to appear in *IAU Symp.* **146** (Paris, France, 1990).

Romeo A. B. (1990b), "Self-Regulation and Secular Heating of Galactic Discs", to appear in *IAU Eur. Reg. Astron. Meeting* **12** (Davos, Switzerland, 1990).

Strömgren B. (1987), in *The Galaxy*, Reidel, Dordrecht, Gilmore G. and Carswell B. eds., p. 229 (Cambridge, UK, 1986).

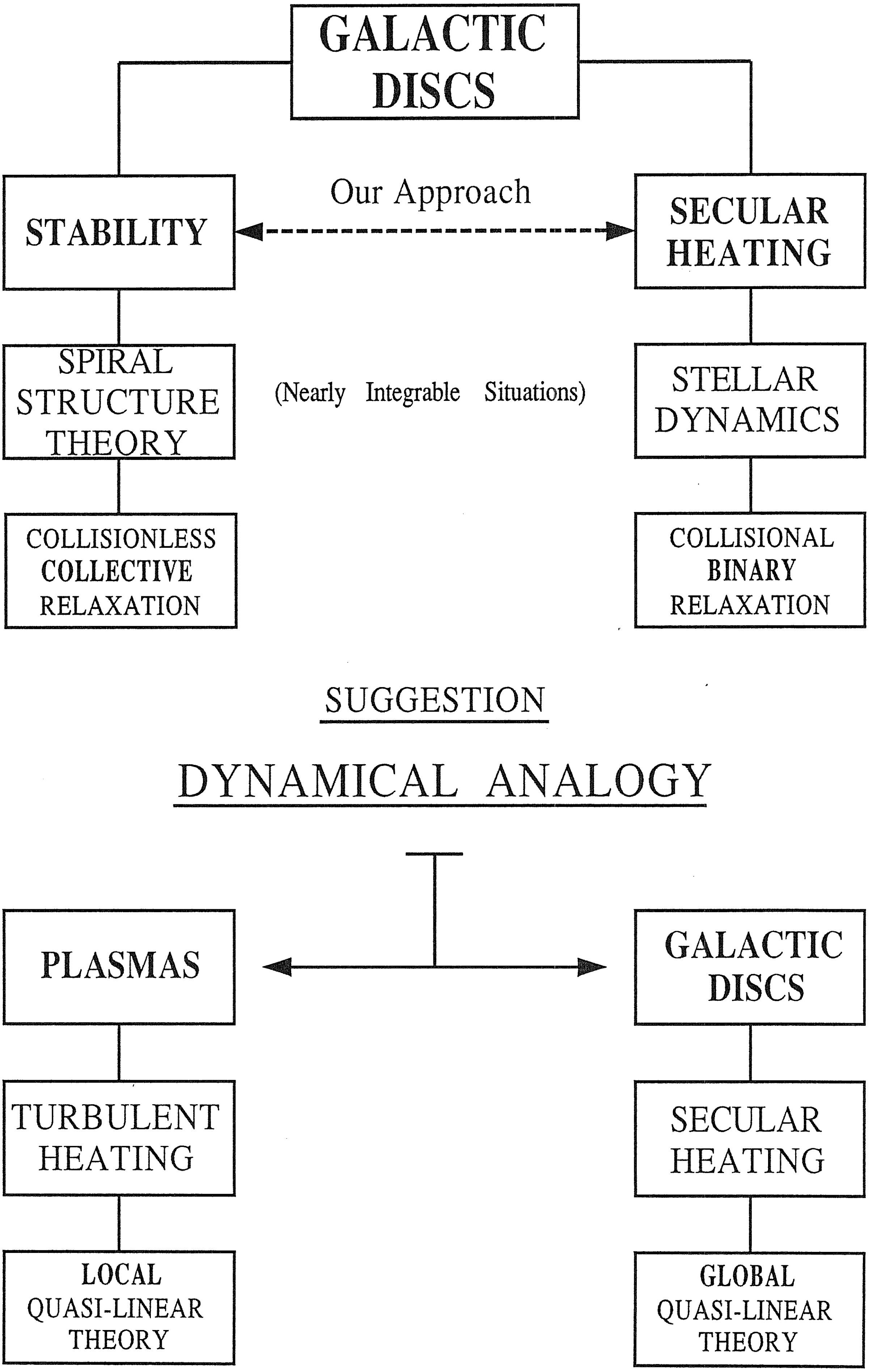
Sygnet J. F., Pellat R. and Tagger M. (1987), *Phys. Fluids* **30**, 1052.

Toomre A. (1964), *Ap. J.* **139**, 1217.

Figure Captions

Figure 6.1. Perspective outline of our theoretical approach to the problem of the secular heating of galactic discs.

Figure 6.1



GALACTIC DISCS
STABILITY
Our Approach
SECULAR HEATING
SPIRAL STRUCTURE THEORY
(Nearly Integrable Situations)
STELLAR DYNAMICS
COLLISIONLESS COLLECTIVE RELAXATION
COLLISIONAL BINARY RELAXATION
SUGGESTION
DYNAMICAL ANALOGY
PLASMAS
GALACTIC DISCS
TURBULENT HEATING
SECULAR HEATING
LOCAL QUASI-LINEAR THEORY
GLOBAL QUASI-LINEAR THEORY

Chapter 7

The Cold Interstellar Gas and the Stability of Galactic Discs: Infinitesimally Thin Systems

Summary

Most theoretical investigations into the spiral structure of galaxies are based on one-component models, because only low-velocity dispersion stars seem to play a fundamental role. However, it has long been recognized that in some cases also the contribution of the cold interstellar gas can be important because of its low turbulent velocity dispersion, although it represents a small fraction of the total mass in normal spiral galaxies. Our analysis is devoted to such cases.

We first perform a local linear stability analysis. It is found that in some regimes of astrophysical interest the role of the cold interstellar gas can even be dominant at short wavelengths.

The results obtained in this context are used to investigate global spiral modes in regimes which are expected to be associated with normal spiral structure. We use two-component equilibrium models which incorporate the essential features of the cold interstellar gas, as suggested by some recent observational surveys. Appreciable modifications to the structure of the modes, with respect to the corresponding one-component cases, are present only when a peaked distribution of molecular hydrogen is simulated. However, even in the cases where no qualitative modifications are present, the basic states which support these modes are characterized by relatively high stellar planar velocity dispersions, i.e. by values of the local stability param-

eter Q larger than unity. Finally, some qualitative predictions concerning the expected structure of global spiral modes in peculiar gas-dominated regimes (where a more complicated global analysis is required) are made.

7.1 Introduction

Galactic discs are found to consist of different populations of stars and gas components. A detailed stability analysis would be extremely complicated, also because it should take into account several physical processes and effects, so that one generally tries to single out the most important contributions from each of them.

From a dynamical point of view (spiral structure), only low-velocity dispersion stars seem to play a fundamental role. It is for this reason that most theoretical investigations into the spiral structure of galaxies are based on one-component models.

However, in some cases also the contribution of the cold interstellar gas can be important (see, e.g., Lin and Shu 1966; Graham 1967; see also Julian 1969; Lynden-Bell 1967; Marochnik and Suchkov 1969; see also Marochnik 1970; Vandervoort 1971; Kato 1972; Morozov 1981; Jog and Solomon 1984a,b; see also Jog 1985; Lubow 1986; Lubow, Balbus and Cowie 1986; Korchagin and Ryabtsev 1987; see also: Sweet 1963; Sweet and McGregor 1964; Harrison 1970; Biermann 1975; Fridman and Polyachenko 1984; Fridman et al. 1985; Contopoulos and Grosbøl 1986; Smith and Miller 1986; Contopoulos 1987; Min 1988; Contopoulos et al. 1989; Li 1990) or even *dominant* (Romeo 1985, 1987, 1988, 1989; Bertin and Romeo 1988) because of its low turbulent velocity dispersion, although it represents a small fraction of the total mass in normal spiral galaxies. Our analysis is devoted to such cases.

We shall now briefly discuss some observational studies which have attracted new interest in the dynamics of two-component systems.

Relatively recent radio data have shown that a substantial fraction of the interstellar gas in the Milky Way and also in some external galaxies could be in the form of molecular hydrogen. It is generally agreed that GMCs exhibit the following physical features:

- A peaked mass density distribution, in a ring-like fashion, in contrast to the essentially flat distribution which characterizes HI regions of neutral atomic hydrogen.
- A turbulent velocity dispersion of $\sim$ 4–8 $\mathrm{km\,s^{-1}}$, smaller than that relevant to HI regions ($\approx$ 8–10 $\mathrm{km\,s^{-1}}$).

However, the local values of their mass density as well as their turbulent velocity dispersion are known with great uncertainty and are still under debate, also because these estimates are based on indirect observational methods (see, e.g., Blitz and Shu 1980; Stark 1984; Bhat et al. 1985; Burton and Deul 1987; Dame et al. 1987; Knapp, Helou and Stark 1987; Solomon et al. 1987; Taylor, Dickman and Scoville 1987; Blitz 1988; Fleck 1988a,b,c; MacLaren, Richardson and Wolfendale 1988; Maloney 1988, 1990; Maloney and Black 1988; Myers and Goodman 1988a,b; Polk et al. 1988; Richardson and Wolfendale 1988a,b; Strong et al. 1988; Weliachew and Lucas 1988; Bloemen 1989; Broadbent, MacLaren and Wolfendale 1989; Elmegreen 1989; Heithausen and Mebold 1989; Kegel 1989; Sodroski et al. 1989; Somerville and Smith 1989; Stark and Brand 1989; Stecker 1989; Blitz, Bazell and Désert 1990; Bloemen, Deul and Thaddeus 1990; Devereux and Young 1990; Digel, Bally and Thaddeus 1990; Issa, MacLaren and Wolfendale 1990a,b; Leisawitz 1990; Martin, Hurwitz and Bowyer 1990; Sage, Shore and Solomon 1990; see also Zasov and Simakov 1988).

Another source of interest in the dynamics of two-component systems derives from the recent progress in measuring the stellar velocity dispersion in galactic discs. Values of the local stability parameter Q larger than unity, as reported for a number of gas-rich galaxies in some recent observational surveys (e.g., Casertano 1985, private communication; Kormendy 1985; van der Kruit and Freeman 1986; Bottema, van der Kruit and Freeman 1987; Freeman 1987; Bottema 1988, 1989a; see also Bottema 1989b, 1990; Lewis and Freeman 1989; see also: Efstathiou, Lake and Negroponte 1982; Sellwood 1985; Martinet 1988), can be explained, consistently with a mechanism of *self-regulation*, by taking the destabilizing role of the cold gas component into account. However, even in gas-poor galaxies relatively large values of the "observed" Q (see, e.g., Kormendy 1984a,b) can be justified by taking other factors into account (see Subsection 6.4.2).

The importance of such self-regulation mechanisms in the dynamics and long-term evolution of spiral galaxies, in which the cold interstellar gas is expected to provide a welcome source of cooling (see, e.g., Lin and Bertin 1985; Ostriker 1985; Shu 1985), has also been suggested by several numerical simulations (e.g., Miller, Prendergast and Quirk 1970; Hohl 1971; Quirk 1971; Carlberg and Sellwood 1983; Sellwood and Carlberg 1984; see also Renz 1985; Carlberg and Freedman 1985; Sellwood 1989; Thomasson et al. 1990).

For more information see Romeo (1985, 1987), Bertin and Romeo (1988), and references cited therein; see also Romeo (1988, 1989). Other more specific references concerning finite-thickness effects in two-component galactic discs will be given in Section 9.1.

7.2 Local Analysis

7.2.1 Assumptions

The System. We consider an infinitesimally thin, two-component, self-gravitating disc system. Each component is described by the standard Euler-continuity equations supplemented by a barotropic equation of state. The only interaction between the two components is taken to occur via the gravitational field (Poisson equation). The two components are denoted by different subscripts, "H" (HOT) and "C" (COLD), in order to recall that they are characterized by different equivalent acoustic speeds. Having in mind cases of astrophysical interest, we refer to them as the stars of the active disc (H) and the cold interstellar gas (C). However, we could also consider the case in which gas is absent but two stellar populations with different velocity dispersions can be identified.

The Basic State. In the basic state the system is taken to be axi-symmetric and in differential rotation $\Omega(r)$.

The Perturbations. Linear perturbations of spiral form

$$f_1 = \hat{f}_1 \exp\left[i\left(\int^r k(r')\,dr' + m\theta - \omega t\right)\right] \tag{7.1}$$

(with this convention trailing waves are characterized by $k > 0$) are studied under the ordering

$$\frac{m}{|k|r} \ll 1\,, \tag{7.2}$$

where m is the number of spiral arms. This is the approximation of tightly wound spiral structure, which is expected to be applicable to early normal spirals. Since we adopt an infinitesimally thin disc model, we should bear in mind, when dealing with observational implications, that the present discussion is restricted to relatively long waves $|k|\langle z\rangle \ll 1$, where $\langle z\rangle$ is the thickness-scale of the system. Moreover, a comparison with the more appropriate kinetic-fluid approach shows that the further restriction $c_{\rm H}/r\kappa \ll 1$ (epicyclic approximation) is required (the winding and the epicyclic parameters are formally taken to be of the same order), where here $c_{\rm H}$ represents the stellar radial velocity dispersion and κ is the epicyclic frequency.

7.2.2 Local Dispersion Relation

To lowest order the local dispersion relation written in a convenient dimensionless form is

$$\hat{\nu}^4 + \hat{\nu}^2|\tilde{k}|\left[|\tilde{k}|\,(1+\beta) - (1+\alpha)\right] + |\tilde{k}|^3\left[\beta|\tilde{k}| - (\alpha+\beta)\right] = 0\,, \tag{7.3}$$

where we have adopted the following scaling and parametrization:

$$\tilde{k} \equiv \frac{k}{\Sigma_{\rm H}}\,, \quad \text{where} \quad \Sigma_{\rm H} \equiv \frac{2\pi G\sigma_{\rm H}}{c_{\rm H}^2}\,; \tag{7.4}$$

$$\hat{\nu}^2 \equiv \frac{Q_{\rm H}^2}{4}(1-\nu^2)\,, \quad \text{where} \quad \nu \equiv \frac{\omega - m\Omega}{\kappa}\,, \quad Q_{\rm H} \equiv \frac{c_{\rm H}\kappa}{\pi G\sigma_{\rm H}}\,; \tag{7.5}$$

$$\alpha \equiv \frac{\sigma_{\rm C}}{\sigma_{\rm H}}\,, \quad \beta \equiv \frac{c_{\rm C}^2}{c_{\rm H}^2} \quad (0 < \alpha < +\infty,\ 0 < \beta < 1)\,. \tag{7.6}$$

In these formulae, σ is the unperturbed surface density, c is the equivalent acoustic speed, ν is the dimensionless Doppler-shifted frequency of the spiral perturbation, and the local parameter $Q_{\rm H}$ related to the hotter component is analogous to the **Local Stability parameter** Q for one-component systems. The cases $\alpha = 0$, $\alpha \to +\infty$ and $\beta = 1$ represent the limit of a one-component system.

Note that with a proper transformation of variables ($|\tilde{k}| \mapsto \hat{k}^2$, $\hat{\nu}^2 \mapsto -\hat{k}^2\hat{\omega}^2$) the discussion of this local dispersion relation can be given in complete analogy with the case of Jeans stability for a two-component, three-dimensional, homogeneous system in the absence of rotation.

For more information see Romeo (1985), where the dispersion properties of the wave branches and the more general case of n-component systems are also discussed, Bertin and Romeo (1988); see also Romeo (1988, 1989).

7.2.3 Local Stability

Marginal Stability Curve. Setting $\hat{\nu}^2 = Q_{\rm H}^2/4$, i.e. $\nu^2 = 0$, in the local dispersion relation (7.3), we obtain the marginal stability condition for the spiral perturbations. This can be seen as determining a value for $Q_{\rm H}^2$ when $|\tilde{k}|$ and α, β have been fixed. This value is positive in the range $0 < |\tilde{k}| < \tilde{k}_{\rm J} \equiv 1 + (\alpha/\beta)$. Adopting the more standard scaling

$$\bar{\lambda} \equiv \frac{k_{\rm H}}{|k|}\,, \quad \text{where} \quad k_{\rm H} \equiv \frac{\kappa^2}{2\pi G\sigma_{\rm H}}\,, \tag{7.7}$$

the marginal stability curve in the $(\bar{\lambda}, Q_{\rm H}^2)$ plane is defined by the following relation:

$$Q_{\rm H}^2 = \left(\frac{2\bar{\lambda}}{\beta}\right)\cdot\Big[(\alpha+\beta) - \bar{\lambda}(1+\beta) + \sqrt{\bar{\lambda}^2(1-\beta)^2 - 2\bar{\lambda}(1-\beta)(\alpha-\beta) + (\alpha+\beta)^2}\,\Big]\,, \tag{7.8}$$

which we consider in the range $0 \le \bar{\lambda} \le 1 + \alpha$.

Local Stability Criterion. From this relation a local stability criterion can be stated in analogy with the case of one-component systems. A function $\bar{Q}^2 = \bar{Q}^2(\alpha, \beta)$ can be defined in such a way that when $Q_{\rm H}^2 > \bar{Q}^2$ the system is locally stable at all wavelengths. For $Q_{\rm H}^2 < \bar{Q}^2$ we expect the system to be locally unstable in ranges of wavelengths defined by the marginal stability condition. The function $\bar{Q}^2$, which reduces to unity when $\alpha = 0$, plays a *crucial* role in the discussion of global spiral modes which we shall give in the next section.

Results. The properties of the marginal stability curve and the behaviour of the function $\bar{Q}^2(\alpha, \beta)$ are summarized in Figures 7.1–7.3, which show how effective a small amount of gas can be in destabilizing the system, provided the gas is sufficiently cold.

Secular Effects. When local self-regulation mechanisms of the kind discussed in Section 6.2 are considered, the stars of the active disc continually heat. In turn β decreases in the process, and the system moves along a horizontal line to the left of the (β, α) plane. Depending on the initial value of α, the system may suddenly experience a "phase transition" by encountering the two-phase region ($\alpha < \alpha_0$), or it may more smoothly move towards a gas-dominated regime directly from the original star-dominated regime ($\alpha > \alpha_0$). These processes of β-evolution are expected to occur secularly on a long timescale. A slow decrease of α can also be included to allow for star-formation processes.

For more information see Romeo (1985), Bertin and Romeo (1988); see also Romeo (1988, 1989).

7.3 Global Analysis

7.3.1 Proposals

The aim of this section is to evaluate whether and possibly how the presence of the cold interstellar gas modifies the morphology of global spiral modes with respect to the case of a one-component system, modelled as an active disc consisting of an equivalent (representative) population of low-velocity dispersion stars. This requires a proper choice of the basic states which support them, as explained below.

7.3.2 Choice of the Equilibrium Models

Methodology. Given a basic one-component equilibrium model

- $\Omega = \Omega(r)$,
- $\sigma = \sigma(r)$,
- $Q = Q(r)$, supposed to be consistent with a mechanism of *self-regulation*,

we modify it into a two-component equilibrium model according to the following prescription:

- $\Omega = \Omega(r)$;
- $\sigma_{\rm C} = \sigma_{\rm C}(r)$, $\sigma_{\rm H} = \sigma_{\rm H}(r) = \sigma(r) - \sigma_{\rm C}(r) \implies \alpha = \alpha(r)$;
- $c_{\rm C} = c_{\rm C}(r)$, $Q_{\rm H} = Q_{\rm H}(r) = Q(r)\,\bar{Q}[\alpha(r), \beta(r)] \implies \beta = \beta(r)$, where the last condition expresses the Self-Regulation constraint and the ratio $Q_{\rm H}/\bar{Q}$ plays the role of an Effective Q-parameter. In fact, the equation

$$\sqrt{\beta} = \frac{c_{\rm C}\kappa}{\pi G \sigma} \frac{1+\alpha}{Q\,\bar{Q}(\alpha,\beta)} \tag{7.9}$$

can be solved iteratively starting from

$$\sqrt{\beta^{(0)}} = \frac{c_{\rm C}\kappa}{\pi G \sigma} \frac{1}{Q}. \tag{7.10}$$

Recall in this context that the cold interstellar gas provides a welcome source of self-regulation for spiral instabilities, by inhibiting excessive heating in the stellar disc. This fact shows again how *crucial* is its role in the physical picture of spiral galaxies.

Practical Applications. The basic one-component equilibrium model, which has been derived from a family of models provided for us by Lowe (1988), is shown in Figure 7.4. Correspondingly, we have derived two two-component equilibrium models which incorporate the essential features of the cold interstellar gas, and whose (β, α) tracks lie outside the two-phase region in Star-dominated regimes. They are shown in Figures 7.5–7.6. Figure 7.5 shows also the (β, α) tracks for models of the galaxies NGC 4565 and NGC 5907 provided for us by Casertano (1985, private communication), where for simplicity it is assumed that all the cold interstellar gas is in the form of neutral atomic hydrogen with $c_{\rm C} = 8\ {\rm km\,s^{-1}}$.

Peculiar two-component equilibrium models involving Gas-dominated regimes will be considered in future applications.

For more information see Romeo (1985), Bertin and Romeo (1988); see also Romeo (1988, 1989).

7.3.3 Discrete Global Spiral Modes

Methodology. For a given equilibrium model the local dispersion relation (7.3) can be used to determine $k = k(r; m, \omega_R)$. The discrete global spiral modes are then calculated by imposing the quantum condition derived in the one-component case, which is expected to hold also for two-component systems in star-dominated regimes:

$$\oint k(r; m, \Omega_p)\, dr = (2n+1)\,\pi\,, \quad \text{where} \quad \Omega_p \equiv \frac{\omega_R}{m}\,, \tag{7.11}$$

taken between the inner turning point r_{ce}, where the bulge terminates, and the corotation circle r_{co}. This equation fixes the pattern frequency Ω_p of the (m, n) mode. The growth rate of the mode is inversely proportional to the propagation time taken along the relevant wave cycle.

Results. The propagation diagrams $k = k(r)$ which identify the relevant wave cycles and excitation mechanisms of the $(m = 2,\ n = 0)$ mode for our two-component equilibrium models are shown in Figure 7.7. When gas is included, we observe a shift of the mode towards a more tightly wound spiral structure with a smaller corotation radius. Note that, even in the case in which no qualitative modifications are present, the basic states which support this mode are characterized by relatively high stellar radial velocity dispersions, as shown in Figure 7.6, and interestingly enough the qualitative behaviour of the local stability parameter Q_H resembles that inferred from observations of gas-rich galaxies (see the references cited in the introductory Section 7.1).

For more information see Romeo (1985), Bertin and Romeo (1988); see also Romeo (1988, 1989).

7.3.4 Peculiar Gas-Dominated Equilibrium Models

How to Construct Them. Consider the reference one-component model used for deriving two-component models in more standard (star-dominated) regimes. The self-regulation constraint (7.9) shows that small values of β require relatively large values of α and small ratios $(c_C\kappa/\pi G\sigma)$, which can be obtained by changing properly the scale of the rotation curve and of the density distribution for a fixed c_C.

Qualitative Predictions. When the (β, α) tracks of the equilibrium model intersect the two-phase region, the use of the previous quantum condition (7.11) is no longer justified because we have a different (more complicated) eigenvalue problem. We expect to have "Double propagation diagrams" characterized by other turning points in addition to r_{ce} and r_{co}. Thus, more complicated wave channels and cycles are available ("Short and Long gaseous waves"). The structure of the modes is expected to be considerably modified. In particular, the presence of the gaseous peak at short wavelengths suggests a *high* degree of winding.

More comments are reported by Romeo (1985); see also Romeo (1988, 1989).

7.4 Astrophysical Implications

- Gas-rich early normal spiral galaxies should exhibit relatively *high* stellar planar velocity dispersions, consistently with a mechanism of self-regulation. *Non-monotonic* profiles are expected for relatively high H_2 densities.

- A new regime of *extremely* tightly wound spirals is expected when the stellar component is much hotter than the gaseous component in gas-rich early-type galaxies (e.g., NGC 488?). Correspondingly, a *discontinuity* in the behaviour of the stellar planar velocity dispersion profile may be observed.

For a more complete discussion see Romeo (1985); see also Romeo (1987, 1988, 1989).

References

Bertin G. and Romeo A. B. (1988), *A. Ap.* **195**, 105.

Bhat C. L., Issa M. R., Houston B. P., Mayer C. J. and Wolfendale A. W. (1985), *Nature* **314**, 511.

Biermann P. (1975), in *IAU Symp.* **69**, Reidel, Dordrecht, Hayli A. ed., p. 321 (Besançon, France, 1974).

Blitz L. (1988), *Preprint, IASSNS Ast. 15*, Princeton, New Jersey USA, to appear in *Millimeter and Submillimeter Astronomy* (Sterling, UK, 1987).

Blitz L. and Shu F. H. (1980), *Ap. J.* **238**, 148.

Blitz L., Bazell D. and Désert F. X. (1990), *Ap. J. Lett.* **352**, L13.

Bloemen H. (1989), *Ann. Rev. A. Ap.* **27**, 469.

Bloemen J. B. G. M., Deul E. R. and Thaddeus P. (1990), *A. Ap.* **233**, 437.

Bottema R. (1988), *A. Ap.* **197**, 105.

Bottema R. (1989a), *A. Ap.* **221**, 236.

Bottema R. (1989b), *A. Ap.* **225**, 358.

Bottema R. (1990), *A. Ap.* **233**, 372.

Bottema R., van der Kruit P. C. and Freeman K. C. (1987), *A. Ap.* **178**, 77.

Broadbent A., MacLaren I. and Wolfendale A. W. (1989), *M.N.R.A.S.* **237**, 1075.

Burton W. B. and Deul E. R. (1987), in *The Galaxy*, Reidel, Dordrecht, Gilmore G. and Carswell B. eds., p. 141 (Cambridge, UK, 1986).

Carlberg R. G. and Freedman W. L. (1985), *Ap. J.* **298**, 486.

Carlberg R. G. and Sellwood J. A. (1983), in *IAU Symp.* **100**, Reidel, Dordrecht, Athanassoula E. ed., p. 127 (Besançon, France, 1982).

Casertano S. (1985), private communication.

Contopoulos G. (1987), in *The Galaxy*, Reidel, Dordrecht, Gilmore G. and Carswell B. eds., p. 199 (Cambridge, UK, 1986).

Contopoulos G. and Grosbøl P. (1986), *A. Ap.* **155**, 11.

Contopoulos G., Gottesman S. T., Hunter J. H. Jr. and England M. N. (1989), *Ap. J.* **343**, 608.

Dame T. M. et al. (1987), *Ap. J.* **322**, 706.

Devereux N. A. and Young J. S. (1990), *Ap. J.* **359**, 42.

Digel S., Bally J. and Thaddeus P. (1990), *Ap. J. Lett.* **357**, L29.

Efstathiou G., Lake G. and Negroponte J. (1982), *M.N.R.A.S.* **199**, 1069.

Elmegreen B. G. (1989), *Ap. J.* **338**, 178.

Fleck R. C. Jr. (1988a), *Ap. J.* **328**, 299.

Fleck R. C. Jr. (1988b), *Ap. J.* **333**, 840.

Fleck R. C. Jr. (1988c), *Ap. J.* **333**, 845.

Freeman K. C. (1987), *Ann. Rev. A. Ap.* **25**, 603.

Fridman A. M. and Polyachenko V. L. (1984), *Physics of Gravitating Systems*, 2 Vols., Springer-Verlag, New York.

Fridman A. M., Morozov A. G., Palous J., Piskunov A. E. and Polyachenko V. L. (1985), in *IAU Symp.* **106**, Reidel, Dordrecht, van Woerden H. et al. eds., p. 509 (Groningen, The Netherlands, 1983).

Graham R. (1967), *M.N.R.A.S.* **137**, 25.

Harrison E. R. (1970), *M.N.R.A.S.* **150**, 443.

Heithausen A. and Mebold U. (1989), *A. Ap.* **214**, 347.

Hohl F. (1971), *Ap. J.* **168**, 343.

Issa M. R., MacLaren I. and Wolfendale A. W. (1990a), *Ap. J.* **352**, 132.

Issa M. R., MacLaren I. and Wolfendale A. W. (1990b), *A. Ap.* **236**, 237.

Jog C. J. (1985), in *IAU Symp.* **106**, Reidel, Dordrecht, van Woerden H. et al. eds., p. 505 (Groningen, The Netherlands, 1983).

Jog C. J. and Solomon P. M. (1984a), *Ap. J.* **276**, 114.

Jog C. J. and Solomon P. M. (1984b), *Ap. J.* **276**, 127.

Julian W. H. (1969), *Ap. J.* **155**, 117.

Kato S. (1972), *Publ. Astron. Soc. Japan* **24**, 61.

Kegel W. H. (1989), *A. Ap.* **225**, 517.

Knapp G. R., Helou G. and Stark A. A. (1987), *A. J.* **94**, 54.

Korchagin V. I. and Ryabtsev A. D. (1987), *Sov. Astron.* **31**, 16.

Kormendy J. (1984a), *Ap. J.* **286**, 116.

Kormendy J. (1984b), *Ap. J.* **286**, 132.

Kormendy J. (1985), in *IAU Symp.* **106**, Reidel, Dordrecht, van Woerden H. et al. eds., p. 541 (Groningen, The Netherlands, 1983).

Leisawitz D. (1990), *Ap. J.* **359**, 319.

Lewis J. R. and Freeman K. C. (1989), *A. J.* **97**, 139.

Li X.-Q. (1990), *A. Ap.* **227**, 317.

Lin C. C. and Bertin G. (1985), in *IAU Symp.* **106**, Reidel, Dordrecht, van Woerden H. et al. eds., p. 513 (Groningen, The Netherlands, 1983).

Lin C. C. and Shu F. H. (1966), *Proc. Natl. Acad. Sci.* **55**, 229.

Lowe S. A. (1988), *Ph. D. Thesis*, MIT, Cambridge, Massachusetts USA.

Lubow S. H. (1986), *Ap. J. Lett.* **307**, L39.

Lubow S. H., Balbus S. A. and Cowie L. L. (1986), *Ap. J.* **309**, 496.

Lynden-Bell D. (1967), in *Relativity Theory and Astrophysics*, Vol. 2, Am. Math. Soc., Providence, Ehlers J. ed., p. 131 (Ithaca, New York USA, 1965).

MacLaren I., Richardson K. M. and Wolfendale A. W. (1988), *Ap. J.* **333**, 821.

Maloney P. (1988), *Ap. J.* **334**, 761.

Maloney P. (1990), *Ap. J. Lett.* **348**, L9.

Maloney P. and Black J. H. (1988), *Ap. J.* **325**, 389.

Marochnik L. S. (1970), in *IAU Symp.* **38**, Reidel, Dordrecht, Becker W. and Contopoulos G. eds., p. 317 (Basel, Switzerland, 1969).

Marochnik L. S. and Suchkov A. A. (1969), *Ap. Space Sci.* **4**, 317.

Martin C., Hurwitz M. and Bowyer S. (1990), *Ap. J.* **354**, 220.

Martinet L. (1988), *A. Ap.* **206**, 253.

Miller R. H., Prendergast K. H. and Quirk W. J. (1970), *Ap. J.* **161**, 903.

Min K. W. (1988), *Ap. Space Sci.* **145**, 167.

Morozov A. G. (1981), *Sov. Astron. Lett.* **7**, 5.

Myers P. C. and Goodman A. A. (1988a), *Ap. J. Lett.* **326**, L27.

Myers P. C. and Goodman A. A. (1988b), *Ap. J.* **329**, 392.

Ostriker J. P. (1985), in *IAU Symp.* **106**, Reidel, Dordrecht, van Woerden H. et al. eds., p. 638 (Groningen, The Netherlands, 1983).

Polk K. S., Knapp G. R., Stark A. A. and Wilson R. W. (1988), *Ap. J.* **332**, 432.

Quirk W. J. (1971), *Ap. J.* **167**, 7.

Renz W. (1985), in *IAU Symp.* **106**, Reidel, Dordrecht, van Woerden H. et al. eds., p. 497 (Groningen, The Netherlands, 1983).

Richardson K. M. and Wolfendale A. W. (1988a), *A. Ap.* **201**, 100.

Richardson K. M. and Wolfendale A. W. (1988b), *A. Ap.* **203**, 289.

Romeo A. B. (1985), *Tesi di Laurea*, University of Pisa and Scuola Normale Superiore, Pisa, Italy.

Romeo A. B. (1987), *M. Phil. Thesis*, SISSA, Trieste, Italy.

Romeo A. B. (1988), *Preprint, SISSA Astro. 167*, Trieste, Italy (enlarged version of Romeo 1989).

Romeo A. B. (1989), in *Dynamics of Astrophysical Discs*, Cambridge University Press, Cambridge, Sellwood J. A. ed., p. 209 (Manchester, UK, 1988).

Sage L. J., Shore S. N. and Solomon P. M. (1990), *Ap. J.* **351**, 422.

Sellwood J. A. (1985), *M.N.R.A.S.* **217**, 127.

Sellwood J. A. (1989), in *Dynamics of Astrophysical Discs*, Cambridge University Press, Cambridge, Sellwood J. A. ed., p. 155 (Manchester, UK, 1988).

Sellwood J. A. and Carlberg R. G. (1984), *Ap. J.* **282**, 61.

Shu F. H. (1985), in *IAU Symp.* **106**, Reidel, Dordrecht, van Woerden H. et al. eds., p. 530 (Groningen, The Netherlands, 1983).

Smith B. F. and Miller R. H. (1986), *Ap. J.* **309**, 522.

Sodroski T. J., Dwek E., Hauser M. G. and Kerr F. J. (1989), *Ap. J.* **336**, 762.

Solomon P. M., Rivolo A. R., Barrett J. and Yahil A. (1987), *Ap. J.* **319**, 730.

Somerville W. B. and Smith C. A. (1989), *M.N.R.A.S.* **238**, 559.

Stark A. A. (1984), *Ap. J.* **281**, 624.

Stark A. A. and Brand J. (1989), *Ap. J.* **339**, 763.

Stecker F. W. (1989), *Preprint, NASA-LHEA 41*, Greenbelt, Maryland USA, to appear in a Festschrift in honour of Shapiro M..

Strong A. W. et al. (1988), *A. Ap.* **207**, 1.

Sweet P. A. (1963), *M.N.R.A.S.* **125**, 285.

Sweet P. A. and McGregor D. D. (1964), *M.N.R.A.S.* **128**, 195.

Taylor D. K., Dickman R. L. and Scoville N. Z. (1987), *Ap. J.* **315**, 104.

Thomasson M., Elmegreen B. G., Donner K. J. and Sundelius B. (1990), *Ap. J. Lett.* **356**, L9.

van der Kruit P. C. and Freeman K. C. (1986), *Ap. J.* **303**, 556.

Vandervoort P. O. (1971), *Ap. J.* **166**, 37.

Weliachew L. N. and Lucas R. (1988), *Rep. Prog. Phys.* **51**, 605.

Zasov A. V. and Simakov S. G. (1988), *Astrophys.* **29**, 518.

Figure Captions

Figure 7.1. Two-fluid marginal stability curves in the $(\bar{\lambda}, Q_{\rm H}^2)$ plane for some values of the local parameter α and fixed $\beta = .01$. Note the quick appearance of the Gaseous peak occurring at short wavelengths $\bar{\lambda} \approx \frac{1}{2}\alpha$ with $Q_{\rm H}^2 \approx (\alpha^2/\beta) + 4\alpha$, and the smoother behaviour of the Stellar peak occurring at intermediate wavelengths $\bar{\lambda} \approx \frac{1}{2}$ with $Q_{\rm H}^2 \approx 1 + 4\alpha$.

Figure 7.2. Phase diagram summarizing the properties of the marginal stability curve. Inside the Two-Phase region of the (β, α) plane, which is the triangular region with vertex (β_0, α_0), the marginal stability curve exhibits two maxima, so that the local stability properties are qualitatively different from those of single-component systems. In the Stellar (Gaseous) regime the stellar (gaseous) peak is dominant. The curve $\alpha = \sqrt{\beta}$ corresponds to the Transition between these two regimes. A cusp formed by the intersection between the upper and the lower boundaries of the two-phase region and the

transition curve occurs at the "Triple point" $(\beta_0, \alpha_0) = (17-12\sqrt{2},\ 3-2\sqrt{2})$. Correspondingly, the marginal stability curve exhibits a single flat maximum at $\bar{\lambda}_0 = 1 - (1/\sqrt{2})$ with $\bar{Q}_0^2 = 2$. Four points are marked corresponding to the cases illustrated in Figure 7.1.

Figure 7.3. $\bar{Q}^2$-contours in the (β, α) plane, showing the transition inside the two-phase region (**a**) and the large-scale behaviour in the ranges $0 < \alpha < 1$, $0 < \beta < 1$ (**b**). Each contour is labeled with the appropriate value of $\bar{Q}^2$. Note the flat behaviour in Star-dominated regimes.

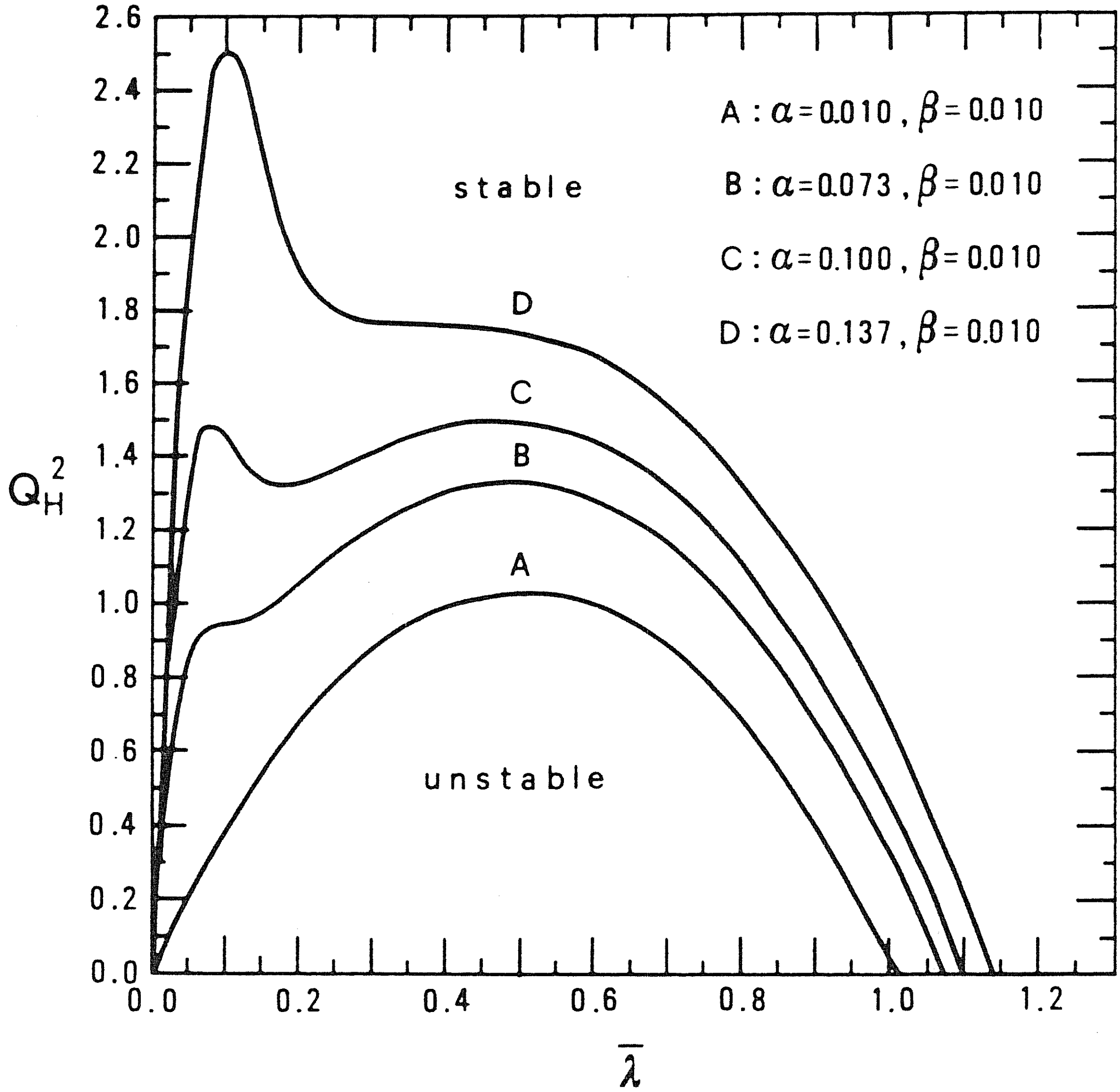
Figure 7.4. Rotation curve, disc density distribution and Q-profile for the one-component equilibrium model $E3_0$. The rotation curve is supported by the combined effect of an active disc and a spheroidal bulge-halo which does not participate in the spiral perturbations (the density distribution of the latter component is not shown here). The Q-profile is consistent with the presence of a bulge in the inner regions and with a mechanism of *self-regulation* in the active disc.

Figure 7.5. The (β, α) tracks for the galaxies NGC 4565, NGC 5907 and for the two-component equilibrium models E3a, E3b. The model E3a is characterized by a flat gas mass distribution $\sigma_C = 3 \cdot 10^6\ M_\odot\,\mathrm{kpc}^{-2}$, simulating the neutral atomic hydrogen distribution, and by a constant equivalent acoustic speed $c_C = 8\ \mathrm{km\,s}^{-1}$. The model E3b is characterized by a gaussian ring overlapped to a flat background $\sigma_C = \{4 + 6\exp[-4(\hat{r}-4)^2]\} \cdot 10^6\ M_\odot\,\mathrm{kpc}^{-2}$ with $\hat{r} \equiv r/1$ kpc, simulating the presence of a ring of molecular hydrogen, and by a constant equivalent acoustic speed $c_C = 6\ \mathrm{km\,s}^{-1}$.

Figure 7.6. The Local Stability parameter Q_H for the two-component equilibrium models E3a and E3b. These profiles have been derived by imposing the condition $Q_H(r) = Q(r)\,\bar{Q}[\alpha(r), \beta(r)]$ (Self-Regulation constraint), $Q(r)$ being specified in the basic one-component equilibrium model $E3_0$. Note that Q_H lies above unity everywhere.

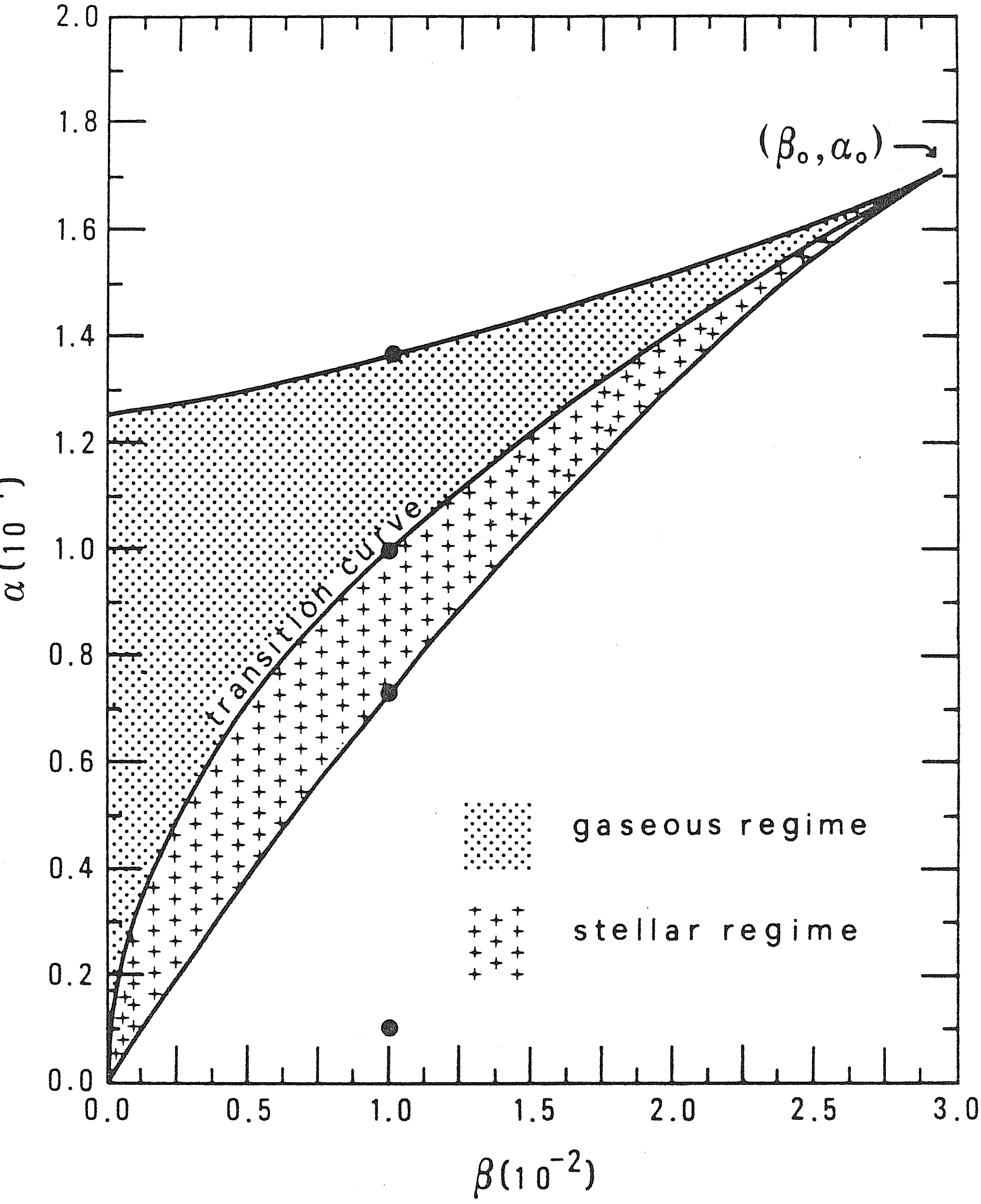
Figure 7.7. Propagation diagrams relative to the $(m = 2,\ n = 0)$ mode for the two-component equilibrium models E3a (**a**) and E3b (**b**), with the indication of the corresponding bulge radius r_{ce}, corotation circle r_{co}, location of the outer Lindblad resonance r_{OLR} and pattern frequency Ω_p. The propagation diagram in **a** essentially coincides with that obtained for the one-component equilibrium model $E3_0$. The distortion affecting appreciably the short-wave branch in **b** is due to the presence of the ring of molecular hydrogen. Thus, when gas is included, there is a shift towards shorter wavelengths and the size of the pattern, as measured by the radial range of the loop, shrinks.

Figure 7.1



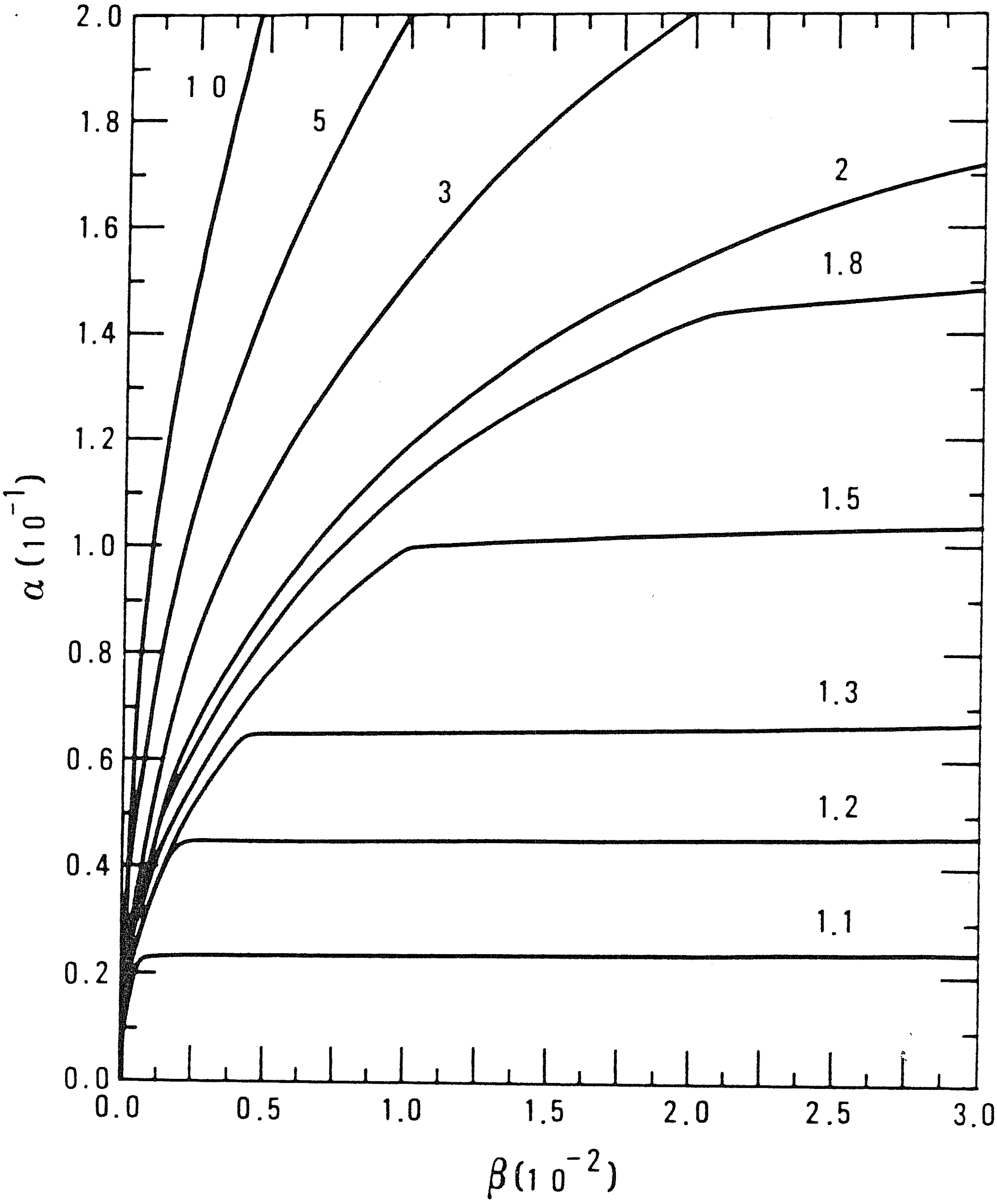
A : α=0.010 , β= 0.010
B : α=0.073 , β= 0.010
C : α=0.100 , β= 0.010
D : α=0.137 , β= 0.010
stable
unstable
D
C
B
A
Q_H^2
λ̄
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2

Figure 7.2



2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
$\alpha\,(10^{-1})$
$\beta\,(10^{-2})$
(β_o, α_o)
transition curve
gaseous regime
stellar regime

Figure 7.3a



2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
α (10⁻¹)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
β (10⁻²)
10
5
3
2
1.8
1.5
1.3
1.2
1.1

Figure 7.3b

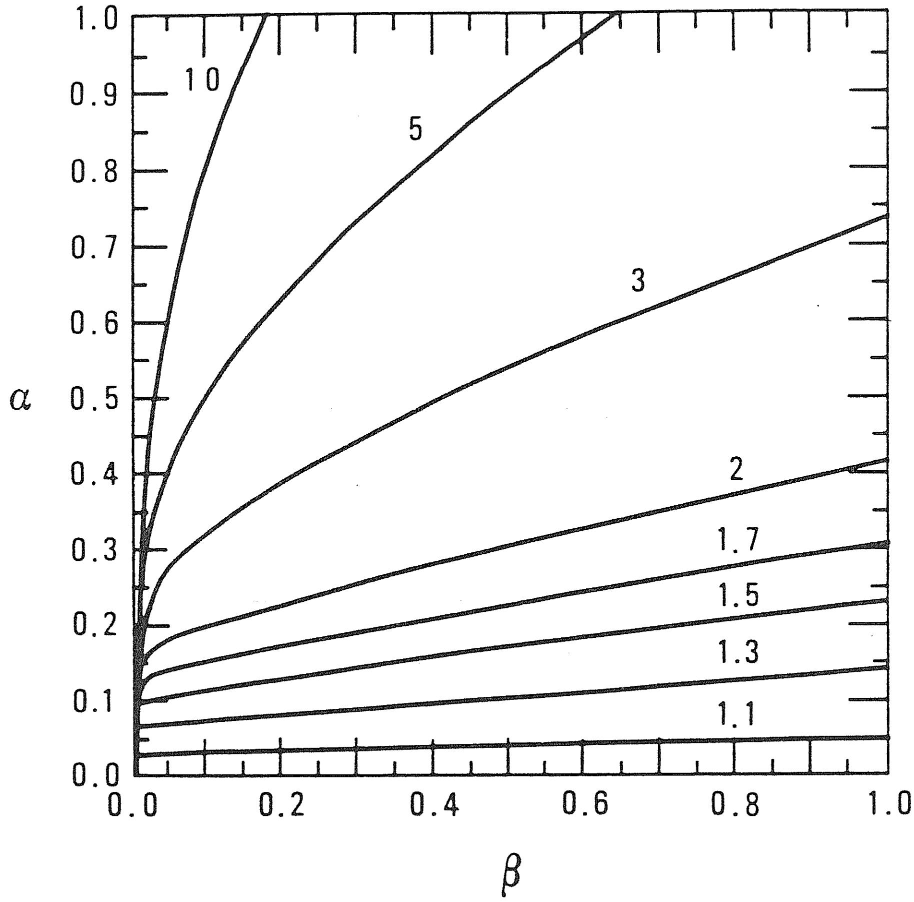
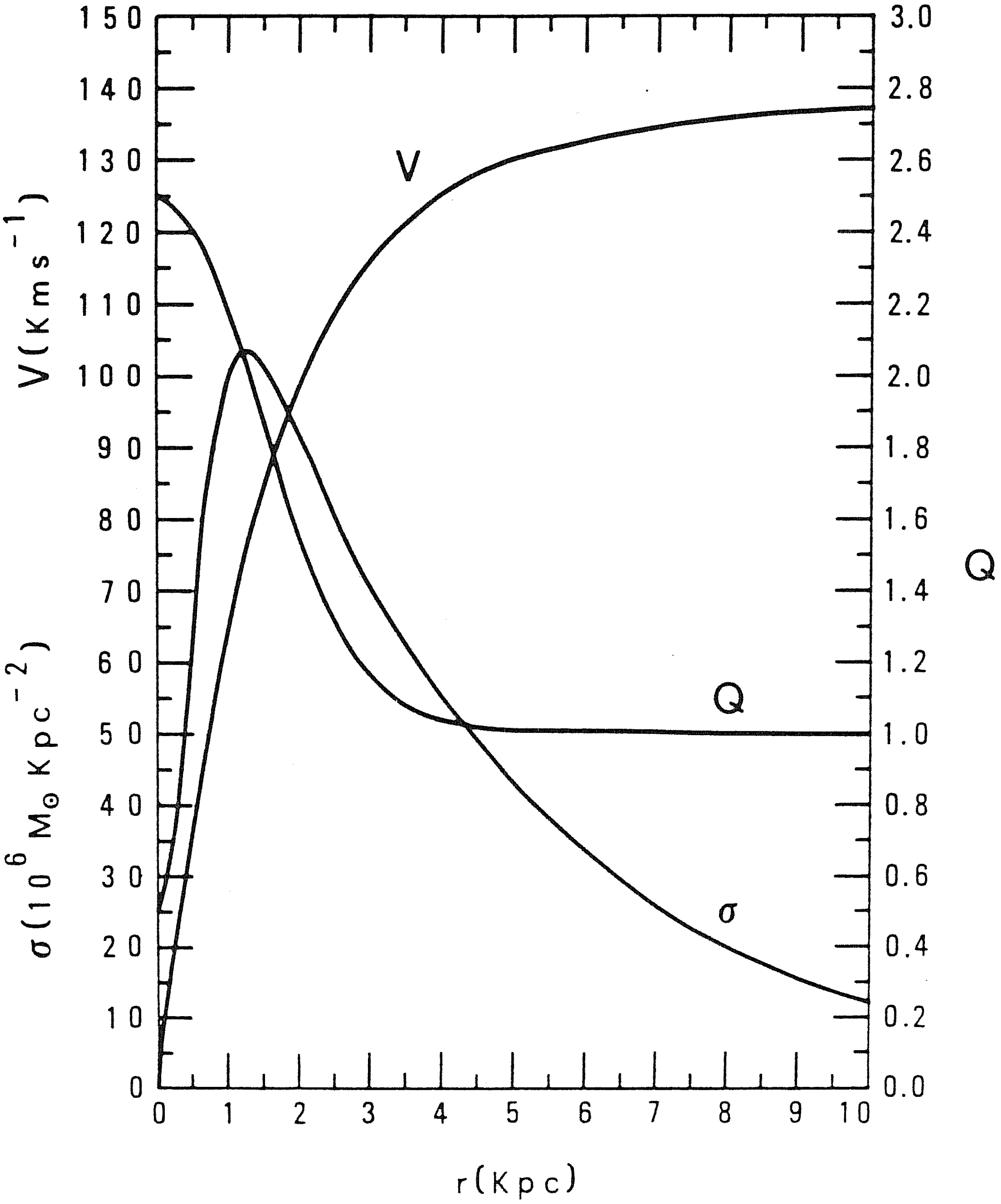
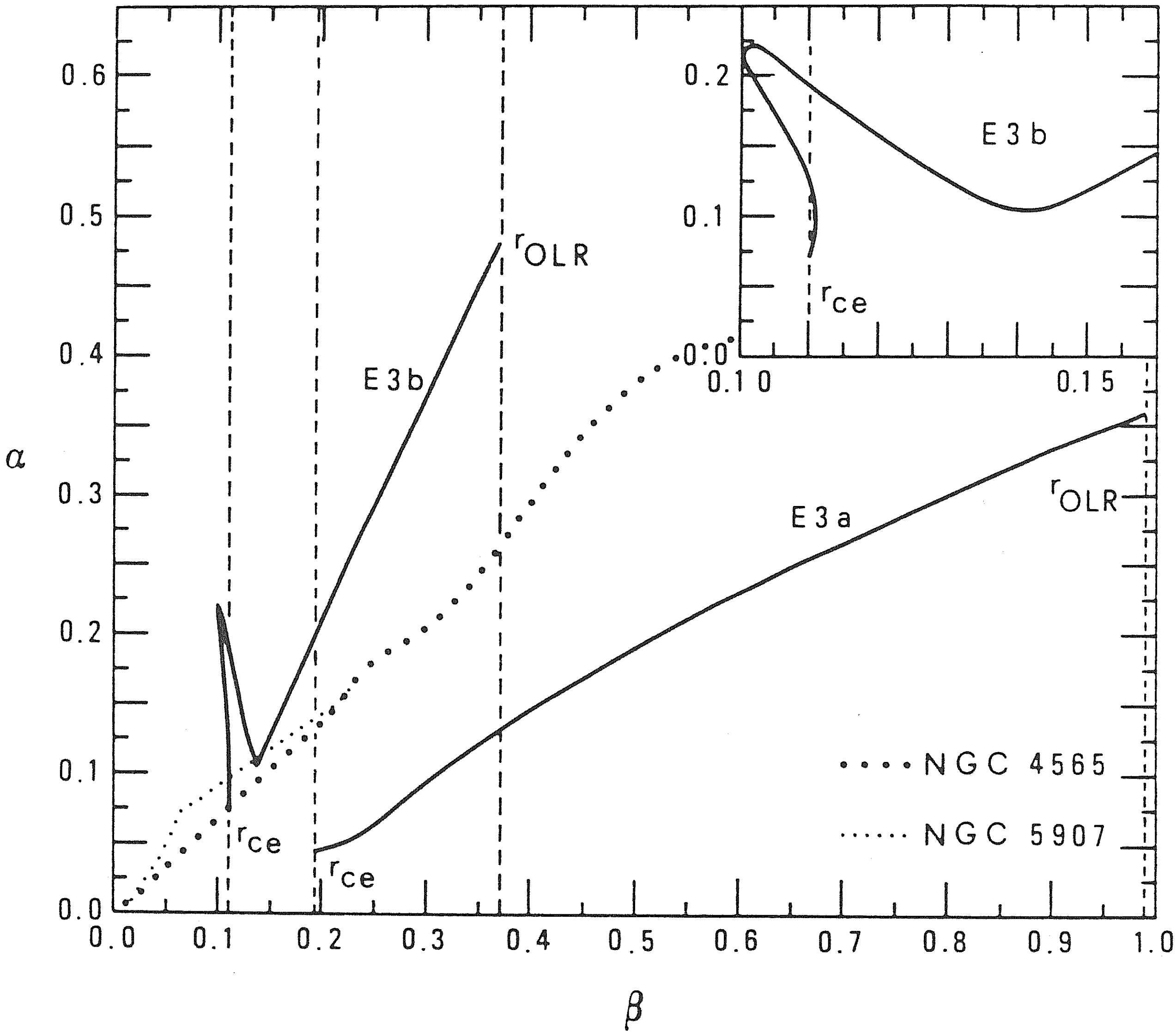


Figure 7.4



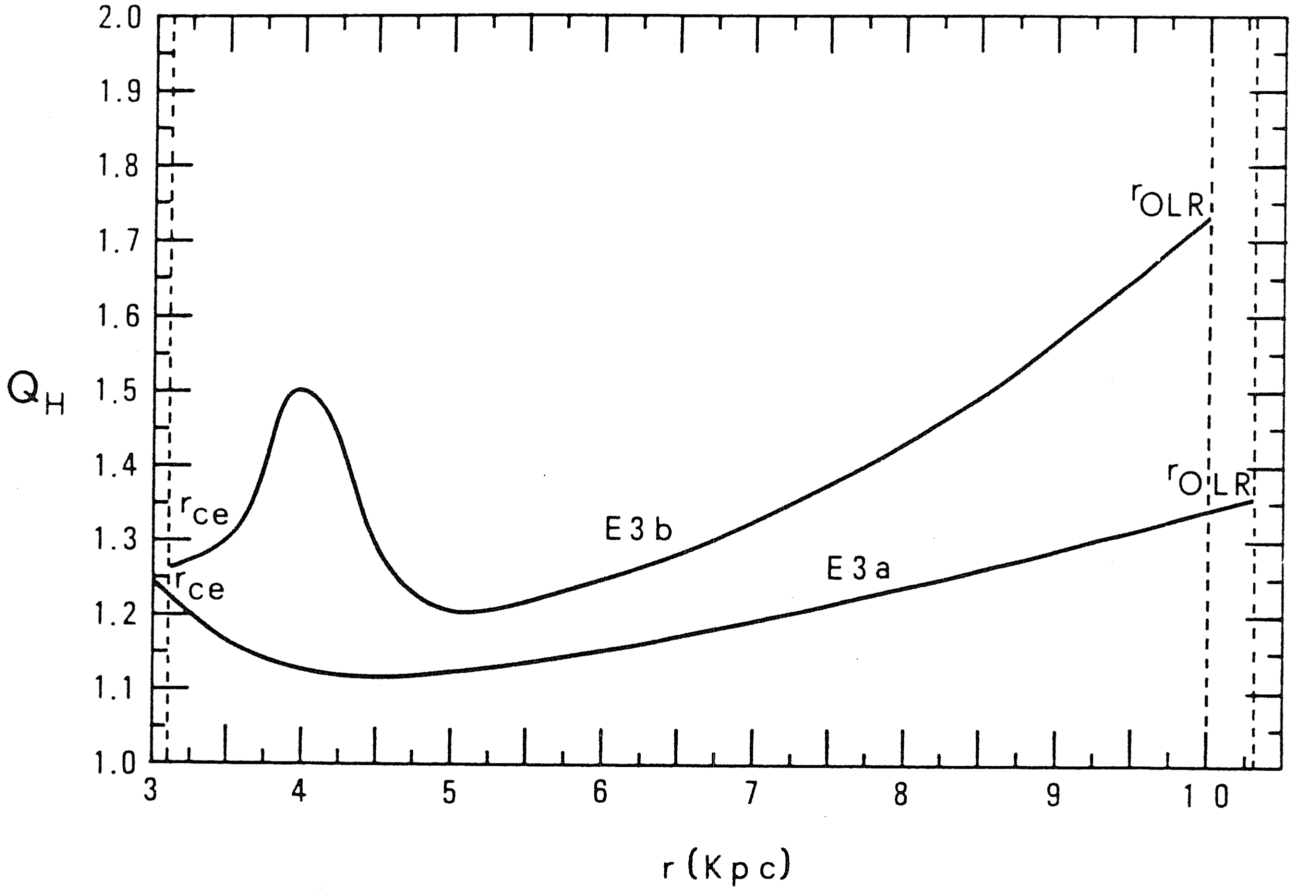
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
V
Q
σ
V(Kms⁻¹)
σ(10⁶ M⊙ Kpc⁻²)
Q
0
1
2
3
4
5
6
7
8
9
10
r(Kpc)

Figure 7.5



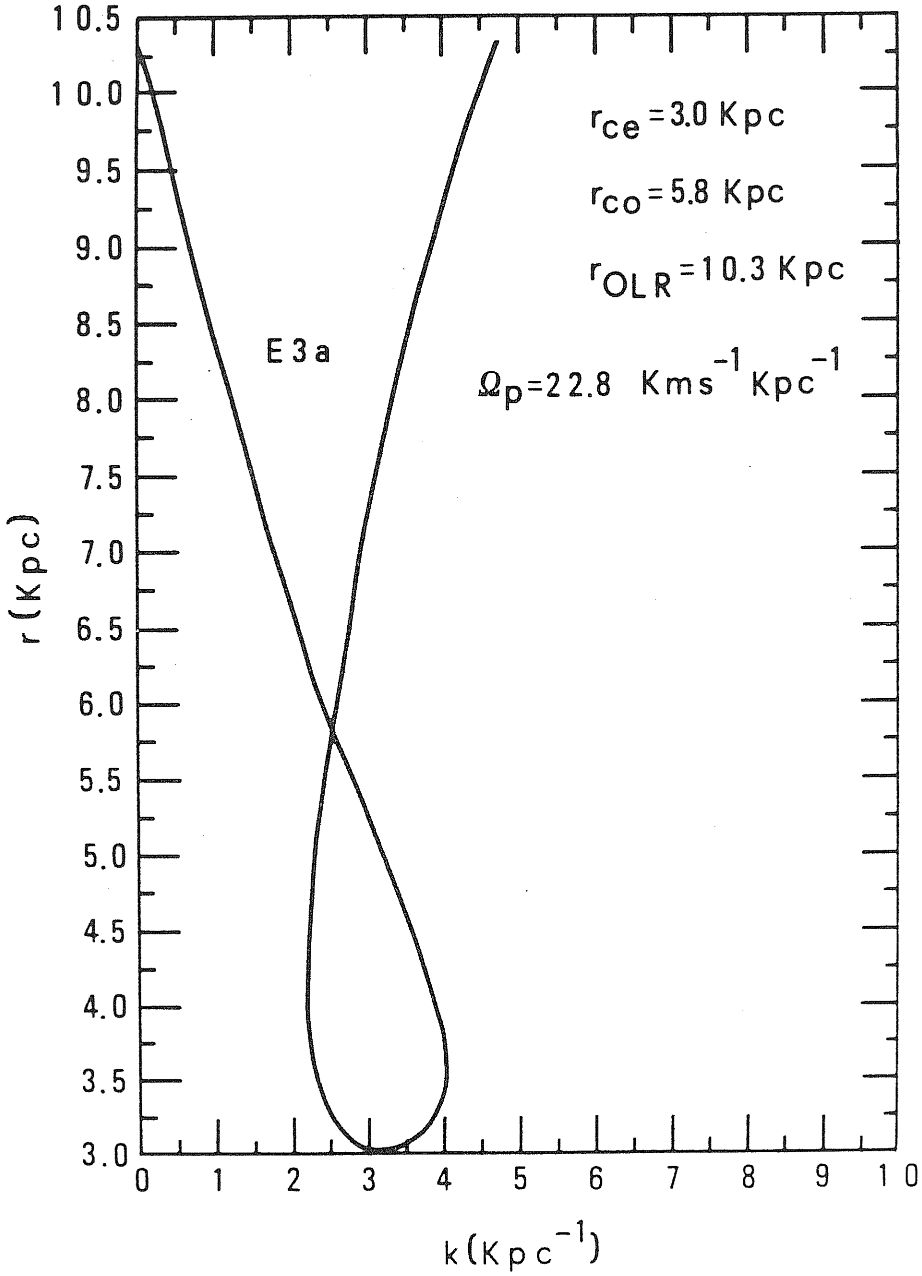
0.6
0.5
0.4
0.3
0.2
0.1
0.0
α
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
β
E3b
E3a
r_{OLR}
r_{ce}
NGC 4565
NGC 5907
0.10
0.15

Figure 7.6



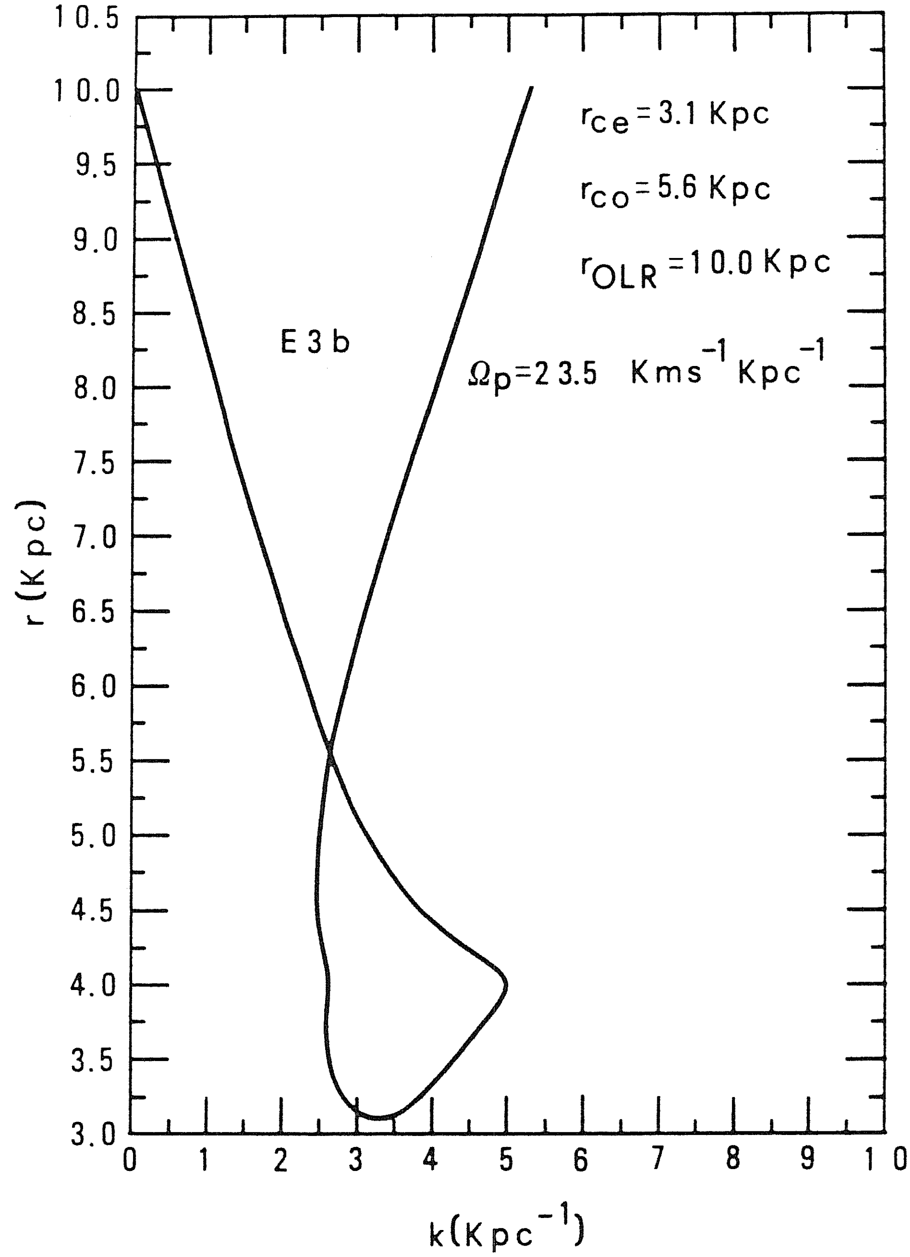
2.0
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
Q_H
r_ce
r_ce
E3b
E3a
r_OLR
r_OLR
3
4
5
6
7
8
9
10
r (Kpc)

Figure 7.7a



10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
r (Kpc)
0
1
2
3
4
5
6
7
8
9
10
k (Kpc^{-1})
E3a
r_{ce} = 3.0 Kpc
r_{co} = 5.8 Kpc
r_{OLR} = 10.3 Kpc
Ω_p=22.8 Kms^{-1} Kpc^{-1}

Figure 7.7b



10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
r (Kpc)
0
1
2
3
4
5
6
7
8
9
10
k(Kpc⁻¹)
E3b
rce = 3.1 Kpc
rco = 5.6 Kpc
rOLR = 10.0 Kpc
Ωp = 23.5 Kms⁻¹ Kpc⁻¹

Chapter 8

The Cold Interstellar Gas and the Stability of Galactic Discs: Vertical Structure at Equilibrium

Summary

The crucial role that the cold interstellar gas can play in the dynamics and structure of early normal spiral galaxies has been shown in Chapter 7, where finite-thickness effects have not been taken into account. In view of the importance that such effects might have in the self-regulation mechanisms which are expected to operate in galactic discs and to be at the basis of their secular heating, we have tried to evaluate them. This can be done only after that their vertical structure at equilibrium has carefully been investigated.

An asymptotic analysis has thus been carried out to study the thickness-scales relevant to both the equilibrium and stability of two-component galactic discs in regimes of astrophysical interest. Two parametrizations have been introduced and examined in view of their relevance to the stability analysis which we shall perform in Chapter 9.

8.1 Introduction

The spiral structure theory relies on a number of working assumptions which allow to make the linearized system of the coupled fluid (or Vlasov) and Poisson equations more tractable. However, in some situations of astrophysical interest the validity of such assumptions may be questioned.

For instance, we know that real galactic discs have finite, although small, thickness and the possibility of regarding them as infinitesimally thin depends on the relevant wavelengths of the perturbations excited. In some cases, when the underlying spiral structure has a high winding degree, finite-thickness effects should be taken into account in the stability analysis. In view of the importance that such effects may have in the *self-regulation* mechanisms which are expected to operate in galactic discs and to be at the basis of their secular heating, we have tried to evaluate them. This can be done only after that their vertical structure at equilibrium has *carefully* been investigated.

In performing this analysis, we have made use of simplified models of galactic discs in which the stars and the cold interstellar gas are treated as two different components (see Chapter 7). Although such models might be thought of as being inaccurate to describe actual galactic discs, which are known to consist of different populations of stars and gas components, they incorporate indeed the most essential features as regard their stability properties. In this context it should be noted that such a single equivalent stellar component is taken to be representative of the whole active stellar disc consisting of low-velocity dispersion stars (high-velocity dispersion stars do not participate appreciably in spiral structure), whereas the single gaseous component is taken to simulate HI regions of neutral atomic hydrogen and giant molecular clouds and complexes.

Generally, even more drastically simplified galactic models are used, in which the cold interstellar gas is not taken into account. In some situations of astrophysical interest this further simplification is not justified, because the cold interstellar gas is expected to play an important or even crucial role in the stability of galactic discs due to its low turbulent velocities (see Chapter 7).

The vertical structure of galactic discs has long been investigated in detail by a number of authors, who made use of multi-component locally isothermal models (e.g., Woolley 1957; Bahcall 1984; Bahcall and Casertano 1984; see Boulares 1989 for a discussion of non-thermal effects). These analyses take into account the fact that galactic discs are nearly self-gravitating perpendicularly to their symmetry plane, so that standard asymptotic expansion techniques can be employed. While in the case of one-component stellar discs this is all that is needed to make the problem analytically tractable (see Vandervoort 1967, 1970 for the most rigorous analysis in this context), when multi-component models are considered, further assumptions are to be made to this end: generally, the component with the largest scale-height is taken to have the largest mass density, so that a perturbative approach can be employed.

In our two-component model this assumption is certainly satisfied, but a perturbative approach of this kind may not always be suitable because in the outermost regions of galactic discs the mass density of the cold interstellar gas becomes comparable to that of disc stars. For this reason we have relaxed this assumption, and this has allowed us to perform a detailed analysis in regimes of astrophysical interest only at small and large distances from the galactic plane (Romeo 1987, 1990a,b). Anyway, as will be shown in Chapter 9, these distances are those most relevant to the stability analysis, and also from an observational point of view. The results obtained in this chapter are at the basis of the further investigation which we shall carry out in Chapter 9.

8.2 One-Component Case

We shall now briefly discuss the one-component case because the investigation of two-component galactic discs, although it is much more complicated, employs similar methods.

The system, which is assumed to be in an axi-symmetric and plane-symmetric equilibrium state and to be locally isothermal perpendicularly to the galactic plane, is described by the Poisson equation

$$\frac{\partial^2 \Phi}{\partial r^2} + \frac{1}{r}\frac{\partial \Phi}{\partial r} + \frac{\partial^2 \Phi}{\partial z^2} = 4\pi G\rho \tag{8.1}$$

supplemented by the "locally isothermal" condition

$$\rho(r,z) = \rho_0(r)\,\exp\left\{-\frac{[\Phi(r,z)-\Phi(r,0)]}{c_z^2(r)}\right\}\,. \tag{8.2}$$

If $\langle r\rangle$ and $\langle z\rangle$ are the characteristic radius and the thickness-scale of the system (defined to be positive) respectively, then $(\partial^2\Phi/\partial r^2)+r^{-1}(\partial\Phi/\partial r) = O(\Phi/\langle r\rangle^2)$ and $(\partial^2\Phi/\partial z^2) = O(\Phi/\langle z\rangle^2)$. Therefore, taking into account the fact that galactic discs are highly flattened, we can perform an asymptotic expansion in powers of the small parameter $\epsilon \equiv \langle z\rangle/\langle r\rangle \ll 1$. We obtain the following hierarchy of Poisson equations (Vandervoort 1967, 1970):

$$\frac{\partial^2\Phi^{(0)}}{\partial z^2} = 0\,, \tag{8.3}$$

$$\frac{\partial^2\Phi^{(1)}}{\partial z^2} = 4\pi G\rho^{(0)}\,, \tag{8.4}$$

$$\frac{\partial^2\Phi^{(n-2)}}{\partial r^2} + \frac{1}{r}\frac{\partial\Phi^{(n-2)}}{\partial r} + \frac{\partial^2\Phi^{(n)}}{\partial z^2} = 4\pi G\rho^{(n-1)} \quad (n\geq 2)\,, \tag{8.5}$$

where in these notations $\Phi^{(n)}, \rho^{(n)} = O(\epsilon^n)$. In the following we shall consider only non-trivial lowest-order contributions, which are represented by equation (8.4).

Combining this equation with the "locally isothermal" condition (8.2) and suppressing the order-indices for simplicity of notations, we find in dimensionless form:

$$\frac{d^2\hat{\Phi}}{d\hat{z}^2} = 2\,e^{-\hat{\Phi}}\,, \tag{8.6}$$

where we have adopted the following scaling:

$$\hat{z} \equiv \frac{z}{\Delta}\,, \quad \text{where} \quad \Delta \equiv \sqrt{\frac{c_z^2}{2\pi G\rho_0}}\,; \tag{8.7}$$

$$\hat{\Phi} \equiv \frac{\Phi - \Phi_0}{c_z^2}\,, \quad \text{where} \quad \Phi_0 \equiv \Phi(r, z = 0)\,. \tag{8.8}$$

This non-linear second-order differential equation can easily be integrated by quadratures (see, e.g., Bahcall 1984). Imposing the boundary conditions

$$\hat{\Phi}(\hat{z} = 0) = 0\,, \quad \left.\frac{d\hat{\Phi}}{d\hat{z}}\right|_{\hat{z}=0} = 0\,, \tag{8.9}$$

we find:

$$\int_0^{\hat{\Phi}} \frac{d\hat{\Phi}'}{\sqrt{1 - e^{-\hat{\Phi}'}}} = 2|\hat{z}|\,, \tag{8.10}$$

whose solution is

$$\hat{\Phi}(\hat{z}) = \ln(\cosh^2 \hat{z})\,. \tag{8.11}$$

The corresponding volume density is

$$\rho(r, z) = \rho_0(r)\,\mathrm{sech}^2\left[\frac{z}{\Delta(r)}\right]\,. \tag{8.12}$$

The following two asymptotic limits are of interest:

$$\rho \underset{|z|\to 0}{\simeq} \rho_0 \exp\left(-\frac{z^2}{z_G^2}\right)\,, \quad \text{where} \quad z_G = \Delta\,; \tag{8.13}$$

$$\rho \underset{|z|\to +\infty}{\simeq} 4\rho_0 \exp\left(-\frac{|z|}{z_E}\right)\,, \quad \text{where} \quad z_E = \frac{1}{2}\,\Delta\,; \tag{8.14}$$

which define the gaussian and the exponential thickness-scales, respectively. Integrating the volume density given in equation (8.12) over z, we find the following expression for the surface density:

$$\sigma = \rho_0\,(2\Delta)\,, \tag{8.15}$$

so that

$$\Delta = \frac{c_z^2}{\pi G \sigma}, \tag{8.16}$$

and in this one-component case the expansion thickness-scale is

$$\langle z \rangle \equiv \frac{\sigma}{2\rho_0} = \Delta \,. \tag{8.17}$$

8.3 Two-Component Case

8.3.1 Formal Integration of the Fundamental Equations

While the investigation of the one-component case is more or less straightforward to the non-trivial lowest order of approximation, the same is not true for two-component galactic discs: several complications arise, some of which are already hidden in the one-component case.

The two components are denoted by different subscripts, "H" (HOT) and "C" (COLD), in order to recall that they are characterized by different vertical velocity dispersions. Having in mind cases of astrophysical interest, we refer to them as the stars of the active disc (H) and the cold interstellar gas (C). However, we could also consider the case in which gas is absent but two stellar populations with different scale-heights can be identified. It is assumed that the system possesses the same symmetry properties as in the one-component case, and that each component is locally isothermal perpendicularly to the galactic plane. The only interaction between the two components is taken to occur via the gravitational field (Poisson equation)

$$\frac{\partial^2 \Phi}{\partial r^2} + \frac{1}{r}\frac{\partial \Phi}{\partial r} + \frac{\partial^2 \Phi}{\partial z^2} = 4\pi G \left(\rho_{\rm H} + \rho_{\rm C}\right) \tag{8.18}$$

supplemented by the "locally isothermal" condition

$$\rho_{\rm i}(r,z) = \rho_{0\rm i}(r)\, \exp\left\{ -\frac{[\Phi(r,z) - \Phi(r,0)]}{c_{z\rm i}^2(r)} \right\} \quad ({\rm i} = {\rm H}, {\rm C})\,. \tag{8.19}$$

Proceeding along the same line as in the one-component case, we find that equation (8.6) is replaced by

$$\frac{d^2\hat{\Phi}}{d\hat{z}^2} = 2\left(e^{-\hat{\Phi}} + \gamma e^{-\beta_z^{-1}\hat{\Phi}}\right), \tag{8.20}$$

where we have adopted the following scaling and parametrization:

$$\hat{z} \equiv \frac{z}{\Delta_{\rm H}}, \quad \text{where} \quad \Delta_{\rm H} \equiv \sqrt{\frac{c_{z\rm H}^2}{2\pi G \rho_{0\rm H}}}\,; \tag{8.21}$$

$$\hat{\Phi} \equiv \frac{\Phi - \Phi_0}{c_{zH}^2}, \quad \text{where} \quad \Phi_0 \equiv \Phi(r, z = 0); \tag{8.22}$$

$$\gamma \equiv \frac{\rho_{0C}}{\rho_{0H}}, \quad \beta_z \equiv \frac{c_{zC}^2}{c_{zH}^2} \quad (0 < \gamma < +\infty, \ 0 < \beta_z < 1). \tag{8.23}$$

The cases $\gamma = 0$, $\gamma \to +\infty$ represent the limit of a one-component system, and the case $\beta_z = 1$ represents the limit of a system in which the two components have the same scale-height. Equation (8.20) can formally be integrated by quadratures imposing the same boundary conditions as in the one-component case:

$$F[\hat{\Phi}] \equiv \int_0^{\hat{\Phi}} \frac{d\hat{\Phi}'}{\sqrt{(1 - e^{-\hat{\Phi}'}) + \gamma\beta_z (1 - e^{-\beta_z^{-1}\hat{\Phi}'})}} = 2|\hat{z}|. \tag{8.24}$$

However, in contrast to the one-component case, the left-hand side of this equation cannot explicitly be expressed in terms of elementary or special functions for arbitrary values of the local parameter β_z (see, e.g., Gradshteyn and Ryzhik 1980; Prudnikov, Brychkov and Marichev 1986).

In order to overcome this difficulty two strategies can be employed:

- One consists in assuming $\gamma\beta_z \ll 1$ and in carrying out a perturbation series expansion in powers of this small quantity (perturbative approach).
- The other consists in investigating the asymptotic behaviour of $F[\hat{\Phi}]$ as $\hat{\Phi} \to 0$ and $\hat{\Phi} \to +\infty$ so as to estimate $\hat{\Phi}(\hat{z})$ at small and large $|\hat{z}|$ (asymptotic approach).

Both methods will be considered and the corresponding results will be compared in such a way as to draw as much information as possible.

8.3.2 Perturbative Approach

Let us first consider the perturbative approach developed by Bahcall (1984).

Performing an asymptotic expansion in powers of the small local parameter $\gamma\beta_z \ll 1$ and retaining only first-order terms, we obtain:

$$\hat{\Phi} \simeq \hat{\Phi}^{(0)} + \gamma\beta_z\hat{\Phi}^{(1)}, \quad \gamma\beta_z\hat{\Phi}^{(1)} \ll \hat{\Phi}^{(0)}; \tag{8.25}$$

$$\hat{\Phi}^{(0)}(\hat{z}) = \ln(\cosh^2 \hat{z}); \tag{8.26}$$

$$\hat{\Phi}^{(1)}(\hat{z}) = I(\hat{z}; \beta_z) \tanh \hat{z}, \quad \text{where} \quad I(\hat{z}; \beta_z) \equiv \int_0^{\hat{z}} \frac{1 - \operatorname{sech}^{2\beta_z^{-1}} \hat{z}}{\tanh^2 \hat{z}} \, d\hat{z}. \tag{8.27}$$

$I(\hat{z};\beta_z)$ cannot explicitly be expressed in terms of elementary or special functions for arbitrary values of the local parameter β_z, so that only its asymptotic behaviour can be investigated analytically. We shall not pursue this line, because this will be done even without the restriction $\gamma\beta_z \ll 1$ in the asymptotic approach which we shall discuss later on. Rather, we shall consider the case in which first-order contributions can be neglected, so that only $\hat{\Phi}^{(0)}$ is relevant to the following analysis.

The volume densities of the two components are:

$$\rho_{\rm H}(r,z) \underset{\substack{\gamma\beta_z \ll 1 \\ (\gamma \ll 1)}}{\simeq} \rho_{0\rm H}(r)\,\mathrm{sech}^2\left[\frac{z}{\Delta_{\rm H}(r)}\right] , \tag{8.28}$$

$$\rho_{\rm C}(r,z) \underset{\substack{\gamma\beta_z \ll 1 \\ (\gamma \ll 1)}}{\simeq} \rho_{0\rm C}(r)\,\mathrm{sech}^{2\beta_z^{-1}}\left[\frac{z}{\Delta_{\rm H}(r)}\right] . \tag{8.29}$$

The further restriction $\gamma \ll 1$ has been derived by comparing the gaussian and the exponential thickness-scales of each component found in this lowest-order approximation to those determined exactly in the asymptotic approach [see equations (8.32) and (8.33) of next subsection]. It can be identified with the self-consistency condition of this perturbative approach.

For an interesting discussion of the observational implications of this analysis see Bahcall (1984) and Bahcall and Casertano (1984).

8.3.3 Asymptotic Approach

Let us now consider the asymptotic approach developed by Romeo (1987, 1990a,b), which turns out to be richer in information than the perturbative approach considered above except in the asymptotic regime $\gamma \ll 1$.

Expanding the integrand function of the implicit equation (8.24) in powers of $\hat{\Phi}$ as $\hat{\Phi} \to 0$ and retaining only first-order terms, we obtain:

$$F[\hat{\Phi}] \underset{\hat{\Phi}\to 0}{\simeq} \frac{2\sqrt{\hat{\Phi}}}{\sqrt{1+\gamma}} \quad \Longrightarrow \quad \hat{\Phi}(\hat{z}) \underset{|\hat{z}|\to 0}{\simeq} (1+\gamma)\,\hat{z}^2 . \tag{8.30}$$

Studying the behaviour of the integrand function in the range $0 < \hat{\Phi} < +\infty$, it can be shown that the following reasonable lower estimate can be given for $F[\hat{\Phi}]$ as $\hat{\Phi} \to +\infty$:

$$F[\hat{\Phi}] \underset{\hat{\Phi}\to +\infty}{\simeq} \frac{\hat{\Phi}}{\sqrt{1+\gamma\beta_z}} \quad \Longrightarrow \quad \hat{\Phi}(\hat{z}) \underset{|\hat{z}|\to +\infty}{\simeq} 2\sqrt{1+\gamma\beta_z}\,|\hat{z}| . \tag{8.31}$$

The volume densities of the two-components have the following asymptotic behaviours:

$$\rho_{\mathrm{i}} \underset{|z|\to 0}{\simeq} \rho_{0\mathrm{i}} \exp\left(-\frac{z^2}{z_{\mathrm{Gi}}^2}\right) , \quad \text{where} \quad z_{\mathrm{Gi}} = \sqrt{\frac{c_{z\mathrm{i}}^2}{2\pi G\left(\rho_{0\mathrm{H}} + \rho_{0\mathrm{C}}\right)}} ; \tag{8.32}$$

$$\rho_{\mathrm{i}} \underset{|z|\to +\infty}{\simeq} 4^{K_{\mathrm{i}}} \rho_{0\mathrm{i}} \exp\left(-\frac{|z|}{z_{\mathrm{Ei}}}\right) , \quad \text{where} \quad z_{\mathrm{Ei}} = \sqrt{\frac{c_{z\mathrm{i}}^4}{8\pi G\left(\rho_{0\mathrm{H}} c_{z\mathrm{H}}^2 + \rho_{0\mathrm{C}} c_{z\mathrm{C}}^2\right)}} ; \tag{8.33}$$

$$\text{and hereafter} \quad (\mathrm{i} = \mathrm{H}, \mathrm{C}) ; \tag{8.34}$$

which define the gaussian and the exponential thickness-scales, respectively. The z-independent exponents K_{i} depend on higher-order corrections, and can only formally be expressed as series in terms of the local parameters γ and β_z:

$$\begin{aligned}
K_{\mathrm{H}} &= \sqrt{1+\gamma\beta_z} \int_0^{+\infty} \left[\frac{1}{\sqrt{(1-e^{-\hat{\Phi}}) + \gamma\beta_z\,(1-e^{-\beta_z^{-1}\hat{\Phi}})}} \right. \\
&\quad \left. - \lim_{\hat{\Phi}\to+\infty} \frac{1}{\sqrt{(1-e^{-\hat{\Phi}}) + \gamma\beta_z\,(1-e^{-\beta_z^{-1}\hat{\Phi}})}} \right] d\hat{\Phi}\, \log_4 e \\
&= \sum_{n=0}^{+\infty} \frac{(2n+1)!!}{(2n+2)!!} \frac{1}{(1+\gamma\beta_z)^{n+1}} \\
&\quad \cdot \sum_{k=0}^{n+1} \binom{n+1}{k} (\gamma\beta_z)^k \frac{\beta_z}{k+\beta_z\,(n+1-k)} \log_4 e \\
&\underset{\gamma\beta_z<1}{=} 1 + \sum_{n=0}^{+\infty} \frac{(2n+1)!!}{(2n+2)!!} \underbrace{\sum_{l=0}^{+\infty}\sum_{k=0}^{n+1}}_{l+k>0} (-1)^l \binom{n+l}{l} \binom{n+1}{k} \\
&\quad \cdot (\gamma\beta_z)^{l+k} \frac{\beta_z}{k+\beta_z\,(n+1-k)} \log_4 e ,
\end{aligned} \tag{8.35}$$

$$\begin{aligned}
K_{\mathrm{C}} &= \frac{\sqrt{1+\gamma\beta_z}}{\beta_z} \int_0^{+\infty} \left[\frac{1}{\sqrt{(1-e^{-\hat{\Phi}}) + \gamma\beta_z\,(1-e^{-\beta_z^{-1}\hat{\Phi}})}} \right. \\
&\quad \left. - \lim_{\hat{\Phi}\to+\infty} \frac{1}{\sqrt{(1-e^{-\hat{\Phi}}) + \gamma\beta_z\,(1-e^{-\beta_z^{-1}\hat{\Phi}})}} \right] d\hat{\Phi}\, \log_4 e \\
&= \sum_{n=0}^{+\infty} \frac{(2n+1)!!}{(2n+2)!!} \frac{1}{(1+\gamma\beta_z)^{n+1}} \\
&\quad \cdot \sum_{k=0}^{n+1} \binom{n+1}{k} (\gamma\beta_z)^k \frac{1}{k+\beta_z\,(n+1-k)} \log_4 e
\end{aligned}$$

$$\underset{\gamma\beta_z<1}{=} \frac{1}{\beta_z} + \sum_{n=0}^{+\infty} \frac{(2n+1)!!}{(2n+2)!!} \underbrace{\sum_{l=0}^{+\infty}\sum_{k=0}^{n+1}}_{l+k>0} (-1)^l \binom{n+l}{l} \binom{n+1}{k}$$

$$\cdot (\gamma\beta_z)^{l+k} \frac{1}{k+\beta_z(n+1-k)} \log_4 e \,. \tag{8.36}$$

The exponential thickness-scales of the two components can be expressed in a more compact and physically transparent form, by integrating the non-trivial lowest-order Poisson equation [the two-component analogue of equation (8.4)] over z

$$\left.\frac{\partial \Phi}{\partial z}\right|_{z\to+\infty} = 2\pi G(\sigma_{\rm H}+\sigma_{\rm C}) \tag{8.37}$$

and making use of the asymptotic form of the "locally isothermal" condition (8.19)

$$z_{\rm Ei}^{-1} \equiv -\left.\frac{\partial \ln \rho_{\rm i}}{\partial z}\right|_{z\to+\infty} = \frac{1}{c_{\rm zi}^2}\left.\frac{\partial \Phi}{\partial z}\right|_{z\to+\infty} . \tag{8.38}$$

The meaning of equations (8.32) and (8.37), (8.38) is clear: the total volume density in the plane determines the Gaussian thickness-scales of the two components

$$z_{\rm Gi} = \sqrt{\frac{c_{\rm zi}^2}{2\pi G\rho_0}}\,, \quad \frac{z_{\rm GC}}{z_{\rm GH}} = \sqrt{\beta_z}\,, \quad \text{where} \quad \rho_0 \equiv \rho_{\rm 0H}+\rho_{\rm 0C}\,, \tag{8.39}$$

while the total surface density determines their Exponential thickness-scales

$$z_{\rm Ei} = \frac{c_{\rm zi}^2}{2\pi G\sigma}\,, \quad \frac{z_{\rm EC}}{z_{\rm EH}} = \beta_z\,, \quad \text{where} \quad \sigma \equiv \sigma_{\rm H}+\sigma_{\rm C}\,. \tag{8.40}$$

The Global Effective thickness-scale of the system, which can be identified with the Expansion thickness-scale $\langle z\rangle$, can easily be expressed in terms of the asymptotic thickness-scales of the two components:

$$\langle z\rangle \equiv \frac{\sigma}{2\rho_0} = \frac{z_{\rm Gi}^2}{2z_{\rm Ei}} = \frac{z_{\rm GH}z_{\rm GC}}{\sqrt{2z_{\rm EH}\cdot 2z_{\rm EC}}}\,. \tag{8.41}$$

Equations (8.39), (8.40) and (8.41) admit a trivial generalization to the case of n-component systems.

As regards the Effective thickness-scales of the two components, the situation is not so straightforward. They can only formally be expressed as series in terms of the local parameters γ and β_z:

$$z_{\rm eff\,H} \equiv \frac{\sigma_{\rm H}}{2\rho_{\rm 0H}} = 2z_{\rm EH}\left[\frac{1}{2}\sqrt{1+\gamma\beta_z}\int_0^1 \frac{du}{\sqrt{(1-u)+\gamma\beta_z(1-u^{\beta_z^{-1}})}}\right]$$

$$= 2z_{\rm EH}\left[\frac{1}{2}+\frac{1}{2}\sum_{n=0}^{+\infty}\frac{(2n+1)!!}{(2n+2)!!}\,\frac{1}{(1+\gamma\beta_z)^{n+1}}\cdot\sum_{k=0}^{n+1}\binom{n+1}{k}(\gamma\beta_z)^k\,\frac{\beta_z}{k+\beta_z\,(n+2-k)}\right]$$

$$\underset{\gamma\beta_z<1}{=} 2z_{\rm EH}\left[1+\frac{1}{2}\sum_{n=0}^{+\infty}\frac{(2n+1)!!}{(2n+2)!!}\cdot\underbrace{\sum_{l=0}^{+\infty}\sum_{k=0}^{n+1}}_{l+k>0}(-1)^l\binom{n+l}{l}\binom{n+1}{k}\cdot(\gamma\beta_z)^{l+k}\,\frac{\beta_z}{k+\beta_z\,(n+2-k)}\right],$$

$$\text{where}\quad u\equiv e^{-\hat{\Phi}}\,; \tag{8.42}$$

$$z_{\rm eff\,C}\equiv\frac{\sigma_{\rm C}}{2\rho_{0\rm C}} = 2z_{\rm EC}\left[\frac{1}{2}\sqrt{1+\gamma\beta_z}\int_0^1\frac{dv}{\sqrt{(1-v^{\beta_z})+\gamma\beta_z\,(1-v)}}\right]$$

$$= 2z_{\rm EC}\left[\frac{1}{2}+\frac{1}{2}\sum_{n=0}^{+\infty}\frac{(2n+1)!!}{(2n+2)!!}\,\frac{1}{(1+\gamma\beta_z)^{n+1}}\cdot\sum_{k=0}^{n+1}\binom{n+1}{k}(\gamma\beta_z)^k\,\frac{1}{k+1+\beta_z\,(n+1-k)}\right]$$

$$\underset{\gamma\beta_z<1}{=} 2z_{\rm EC}\left[\frac{1}{2}\beta_z^{-1}\underbrace{\mathcal{B}(\beta_z^{-1},\tfrac{1}{2})}_{\text{Beta function}}+\frac{1}{2}\sum_{n=0}^{+\infty}\frac{(2n+1)!!}{(2n+2)!!}\cdot\underbrace{\sum_{l=0}^{+\infty}\sum_{k=0}^{n+1}}_{l+k>0}(-1)^l\binom{n+l}{l}\binom{n+1}{k}\cdot(\gamma\beta_z)^{l+k}\,\frac{1}{k+1+\beta_z\,(n+1-k)}\right],$$

$$\text{where}\quad v\equiv e^{-\beta_z^{-1}\hat{\Phi}}\,. \tag{8.43}$$

Similar expressions can be given for $z_{\rm eff\,i}$ scaled in terms of $z_{\rm Gi}$ instead of $z_{\rm Ei}$, making use of the following relations between the asymptotic thickness-scales of the two components:

$$z_{\rm GH} = 2z_{\rm EH}\left[\frac{\sqrt{1+\gamma\beta_z}}{\sqrt{1+\gamma}}\right], \tag{8.44}$$

$$z_{\rm GC} = 2z_{\rm EC}\left[\frac{\sqrt{1+\gamma\beta_z}}{\sqrt{\beta_z\,(1+\gamma)}}\right]. \tag{8.45}$$

Such series representations of $z_{\mathrm{eff\,i}}$ in terms of the local parameters γ and β_z are useful for investigating asymptotic regimes of astrophysical interest. Note that the following inequalities hold between the relevant thickness-scales:

$$2z_{\mathrm{EC}} < z_{\mathrm{eff\,C}} < z_{\mathrm{GC}} < \langle z \rangle < z_{\mathrm{GH}} < z_{\mathrm{eff\,H}} < 2z_{\mathrm{EH}}\,. \qquad (8.46)$$

The determination of the effective thickness-scales of the two components is indeed a *delicate* point. Some authors (e.g., Talbot and Arnett 1975; see also Jog and Solomon 1984), in fact, have superficially generalized equation (8.15) found in the one-component case obtaining the wrong relation $\sigma_{\mathrm{i}} = \rho_{0\mathrm{i}}\,(2z_{\mathrm{Gi}})$, in which the effective thickness-scales are identified with the gaussian thickness-scales. On the other hand, the tricky relation

$$\sigma = \rho_{0\mathrm{H}}\,(4z_{\mathrm{EH}}) + \rho_{0\mathrm{C}}\,(4z_{\mathrm{EC}}) \qquad (8.47)$$

deduced combining equations (8.33) and (8.40) does not imply the further relation $\sigma_{\mathrm{i}} = \rho_{0\mathrm{i}}\,(4z_{\mathrm{Ei}})$, in which the effective thickness-scales are identified with twice the exponential thickness-scales (cf. Mihalas and Binney 1981). The reason for the existence of such a simple relation involving $\rho_{0\mathrm{i}}$ and z_{Ei}, as expressed by equation (8.47), lies in the fact that, while the volume densities are not separately integrable in terms of elementary or special functions, a proper linear combination of them is elementarily integrable. The gaussian and the (self-consistent) exponential approximations to the effective thickness-scales of the two components are thoroughly justified only in the asymptotic regimes in which the system can suitably be represented by a single equivalent component or by two components with the same vertical velocity dispersions. This difficulty is not apparent in the one-component case, where the gaussian, the exponential and the effective thickness-scales are identical apart from numerical factors [see equations (8.13), (8.14) and (8.17)].

The behaviour of the relevant thickness-scales, which play a *significant* role in the discussion of the stability properties of galactic discs when finite-thickness effects are taken into account (see Chapter 9), has been investigated numerically in regimes of the local parameters γ and β_z which are of astrophysical interest, and it is shown in Figures 8.1–8.3.

It is apparent that in general the gaussian and the (self-consistent) exponential approximations to the effective thickness-scales do not work well for both components, except in the asymptotic regimes in which the system approaches the limiting cases mentioned above. Substantial differences with respect to the relevant limiting cases can be ascribed to the strong gravitational coupling between the two components. However, the use of such analytical approximations is reasonably justified for each component. More precisely, the gaussian approximation works fairly well for the cold component, whereas the exponential approximation works fairly well for the hot component.

Finally, note the non-monotonic behaviour characterizing the hot component shown in Figure 8.1, which is not intuitive at all. However, it has no significant implications on the behaviour of the local parameters defined below, whose monotonicity properties turn out to be preserved. Such a peculiar behaviour can also be deduced, together with other qualitative properties, by a careful analytical inspection of the relevant asymptotic regimes.

Let us now introduce the local parameters

$$\alpha \equiv \frac{\sigma_{\rm C}}{\sigma_{\rm H}} = \gamma\beta_{z\,\rm eff}\,, \quad \beta_{z\,\rm eff} \equiv \frac{z_{\rm eff\,C}}{z_{\rm eff\,H}} \quad (0 < \alpha < +\infty,\ 0 < \beta_{z\,\rm eff} < 1)\,, \tag{8.48}$$

since in a flat galactic disc the surface densities are more relevant than the volume densities. More restrictive limitations can be given by considering the chain of inequalities (8.46):

$$\gamma\beta_z < \alpha < \gamma\sqrt{\beta_z}\,, \quad \beta_z < \beta_{z\,\rm eff} < \sqrt{\beta_z}\,, \tag{8.49}$$

where the lower and the upper bounds are nothing but the estimates provided by the (Self-Consistent) Exponential approximation

$$z_{\rm eff\,i} \sim 2z_{\rm Ei} \quad \Longrightarrow \quad \alpha \sim \gamma\beta_z\,, \quad \beta_{z\,\rm eff} \sim \beta_z \tag{8.50}$$

and the Gaussian approximation

$$z_{\rm eff\,i} \sim z_{\rm Gi} \quad \Longrightarrow \quad \alpha \sim \gamma\sqrt{\beta_z}\,, \quad \beta_{z\,\rm eff} \sim \sqrt{\beta_z}\,, \tag{8.51}$$

respectively. The Mixed approximation

$$z_{\rm eff\,H} \approx 2z_{\rm BH}\,, \quad z_{\rm eff\,C} \approx z_{\rm GC} \quad \Longrightarrow$$

$$\alpha \approx \gamma\sqrt{\beta_z}\,\sqrt{\frac{1+\gamma\beta_z}{1+\gamma}}\,, \quad \beta_{z\,\rm eff} \approx \sqrt{\beta_z}\,\sqrt{\frac{1+\gamma\beta_z}{1+\gamma}}\,, \tag{8.52}$$

which should provide more reliable and accurate estimates, as mentioned in the previous paragraph, suggests that the local parameters α and $\beta_{z\,\rm eff}$ are indeed closer to their gaussian approximations in regimes of astrophysical interest.

The behaviour of these local parameters is shown in Figures 8.4–8.5. In particular, in Figure 8.4 the "exact" α is compared to its gaussian and exponential approximations. It is apparent that, although only the exponential approximation is self-consistent, as expressed by equation (8.47), the gaussian approximation works better in regimes of astrophysical interest. Interestingly enough, note the regularization of the divergences characterizing the cold component in the singular limit $\beta_z \to 0$ shown in Figure 8.2.

A simple numerical approximation, much more *accurate* than those considered so far, has been derived employing a rough fitting procedure in the ranges $0 < \gamma \leq 1$ and $0 < \beta_z < 1$, which are those most relevant from an astrophysical point of view. In the following we shall refer to it as the Fit approximation:

$$\alpha \simeq \gamma \beta_z^{3/5} , \quad \beta_{z\,\mathrm{eff}} \simeq \beta_z^{3/5} . \tag{8.53}$$

No intuitive physical interpretation can be given for this specific power-law. The remarkable accuracy of such a simple approximation is shown in Figure 8.6. When $\gamma \gtrsim 1$ and $\beta_z \lesssim 5 \cdot 10^{-2}$ (these estimates are quite rough and mutually dependent) this approximation does no more adequately represent the real physical situation. More accurate approximations can be obtained by sacrificing the mathematical feature which it shares with the gaussian and the exponential approximations and makes it so attractive: it avoids the γ-dependence of $\beta_{z\,\mathrm{eff}}$.

Using the fit approximation, we have expressed all the dimensionless quantities introduced so far (the ratios between the relevant thickness-scales and the remaining local parameters) in terms of α, β_z and α, $\beta_{z\,\mathrm{eff}}$. These parametrizations will be used in Chapter 9 for investigating the stability of galactic discs when finite-thickness effects are taken into account. Note that this "inversion of coordinates" is made possible by the fact that the three parametrizations considered are diffeomorphic, as shown in Figures 8.4–8.5.

The results are shown in Figures 8.7–8.11 for the parametrization involving α, β_z and in Figures 8.14–8.18 for the parametrization involving α, $\beta_{z\,\mathrm{eff}}$. Non-monotonic features are present for both components, and even in these cases can easily be justified by simple analytical considerations concerning the relevant asymptotic regimes. The location and height of the corresponding minima occurring at relatively low β_z or $\beta_{z\,\mathrm{eff}}$ is expected to be inaccurate, also because they imply relatively high γ at fixed α.

Finally, we have performed a test to check "a posteriori" the accuracy of the fit approximation. We have first recalculated the relevant local parameters at the second step of iteration, and then we have compared these values to those obtained at the first step of iteration. The result of this test is shown in Figures 8.12–8.13 for the parametrization involving α, β_z and in Figures 8.19–8.20 for the parametrization involving α, $\beta_{z\,\mathrm{eff}}$. In standard regimes the agreement is satisfactory, whereas significant discrepancies occur near the "forbidden" asymptotic regimes mentioned in the previous paragraph, as expected.

References

Bahcall J. N. (1984), *Ap. J.* **276**, 156.
Bahcall J. N. and Casertano S. (1984), *Ap. J. Lett.* **284**, L35.
Boulares A. (1989), *A. Ap.* **209**, 21.

Gradshteyn I. S. and Ryzhik I. M. (1980), *Table of Integrals, Series, and Products*, Academic Press, New York.

Jog C. J. and Solomon P. M. (1984), *Ap. J.* **276**, 114.

Mihalas D. and Binney J. (1981), *Galactic Astronomy*, Freeman, San Francisco.

Prudnikov A. P., Brychkov Yu. A. and Marichev O. I. (1986), *Integrals and Series*, 2 Vols., Gordon and Breach, New York.

Romeo A. B. (1987), *M. Phil. Thesis*, SISSA, Trieste, Italy.

Romeo A. B. (1990a), in *Chemical and Dynamical Evolution of Galaxies*, Giardini, Pisa, Ferrini F., Franco J. and Matteucci F. eds., p. 463 (Elba Island, Italy, 1989).

Romeo A. B. (1990b), "Self-Regulation and Secular Heating of Galactic Discs", to appear in *IAU Eur. Reg. Astron. Meeting* **12** (Davos, Switzerland, 1990).

Talbot R. J. Jr. and Arnett W. D. (1975), *Ap. J.* **197**, 551.

Vandervoort P. O. (1967), *Ap. J.* **147**, 91.

Vandervoort P. O. (1970), *Ap. J.* **161**, 67.

Woolley R. v. d. R. (1957), *M.N.R.A.S.* **117**, 198.

Figure Captions

Figure 8.1. Ratio of the gaussian to the exponential thickness-scale and ratios of the effective thickness-scale to the asymptotic thickness-scales of the hot component as functions of the local parameters γ and β_z in regimes of astrophysical interest. Note the non-monotonic behaviour of $z_{\mathrm{eff\,H}}/2z_{\mathrm{EH}}$ with varying β_z at fixed γ, which does not appear in the case of the cold component.

Figure 8.2. Same as Figure 8.1 for the cold component.

Figure 8.3. Ratios of the global effective thickness-scale of the system to the effective thickness-scales of the two components as functions of the local parameters γ and β_z in regimes of astrophysical interest.

Figure 8.4. Ratio between the surface densities of the two components as a function of the local parameters γ and β_z in regimes of astrophysical interest, compared to its gaussian and (self-consistent) exponential approximations. Note that none of such analytical approximations is particularly accurate throughout the ranges considered.

Figure 8.5. Ratio between the effective thickness-scales of the two components as a function of the local parameters γ and β_z in regimes of astrophysical interest.

Figure 8.6. Comparison between the "exact" α and its fit approximation. Note the *remarkable* accuracy of such a simple approximation throughout the ranges considered, except for $\gamma \gtrsim 1$ and $\beta_z \lesssim 5 \cdot 10^{-2}$.

Figure 8.7. Same as Figure 8.1 for the parametrization involving α, β_z.

Figure 8.8. Same as Figure 8.2 for the parametrization involving α, β_z. Note the non-monotonic behaviour of $z_{\mathrm{eff\,C}}/z_{\mathrm{GC}}$ with varying β_z at fixed α, which does not appear in the case of the parametrization involving γ, β_z.

Figure 8.9. Same as Figure 8.3 for the parametrization involving α, β_z.

Figure 8.10. Ratio between the volume densities of the two components at $z = 0$ as a function of the local parameters α and β_z in regimes of astrophysical interest.

Figure 8.11. Ratio between the effective thickness-scales of the two components as a function of the local parameter β_z.

Figure 8.12. Result of the test performed to check "a posteriori" the accuracy of the fit approximation to γ. Note the satisfactory agreement in standard regimes.

Figure 8.13. Same as Figure 8.12 for $\beta_{z\,\mathrm{eff}}$. Significant discrepancies occur at low β_z, as expected.

Figure 8.14. Same as Figure 8.7 for the parametrization involving α, $\beta_{z\,\mathrm{eff}}$.

Figure 8.15. Same as Figure 8.8 for the parametrization involving α, $\beta_{z\,\mathrm{eff}}$.

Figure 8.16. Same as Figure 8.9 for the parametrization involving α, $\beta_{z\,\mathrm{eff}}$.

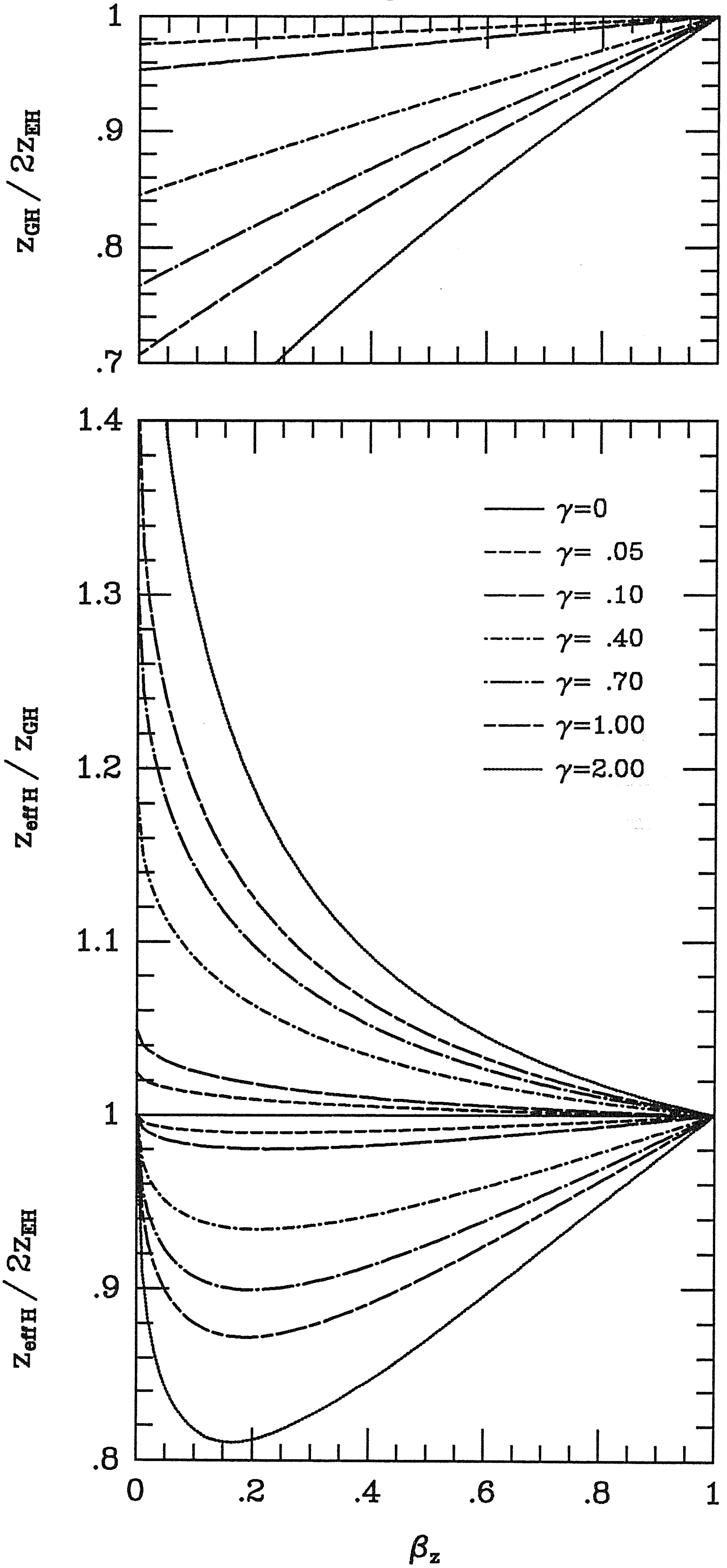
Figure 8.17. Same as Figure 8.10 for the parametrization involving α, $\beta_{z\,\mathrm{eff}}$.

Figure 8.18. Ratio between the vertical velocity dispersions of the two components as a function of the local parameter $\beta_{z\,\mathrm{eff}}$.

Figure 8.19. Same as Figure 8.12 for the parametrization involving α, $\beta_{z\,\mathrm{eff}}$.

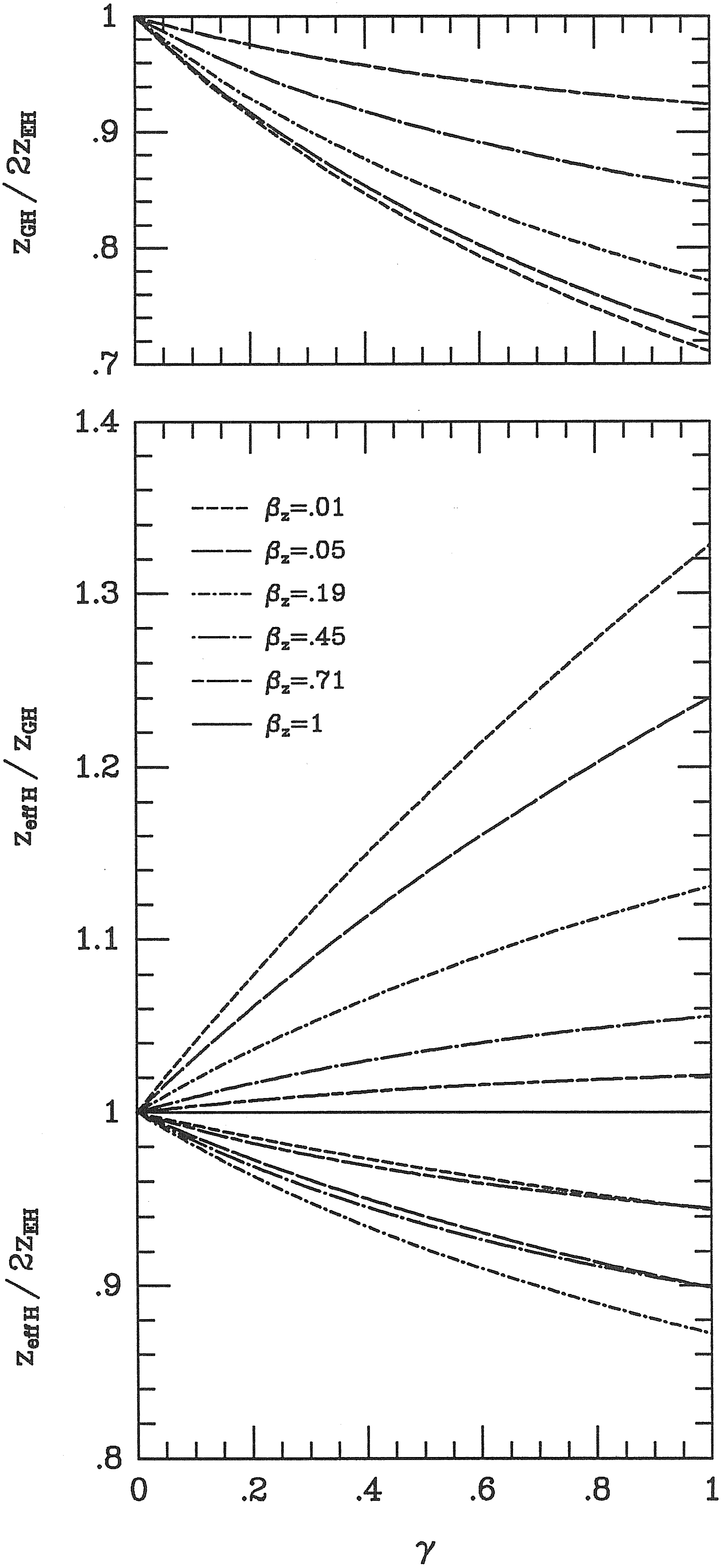
Figure 8.20. Result of the test performed to check "a posteriori" the accuracy of the fit approximation to β_z.

Figure 8.1



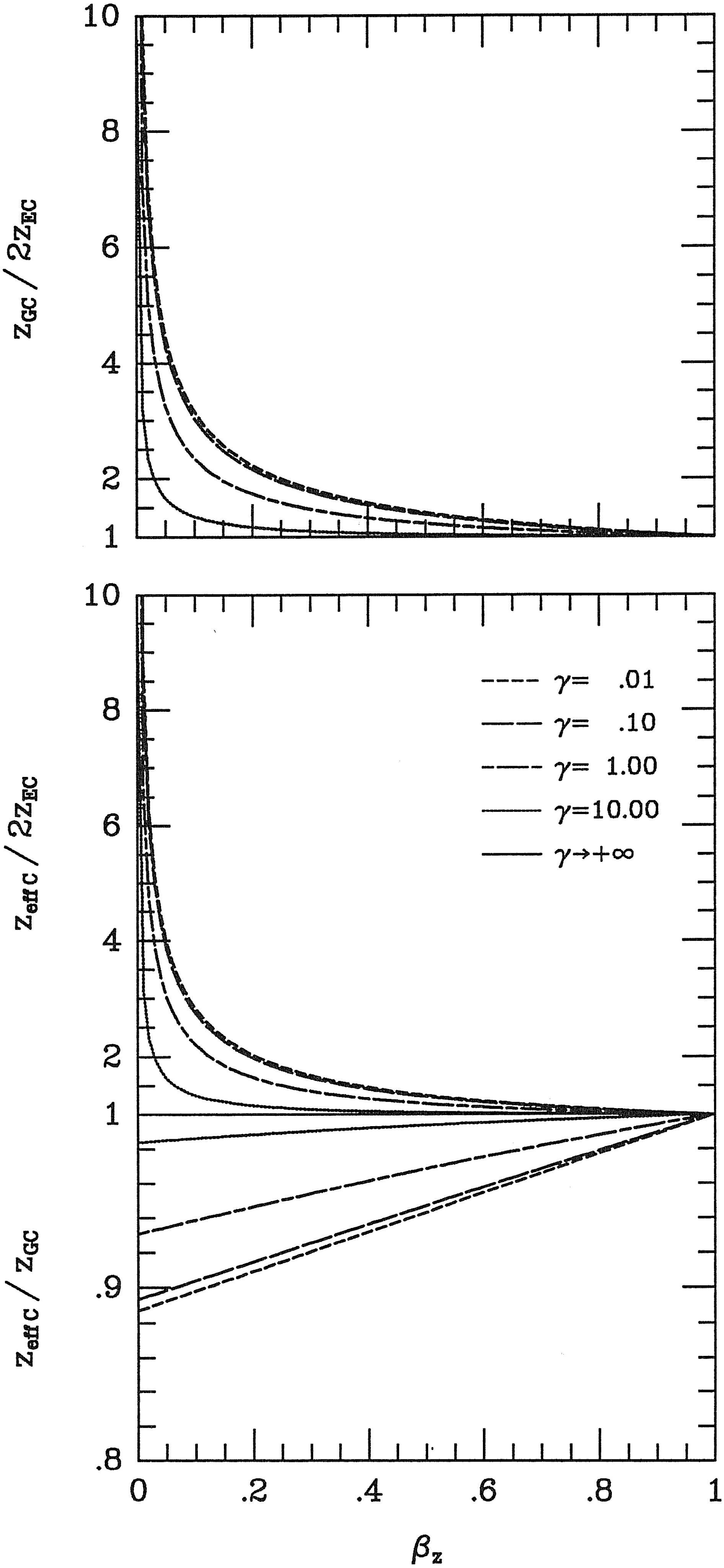
$z_{GH} / 2z_{EH}$
1
.9
.8
.7
1.4
1.3
1.2
1.1
1
.9
.8
$z_{eff\,H} / z_{GH}$
$z_{eff\,H} / 2z_{EH}$
γ=0
γ= .05
γ= .10
γ= .40
γ= .70
γ=1.00
γ=2.00
0
.2
.4
.6
.8
1
β_z

Figure 8.1



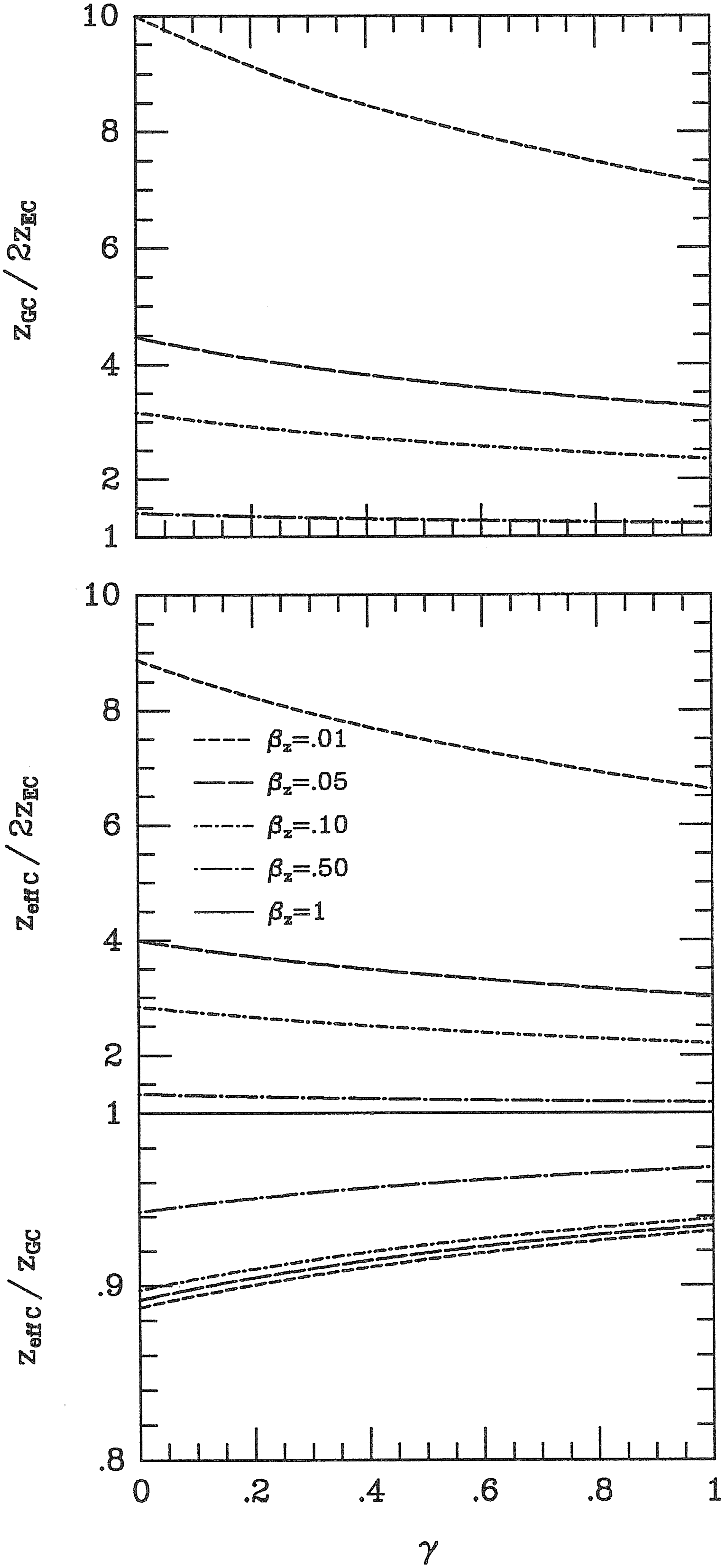
1
.9
.8
.7
$z_{GH} / 2z_{EH}$
1.4
1.3
1.2
1.1
1
.9
.8
β_z=.01
β_z=.05
β_z=.19
β_z=.45
β_z=.71
β_z=1
$z_{eff\,H} / z_{GH}$
$z_{eff\,H} / 2z_{EH}$
0
.2
.4
.6
.8
1
γ

Figure 8.2



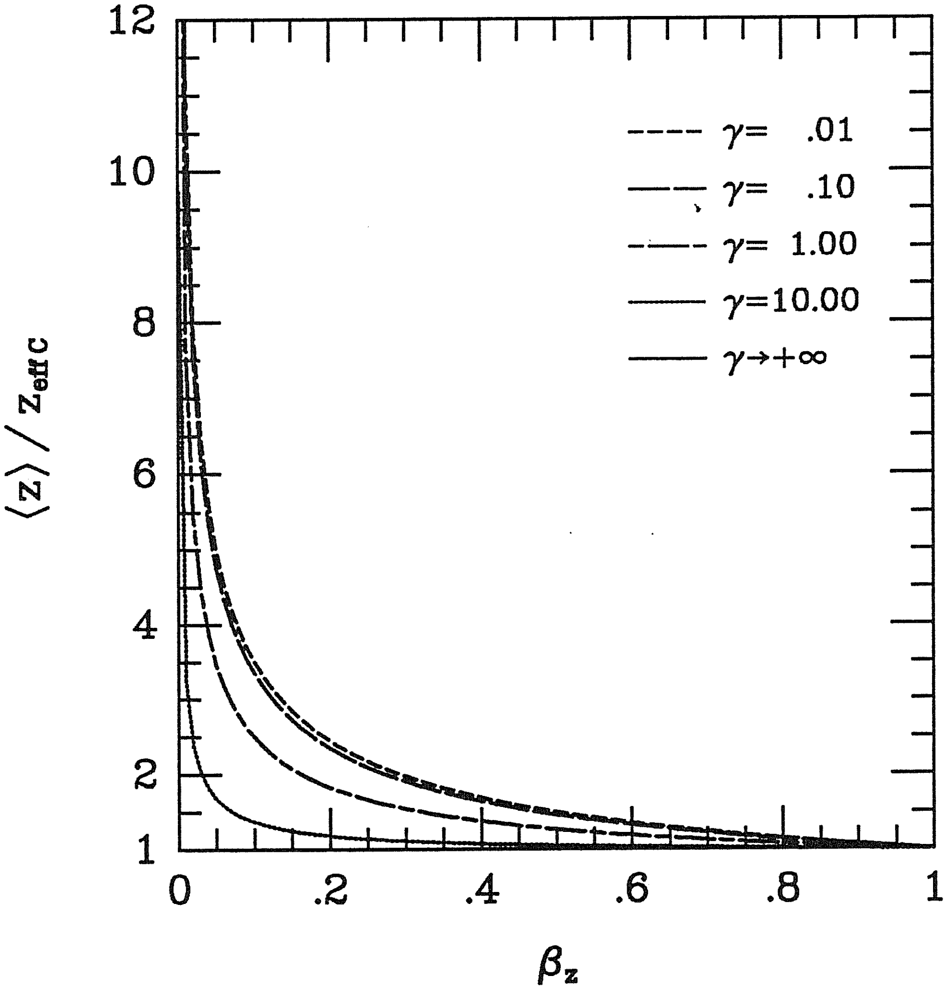
10
8
6
4
2
1
$z_{GC} / 2z_{EC}$
10
8
6
4
2
1
.9
.8
$z_{eff\,C} / 2z_{EC}$
$z_{eff\,C} / z_{GC}$
γ= .01
γ= .10
γ= 1.00
γ=10.00
γ→+∞
0
.2
.4
.6
.8
1
β_z

Figure 8.2



10
8
6
4
2
1
$z_{GC} / 2z_{EC}$
10
8
6
4
2
1
.9
.8
$z_{eff\,C} / 2z_{EC}$
$z_{eff\,C} / z_{GC}$
β_z=.01
β_z=.05
β_z=.10
β_z=.50
β_z=1
0
.2
.4
.6
.8
1
γ

Figure 8.3


12
10
8
6
4
2
1
0
.2
.4
.6
.8
1
γ= .01
γ= .10
γ= 1.00
γ=10.00
γ→+∞
⟨z⟩ / z_eff C
β_z

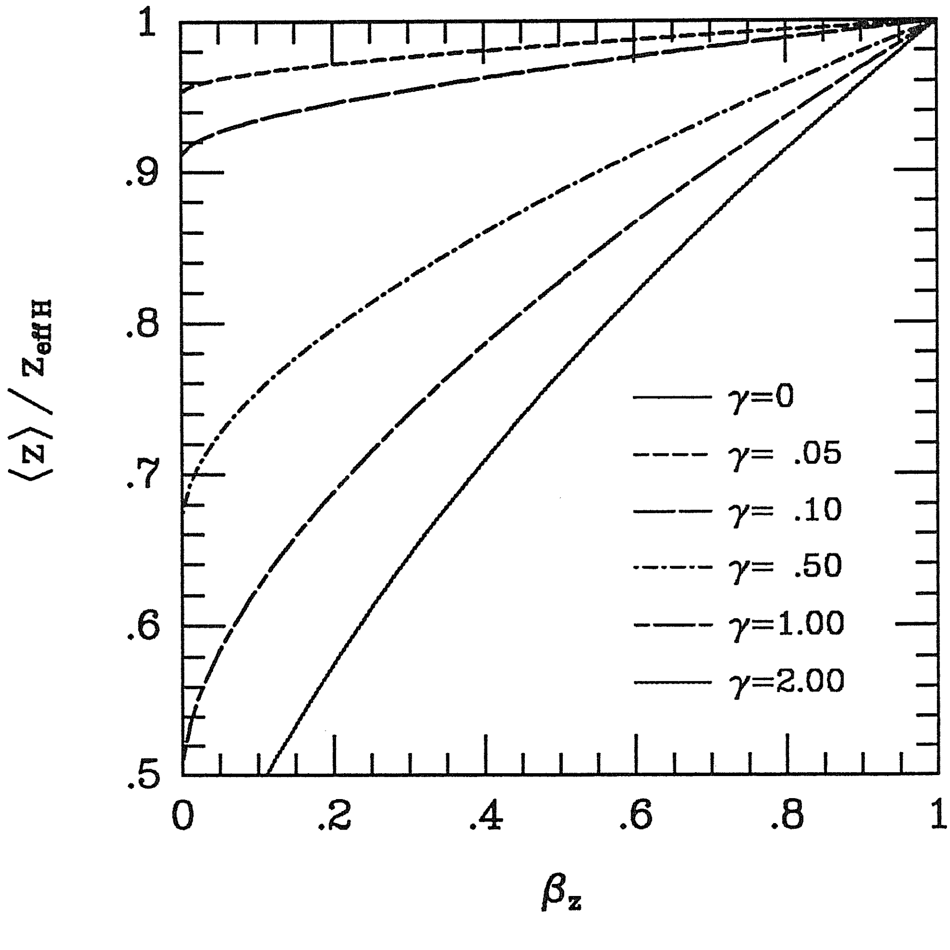
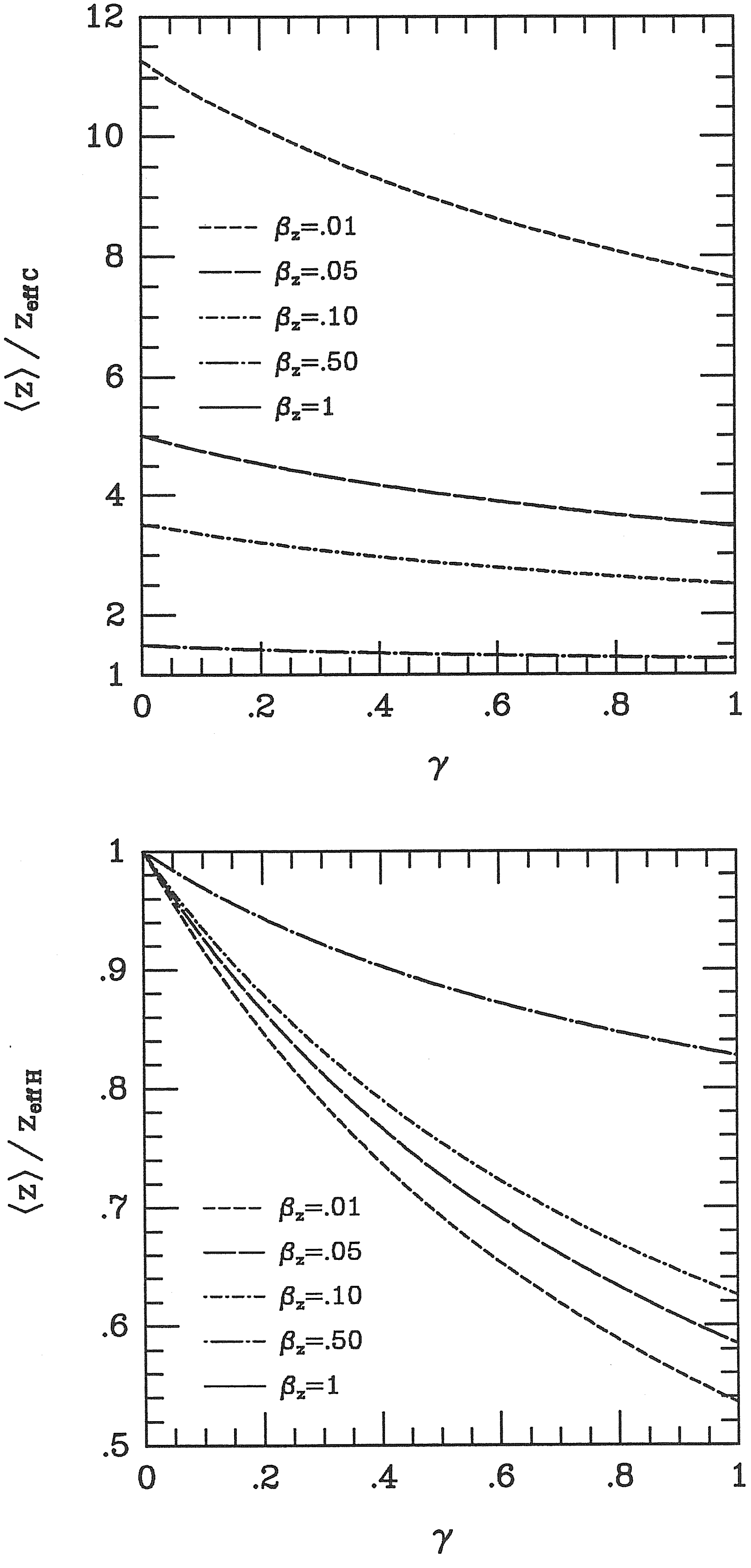
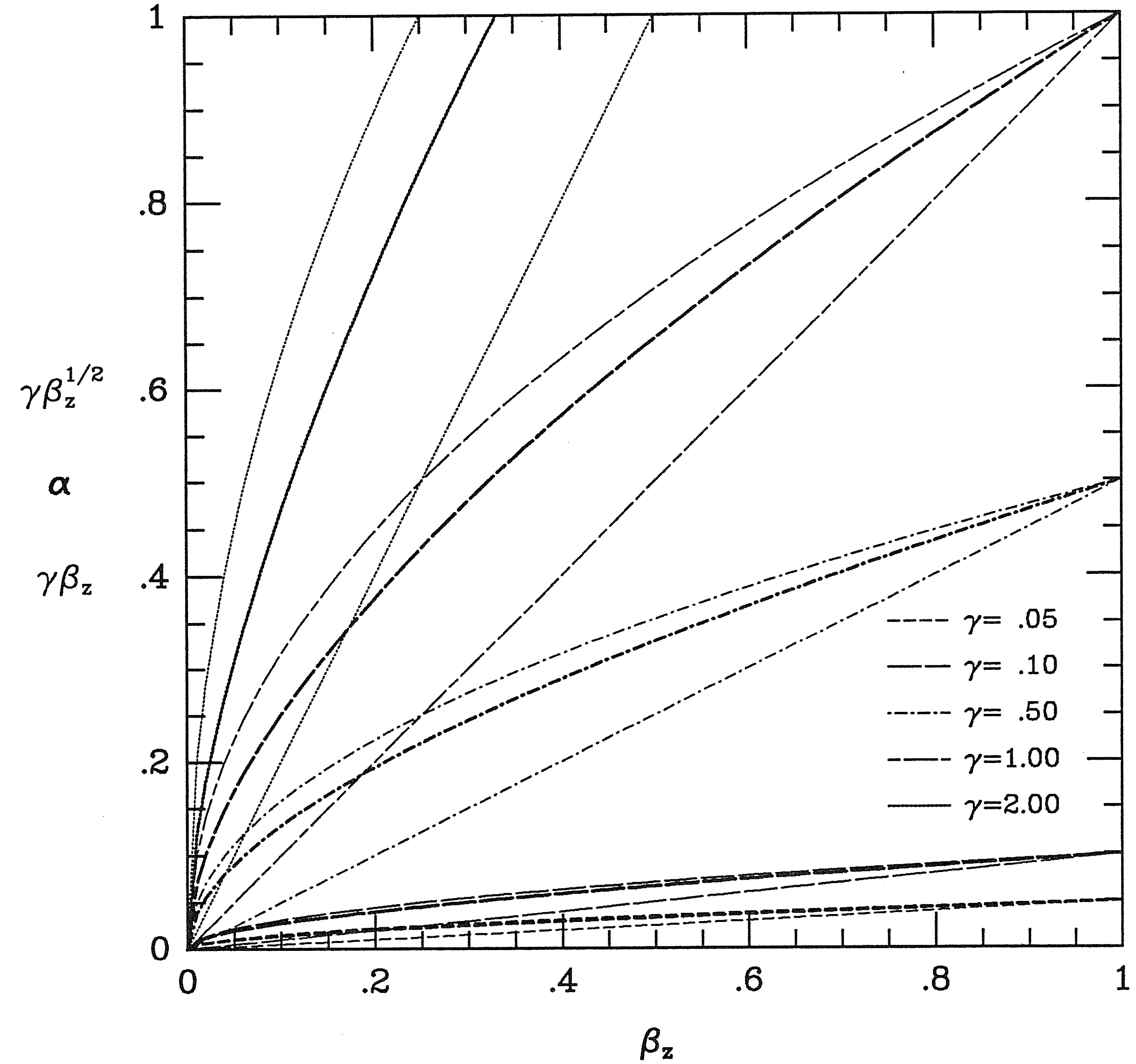

1
.9
.8
.7
.6
.5
0
.2
.4
.6
.8
1
γ=0
γ= .05
γ= .10
γ= .50
γ=1.00
γ=2.00
⟨z⟩ / z_eff H
β_z

Figure 8.3



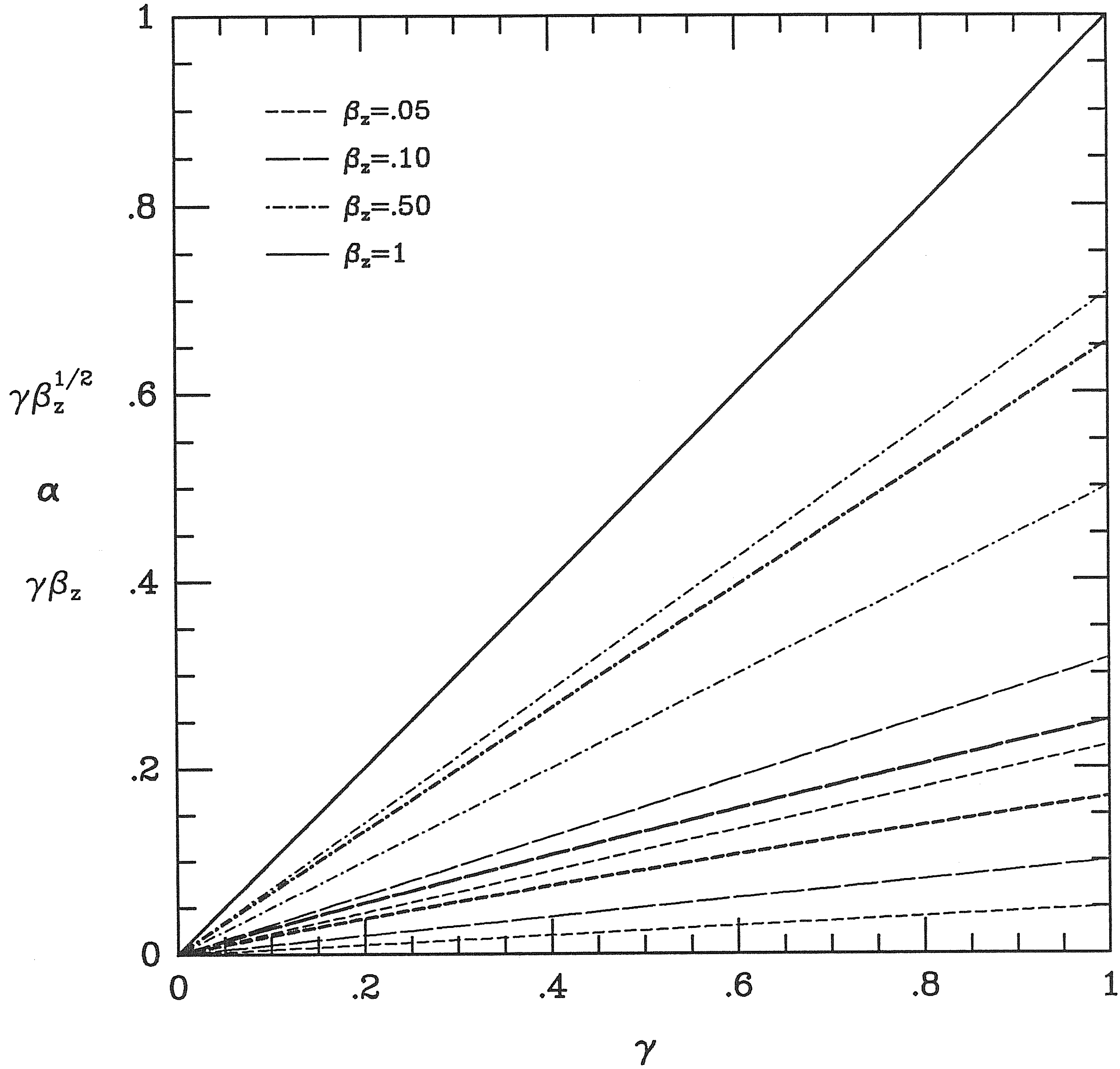
12
10
8
6
4
2
1
0
.2
.4
.6
.8
1
$\langle z \rangle / z_{\mathrm{eff\,C}}$
γ
β_z=.01
β_z=.05
β_z=.10
β_z=.50
β_z=1
1
.9
.8
.7
.6
.5
0
.2
.4
.6
.8
1
$\langle z \rangle / z_{\mathrm{eff\,H}}$
γ
β_z=.01
β_z=.05
β_z=.10
β_z=.50
β_z=1

Figure 8.4



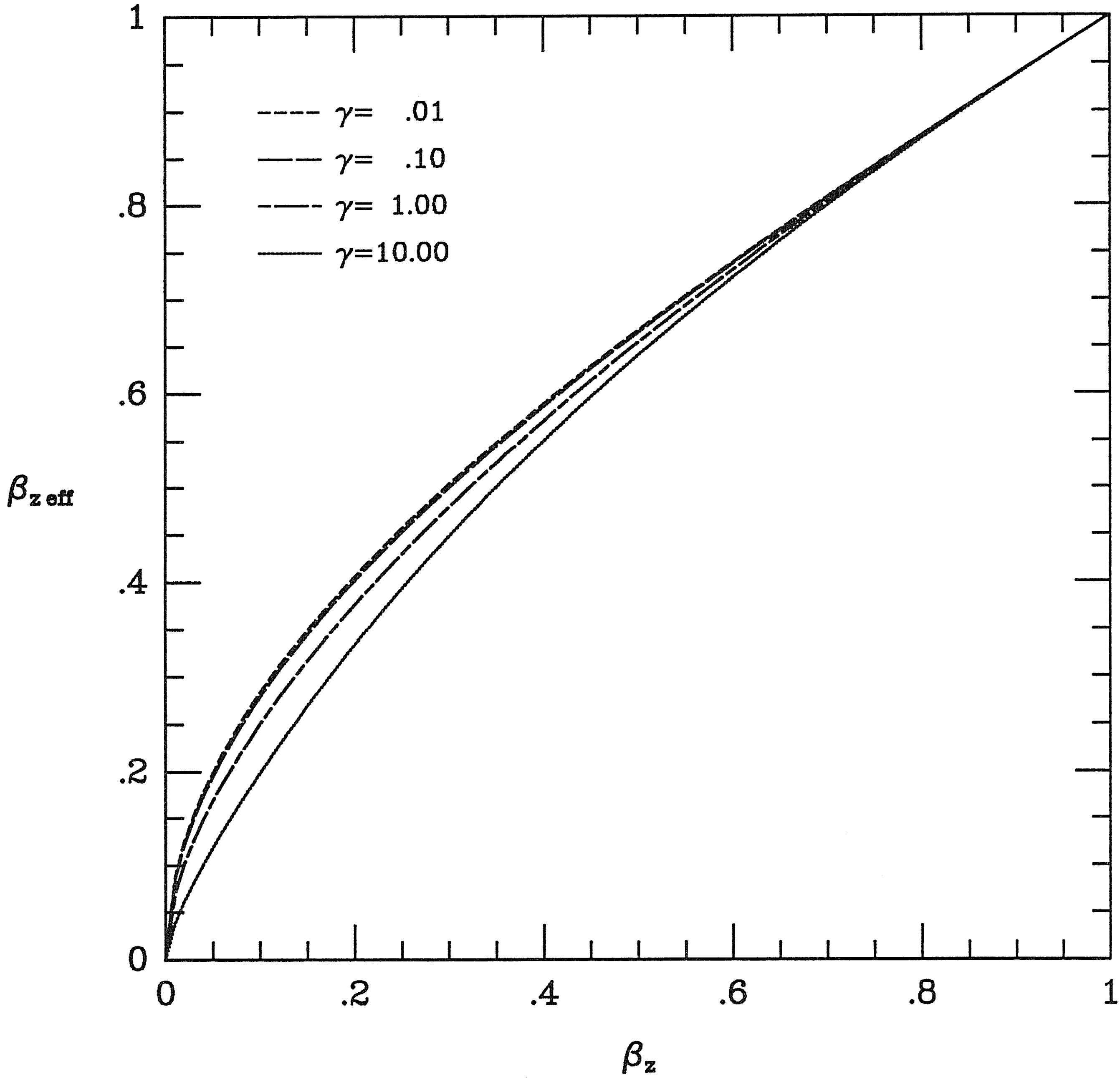
1
.8
.6
.4
.2
0
$\gamma\beta_z^{1/2}$
α
$\gamma\beta_z$
γ= .05
γ= .10
γ= .50
γ=1.00
γ=2.00
0
.2
.4
.6
.8
1
β_z

Figure 8.4



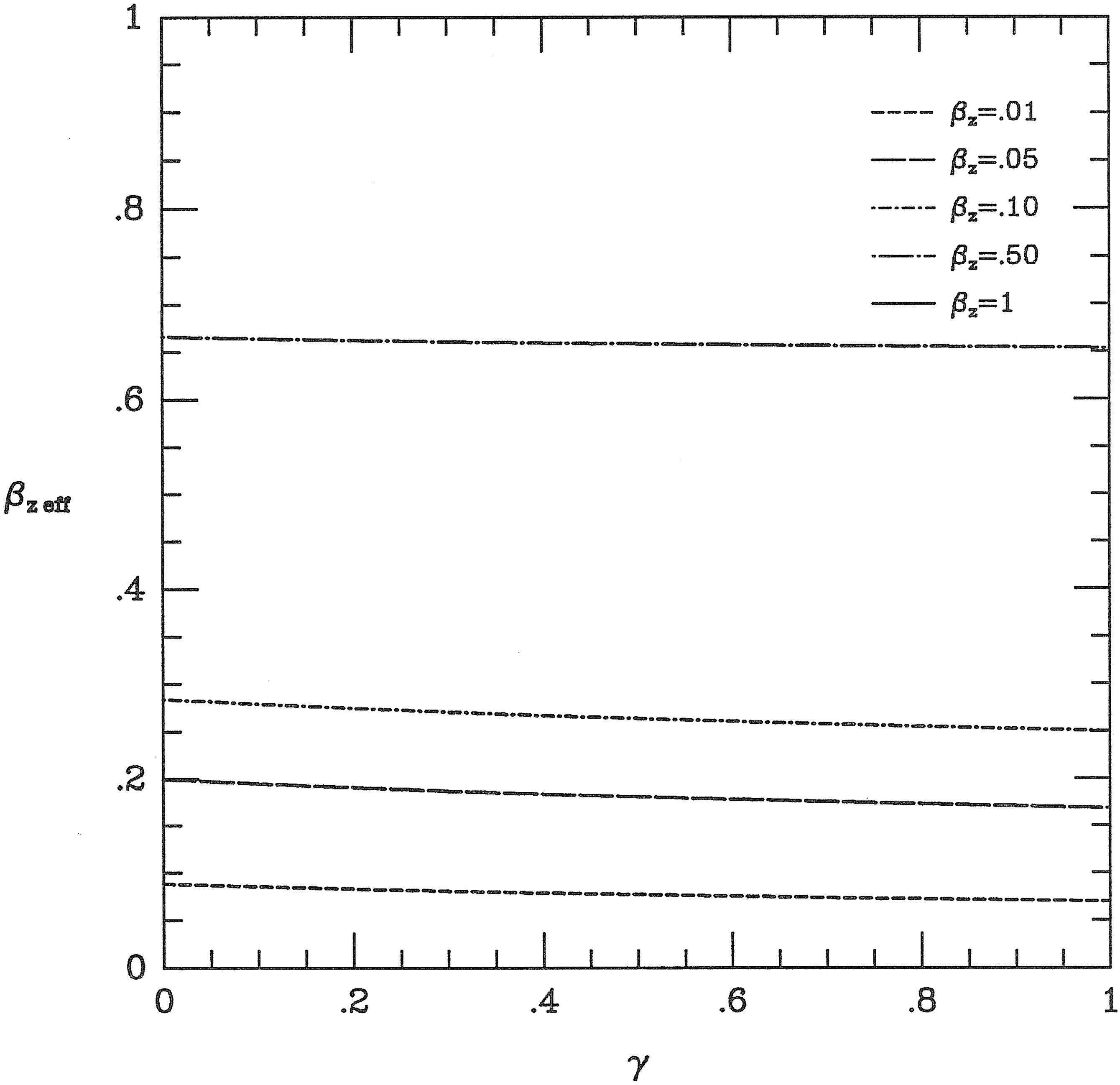
1
.8
.6
.4
.2
0
β_z=.05
β_z=.10
β_z=.50
β_z=1
$\gamma\beta_z^{1/2}$
α
$\gamma\beta_z$
0
.2
.4
.6
.8
1
γ

Figure 8.5



1
.8
.6
.4
.2
0
$\beta_{z\,\mathrm{eff}}$
γ= .01
γ= .10
γ= 1.00
γ=10.00
0
.2
.4
.6
.8
1
β_z

Figure 8.5



1
.8
.6
.4
.2
0
0
.2
.4
.6
.8
1
$\beta_{z\,eff}$
γ
β_z=.01
β_z=.05
β_z=.10
β_z=.50
β_z=1

Figure 8.6

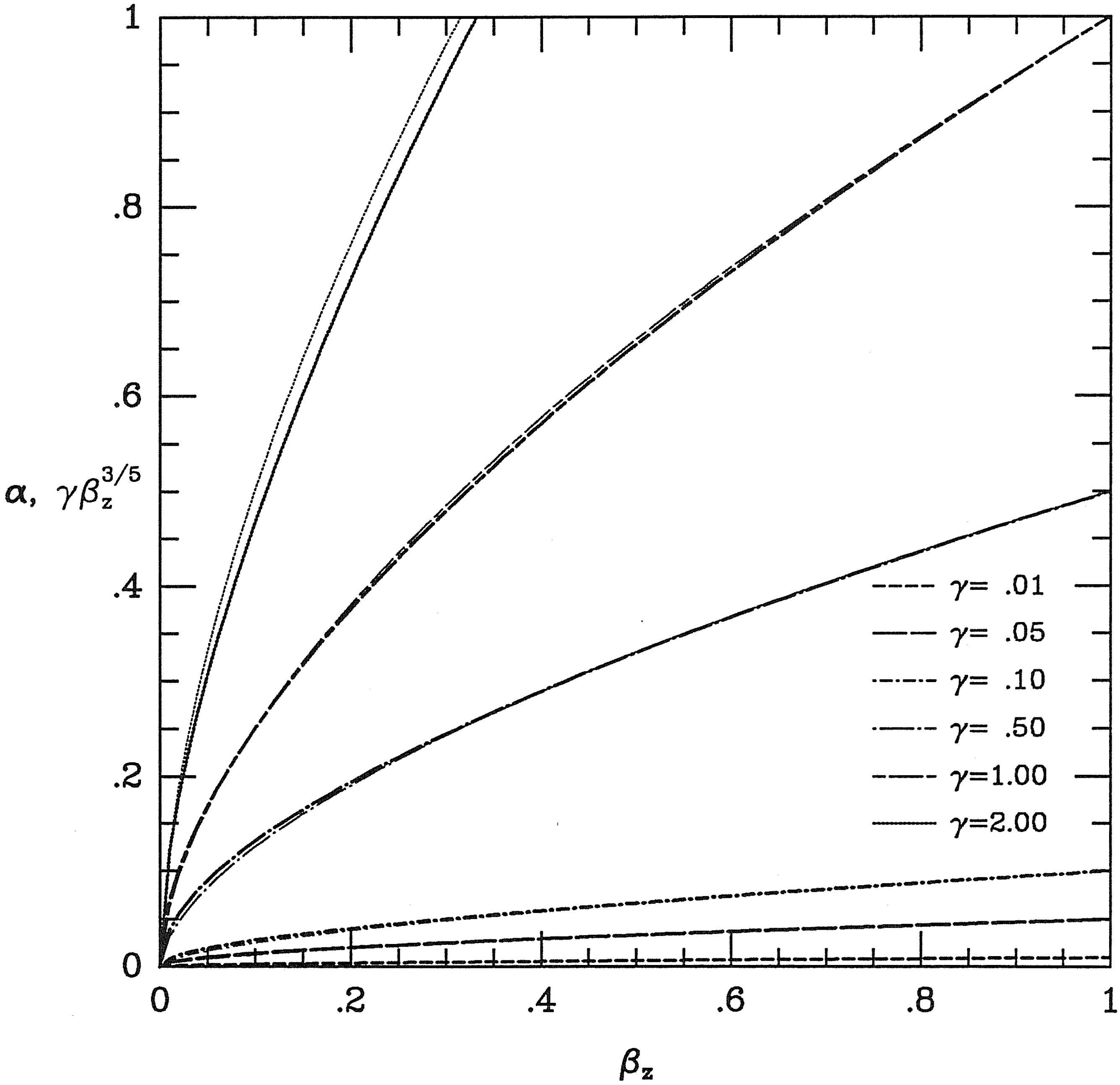
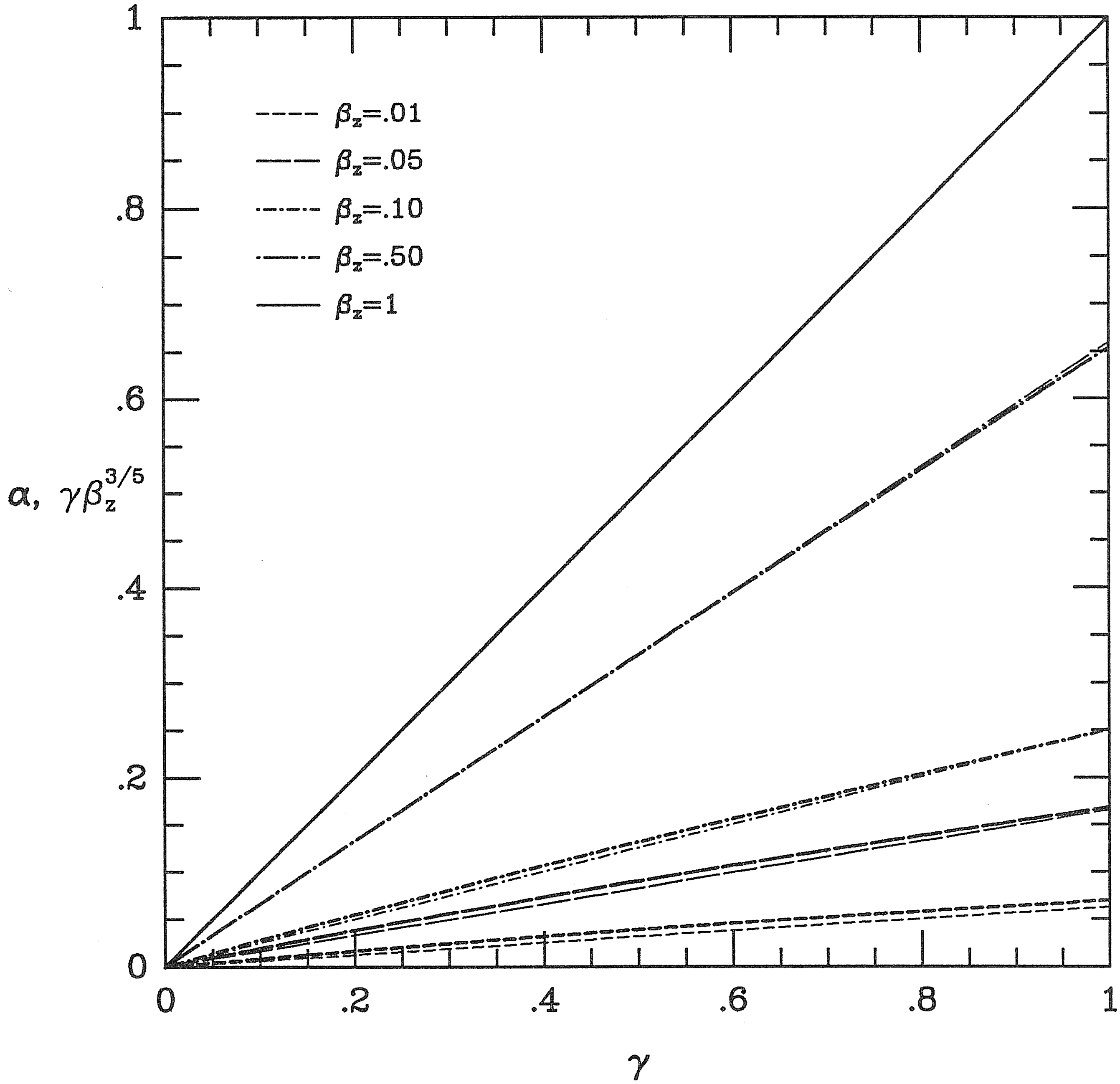
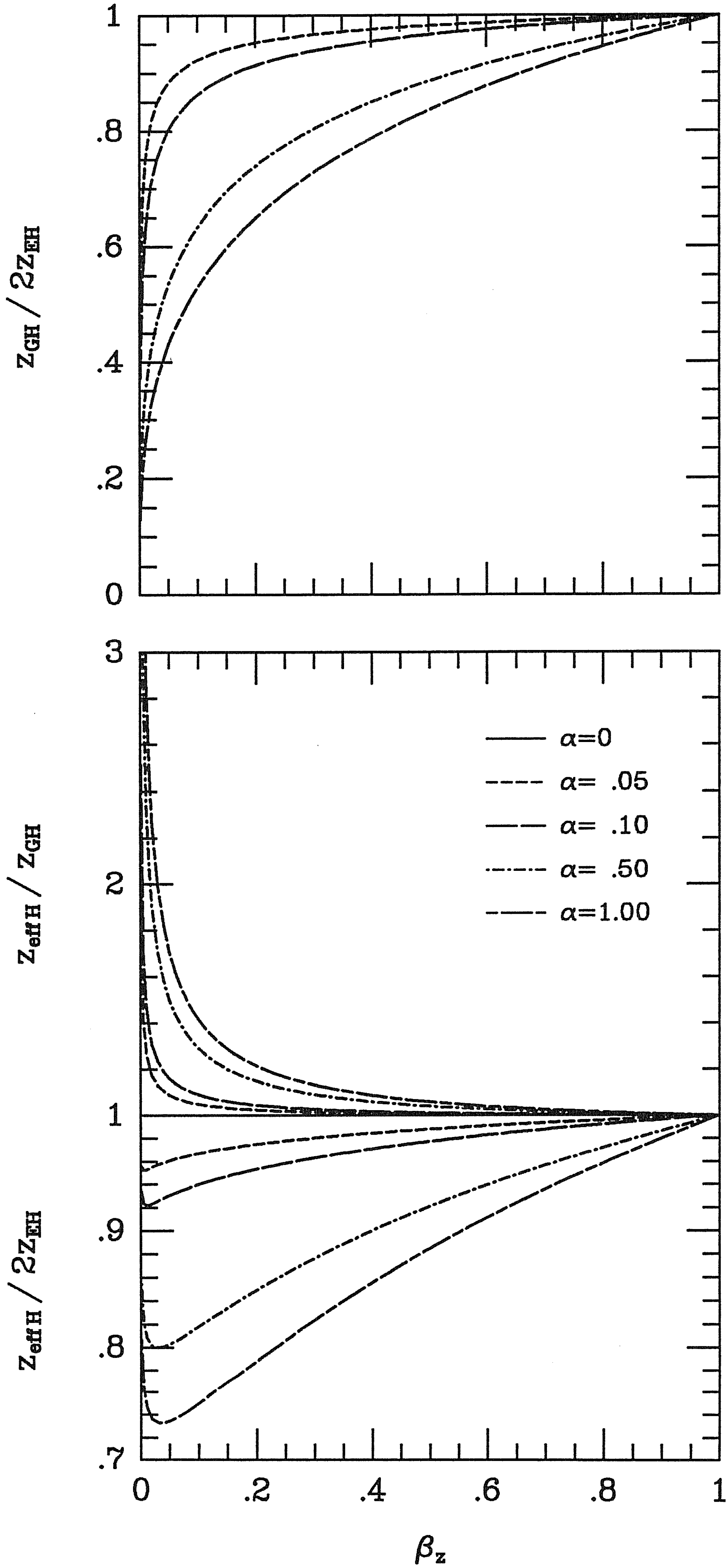


Figure 8.6



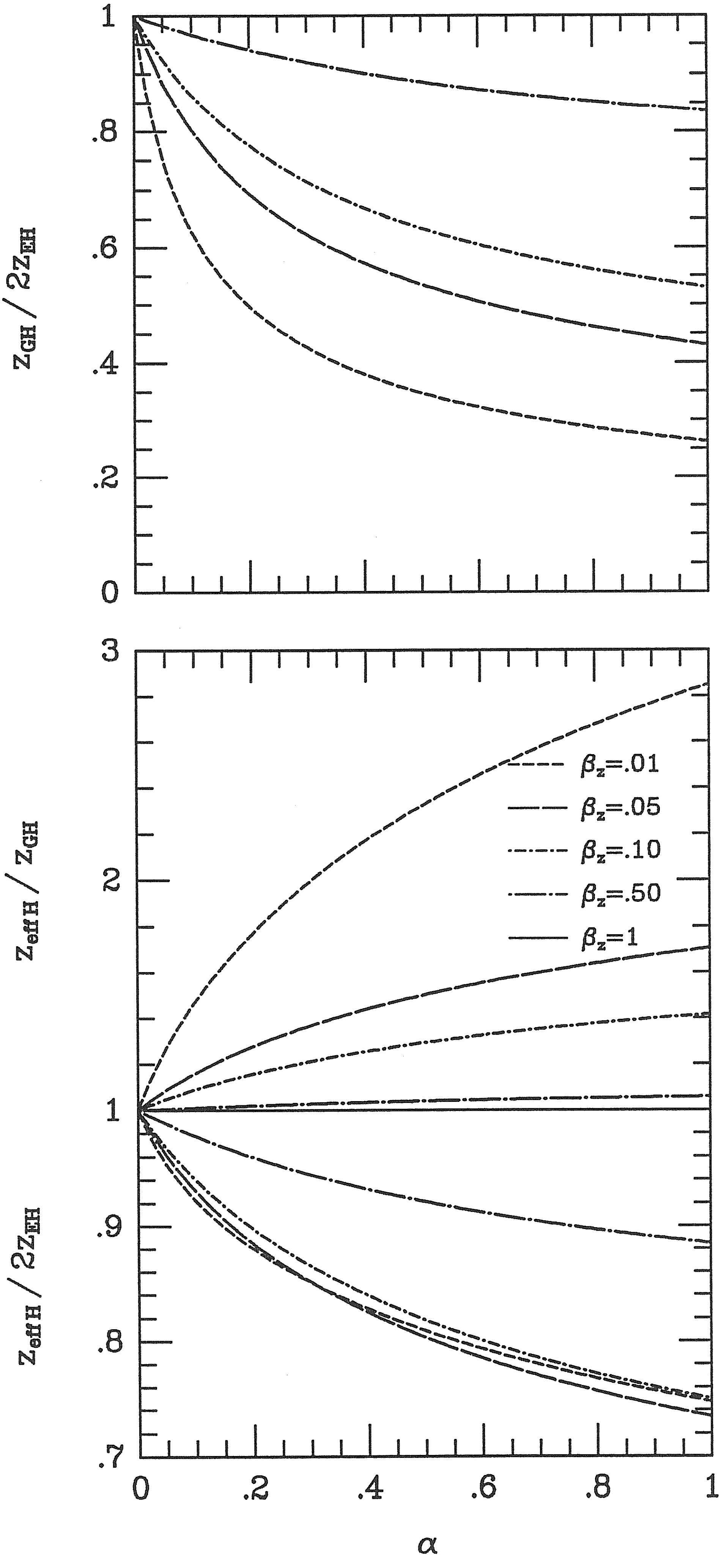
β_z=.01
β_z=.05
β_z=.10
β_z=.50
β_z=1
$\alpha,\ \gamma\beta_z^{3/5}$
γ
0
.2
.4
.6
.8
1

Figure 8.7



$Z_{GH} / 2Z_{EH}$
$Z_{eff\,H} / Z_{GH}$
$Z_{eff\,H} / 2Z_{EH}$
β_z
α=0
α= .05
α= .10
α= .50
α=1.00
0
.2
.4
.6
.8
1
.7
.8
.9
2
3

Figure 8.7



$z_{GH}/2z_{EH}$
$z_{eff\,H}/z_{GH}$
$z_{eff\,H}/2z_{EH}$
β_z=.01
β_z=.05
β_z=.10
β_z=.50
β_z=1
α

Figure 8.8

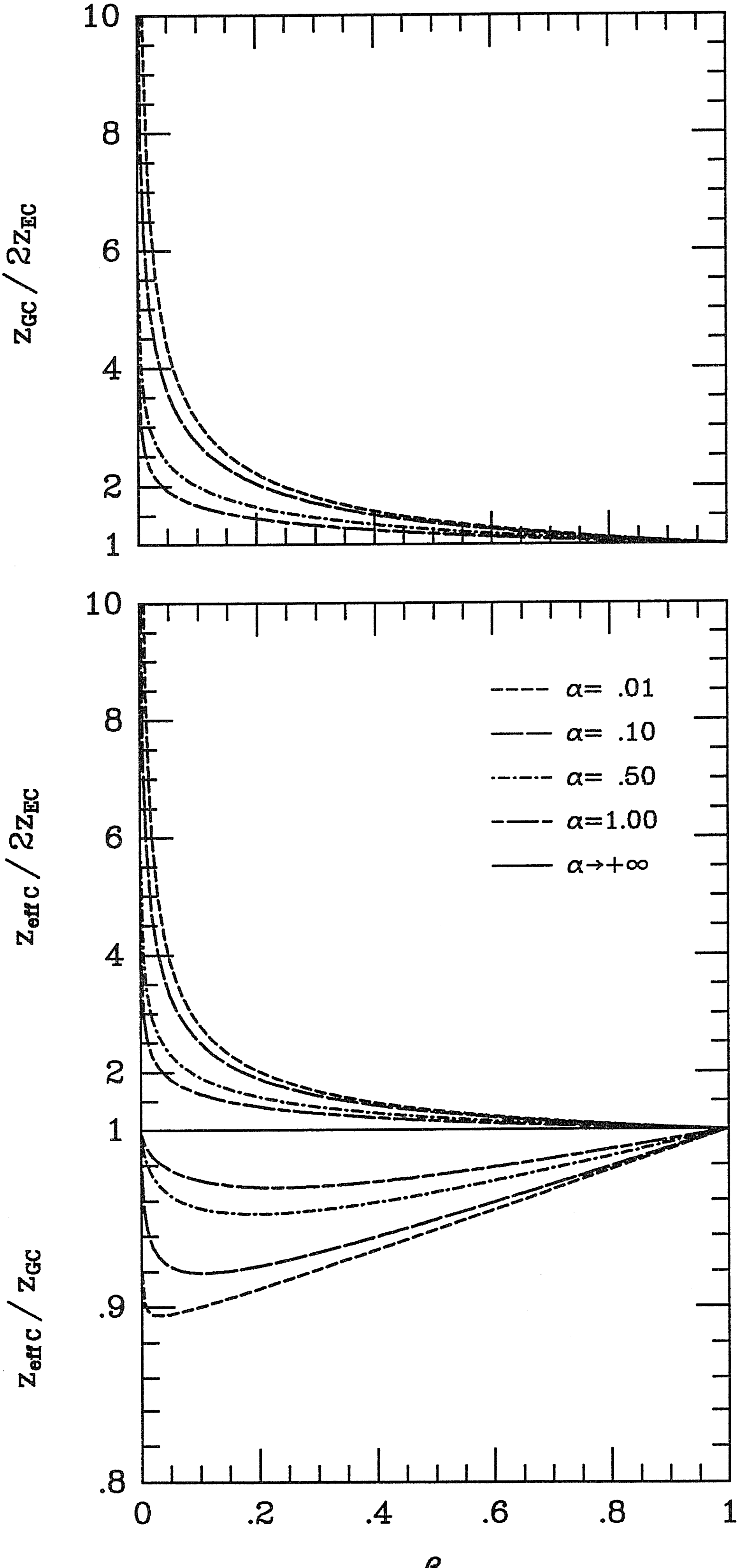
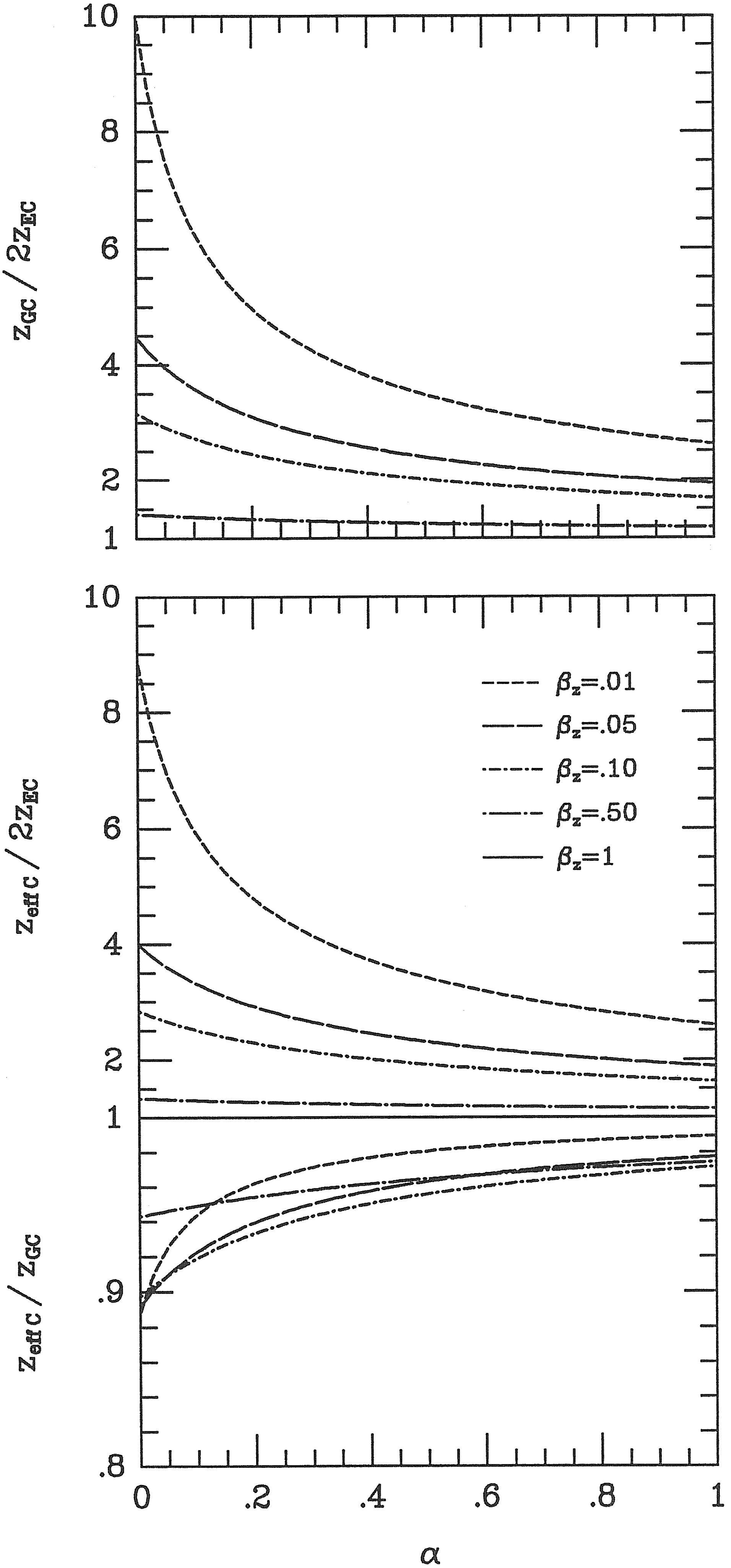
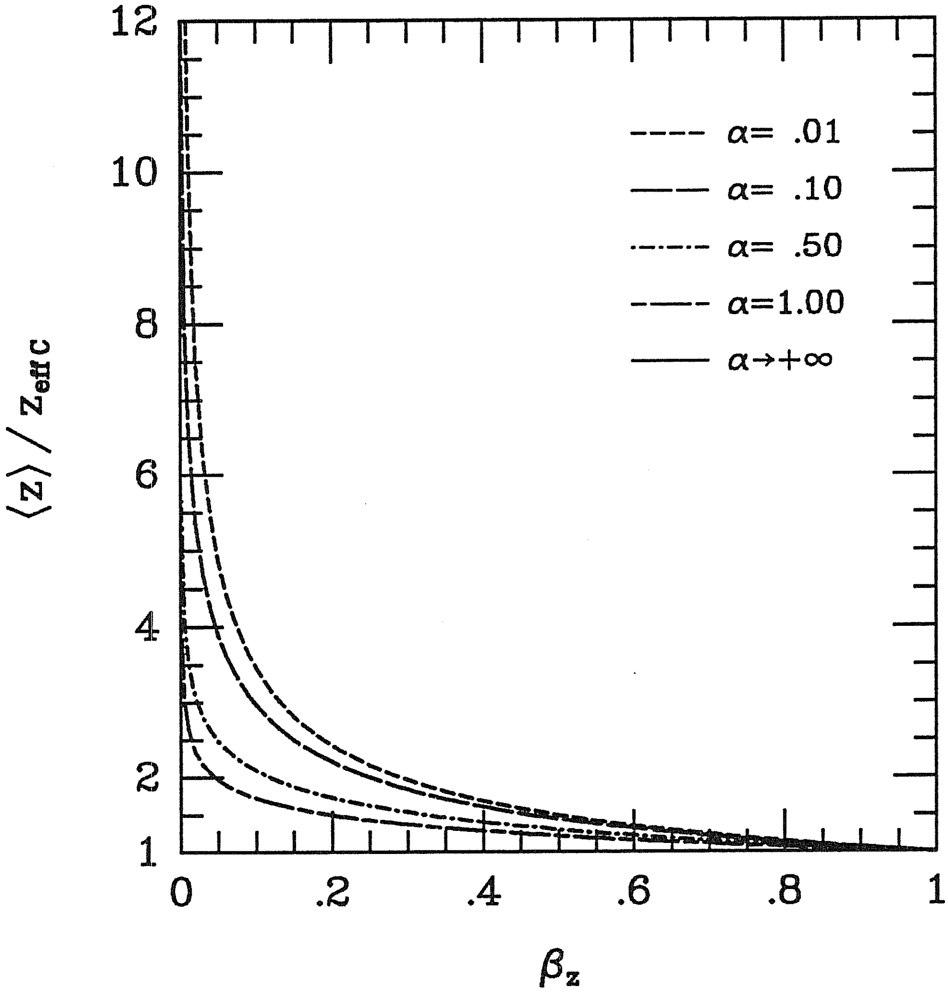

10
8
6
4
2
1
$z_{GC} / 2z_{EC}$
10
8
6
4
2
1
.9
.8
$z_{eff\,C} / 2z_{EC}$
$z_{eff\,C} / z_{GC}$
α= .01
α= .10
α= .50
α=1.00
α→+∞
0
.2
.4
.6
.8
1
β_z

Figure 8.8

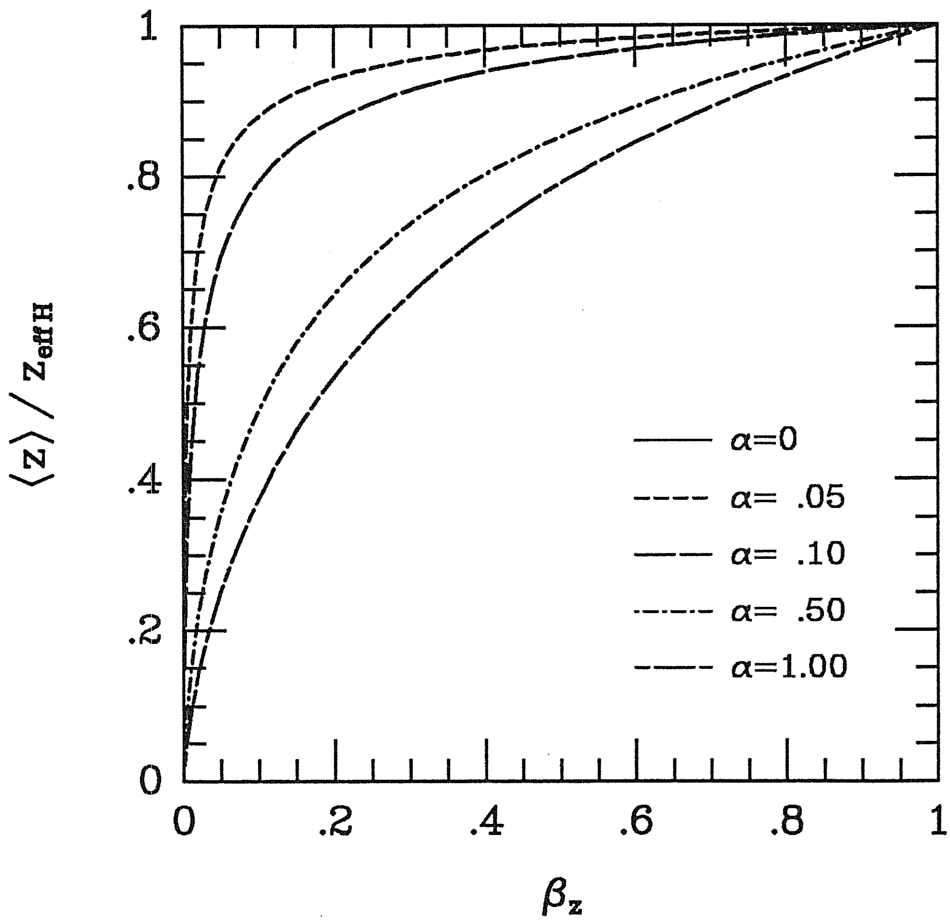


10
8
6
4
2
1
$Z_{GC} / 2Z_{EC}$
β_z=.01
β_z=.05
β_z=.10
β_z=.50
β_z=1
$Z_{eff\,C} / 2Z_{EC}$
$Z_{eff\,C} / Z_{GC}$
.9
.8
0
.2
.4
.6
.8
1
α

Figure 8.9

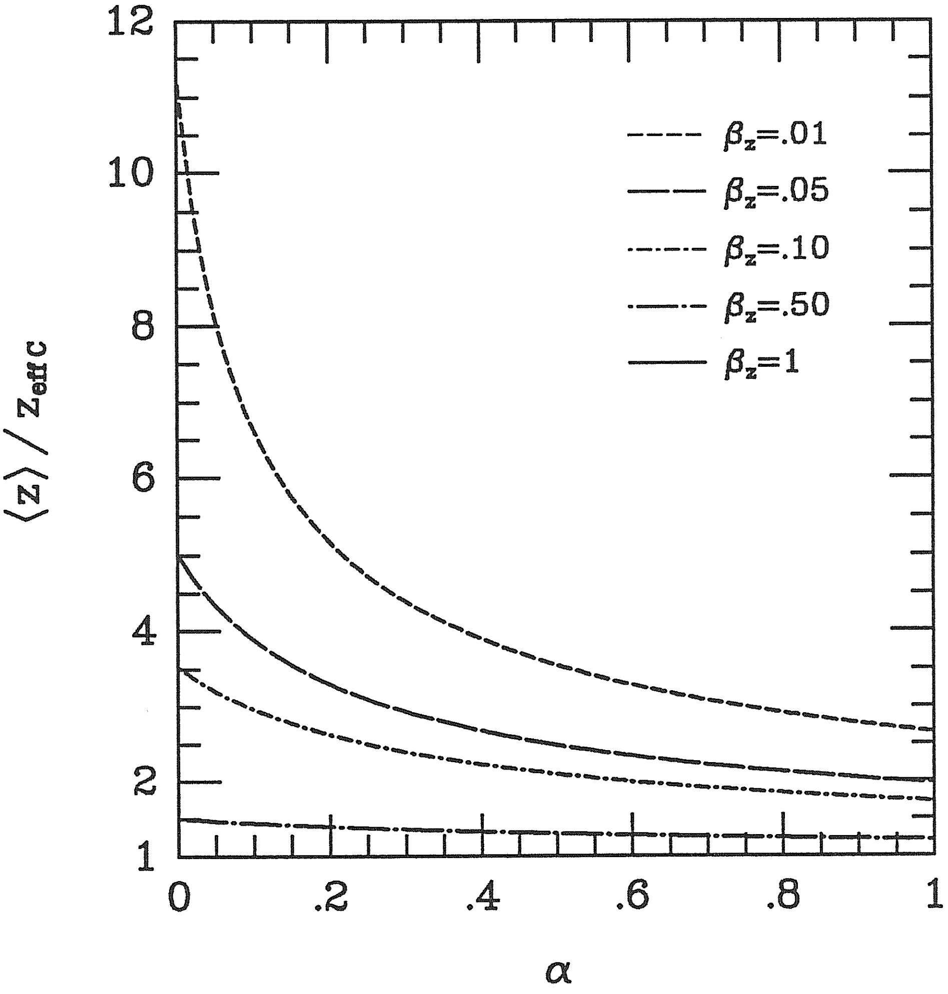


12
10
8
6
4
2
1
0
.2
.4
.6
.8
1
⟨z⟩ / z_eff C
β_z
α= .01
α= .10
α= .50
α=1.00
α→+∞

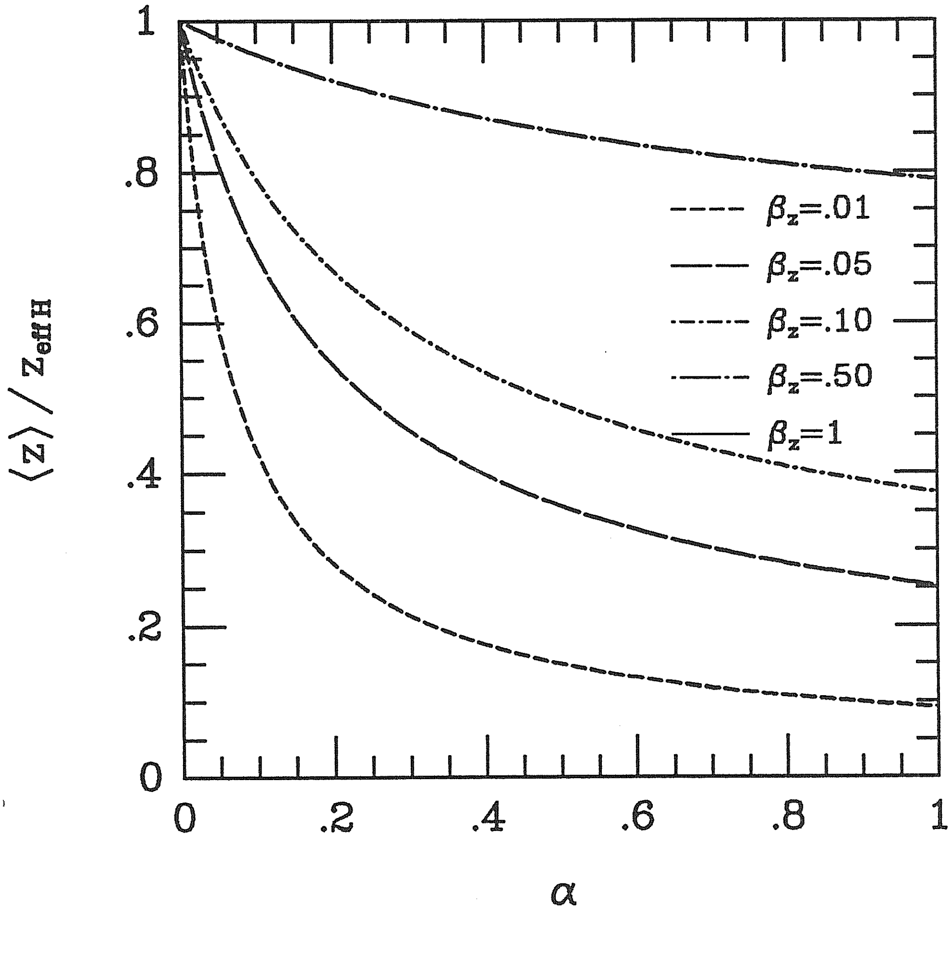


1
.8
.6
.4
.2
0
0
.2
.4
.6
.8
1
⟨z⟩ / z_eff H
β_z
α=0
α= .05
α= .10
α= .50
α=1.00

Figure 8.9

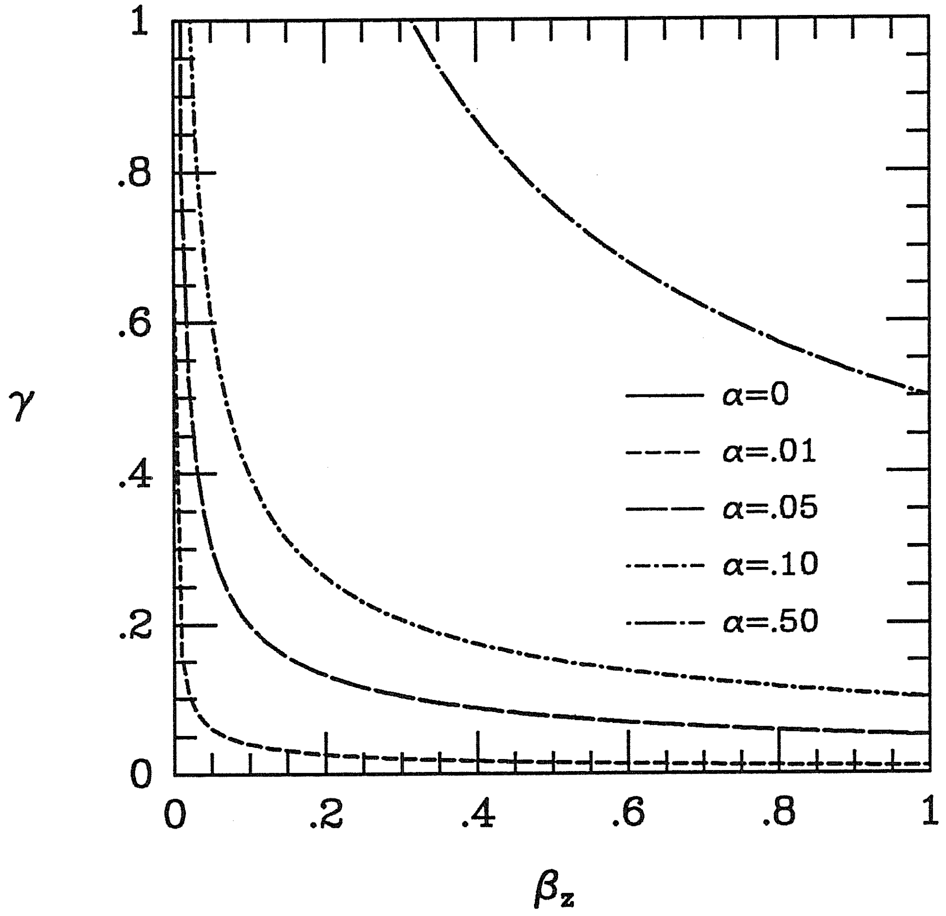


12
10
8
6
4
2
1
0
.2
.4
.6
.8
1
$\langle z \rangle / z_{eff\,C}$
α
β_z=.01
β_z=.05
β_z=.10
β_z=.50
β_z=1

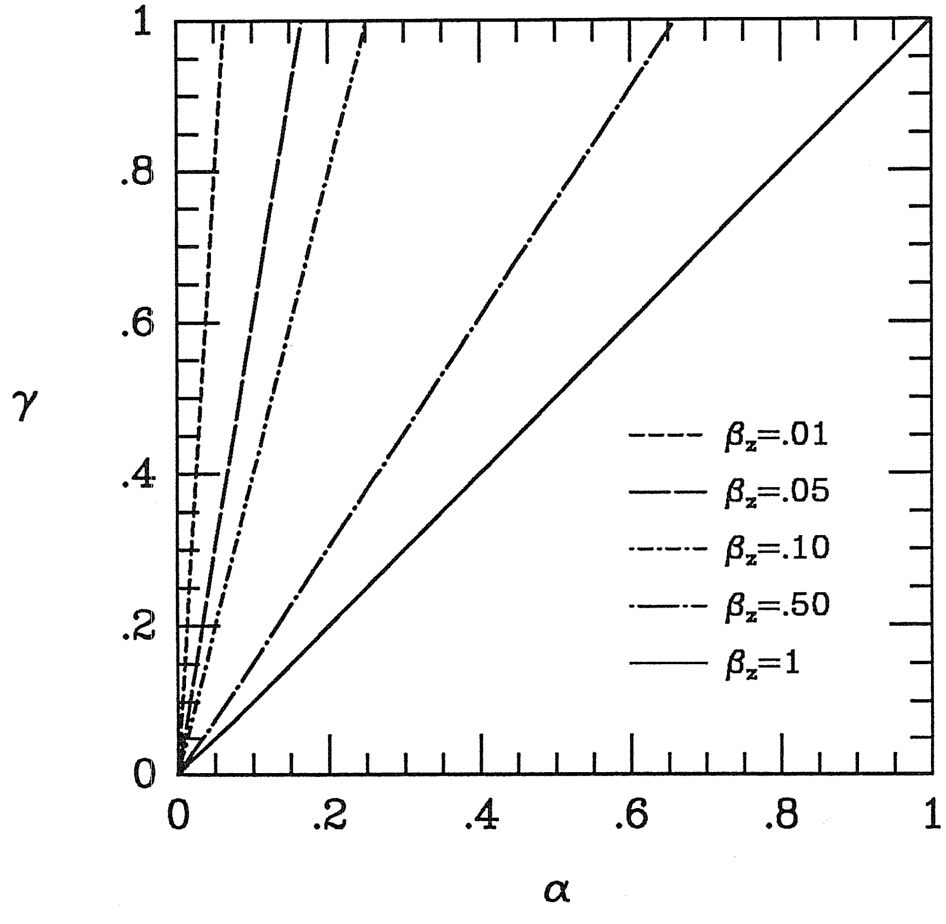


1
.8
.6
.4
.2
0
0
.2
.4
.6
.8
1
$\langle z \rangle / z_{eff\,H}$
α
β_z=.01
β_z=.05
β_z=.10
β_z=.50
β_z=1

Figure 8.10



1
.8
.6
.4
.2
0
γ
0
.2
.4
.6
.8
1
β_z
α=0
α=.01
α=.05
α=.10
α=.50



1
.8
.6
.4
.2
0
γ
0
.2
.4
.6
.8
1
α
β_z=.01
β_z=.05
β_z=.10
β_z=.50
β_z=1

Figure 8.11

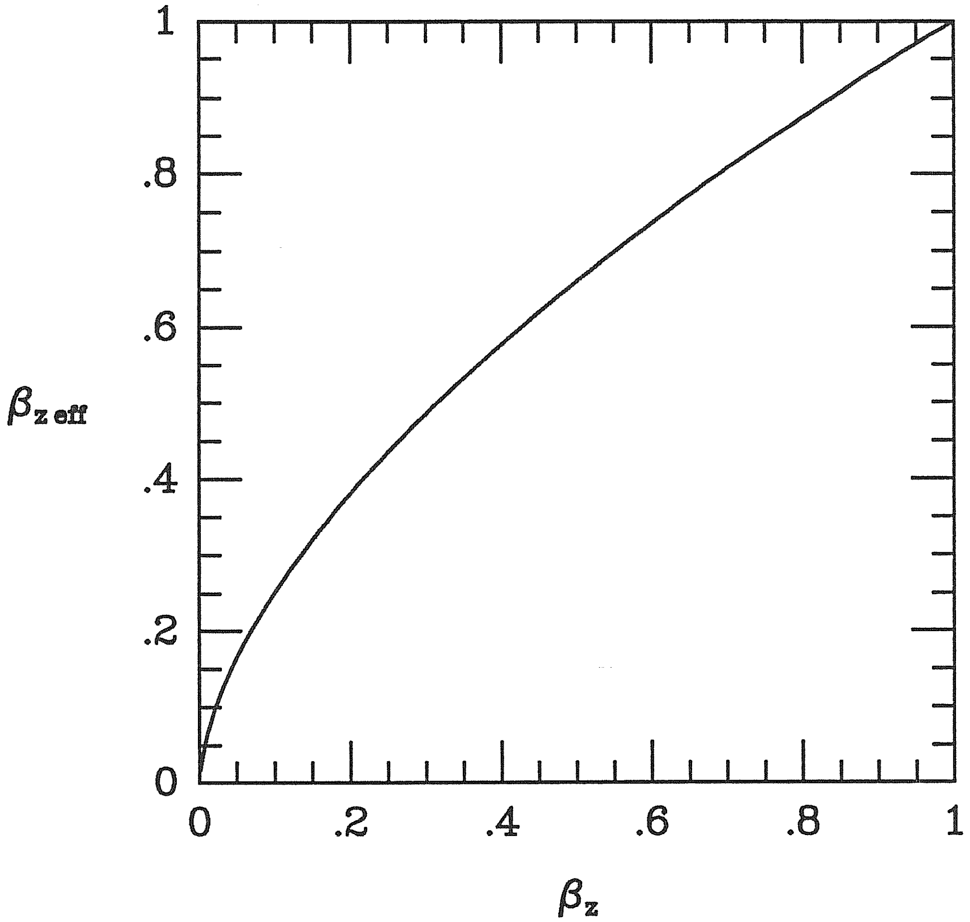
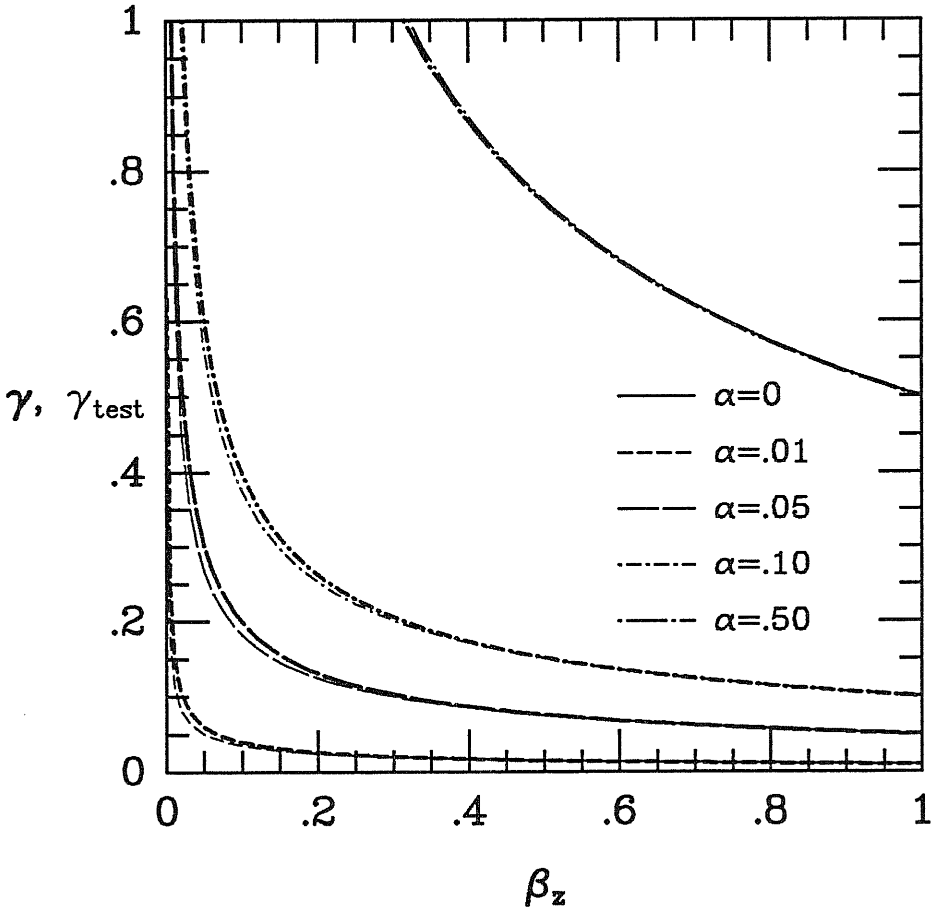


Figure 8.12


1
.8
.6
.4
.2
0
γ, γ_{test}
0
.2
.4
.6
.8
1
β_z
α=0
α=.01
α=.05
α=.10
α=.50

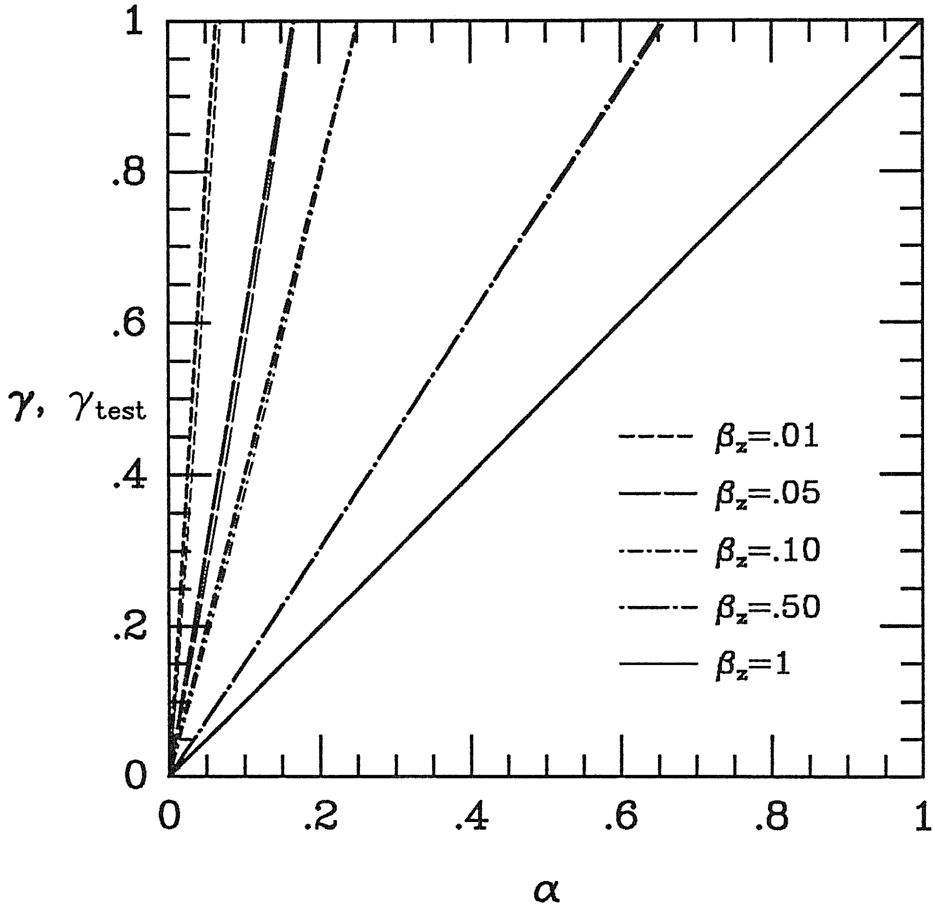
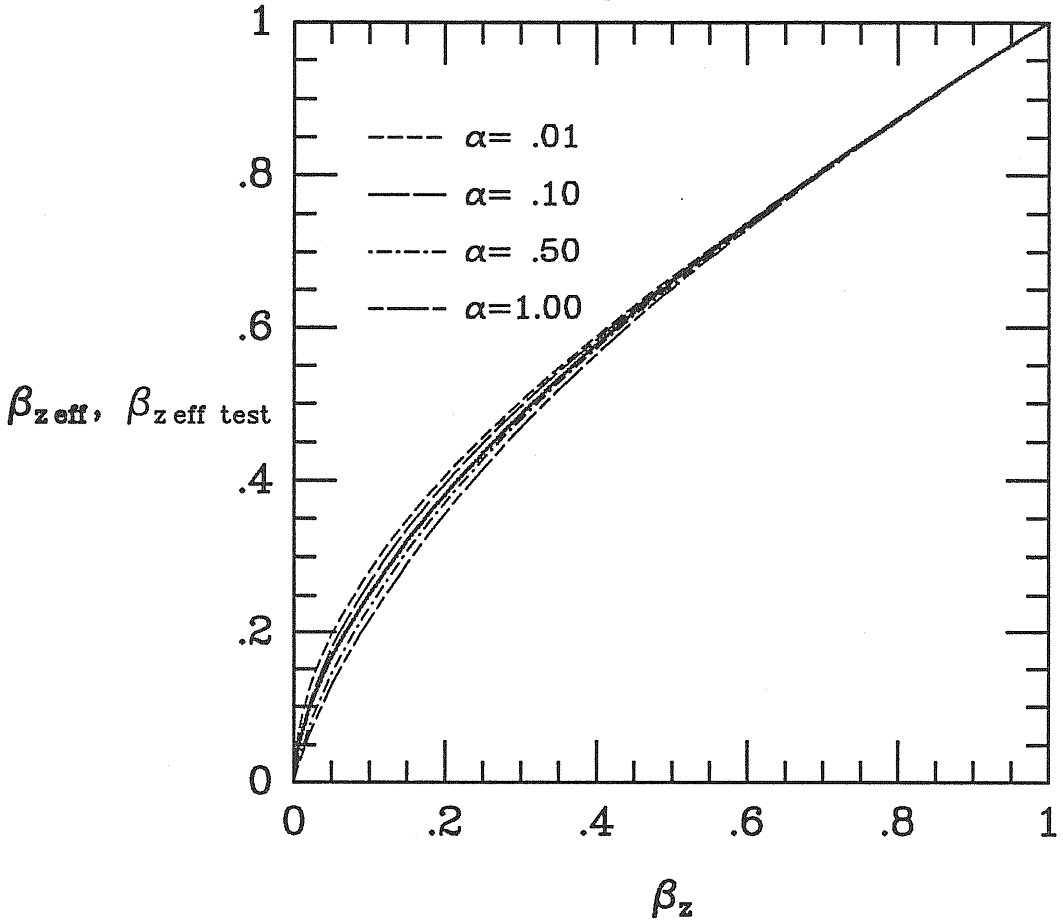
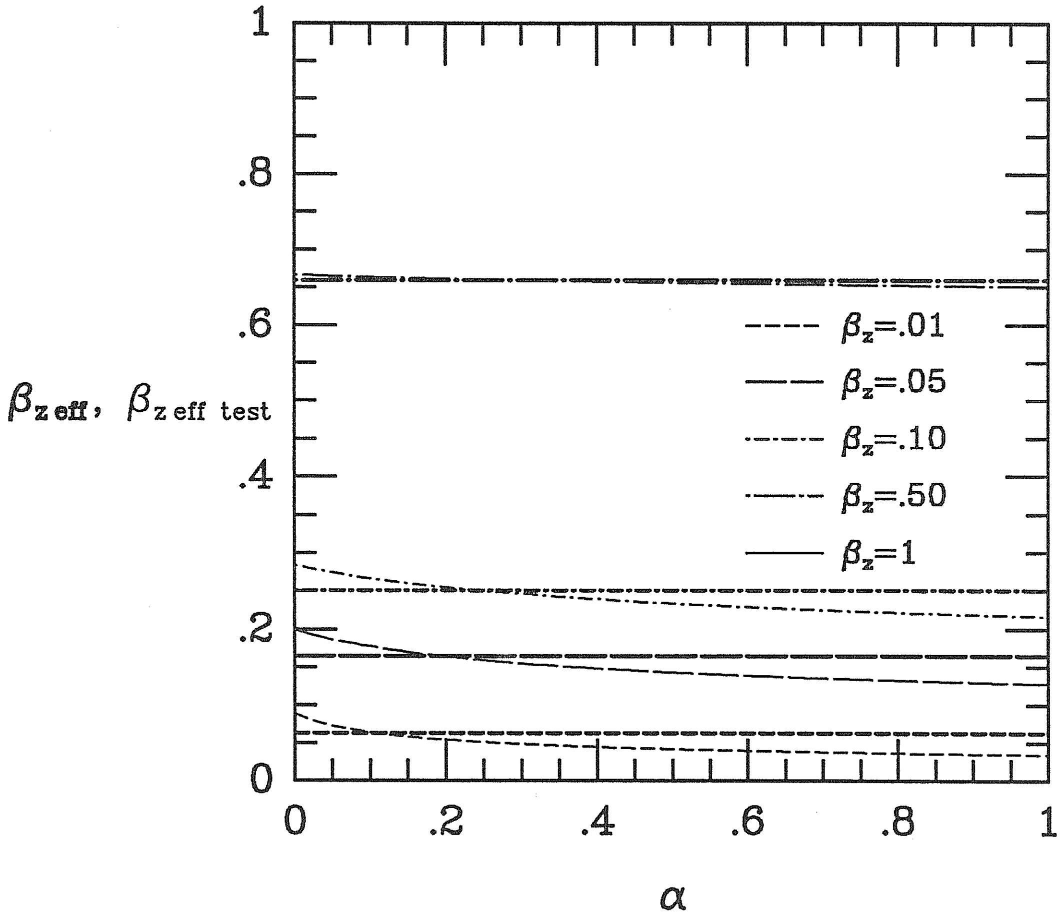

1
.8
.6
.4
.2
0
γ, γ_{test}
0
.2
.4
.6
.8
1
α
β_z=.01
β_z=.05
β_z=.10
β_z=.50
β_z=1

Figure 8.13

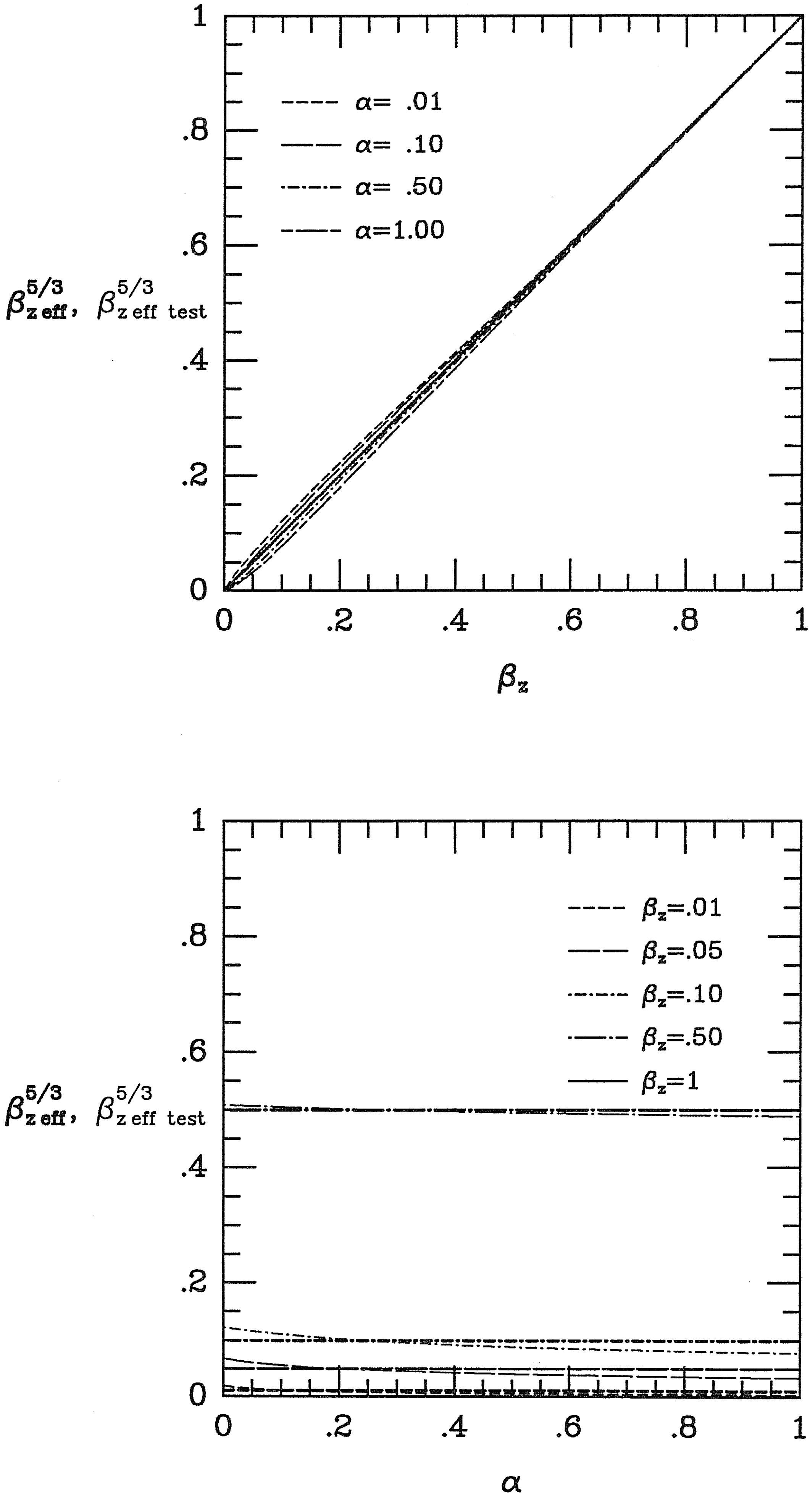


α= .01
α= .10
α= .50
α=1.00
$\beta_{z\,eff}$, $\beta_{z\,eff\ test}$
β_z
0
.2
.4
.6
.8
1



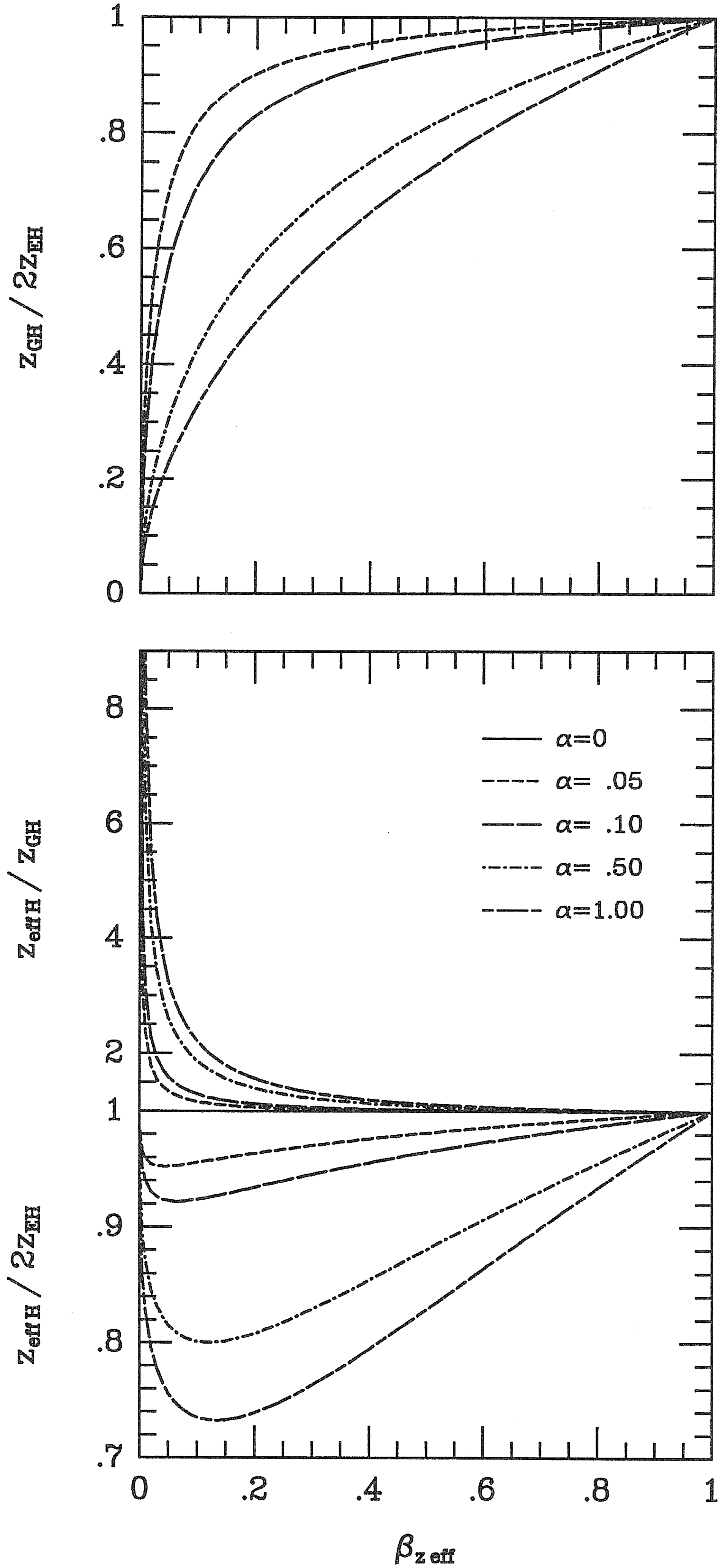
β_z=.01
β_z=.05
β_z=.10
β_z=.50
β_z=1
$\beta_{z\,eff}$, $\beta_{z\,eff\ test}$
α
0
.2
.4
.6
.8
1

Figure 8.13



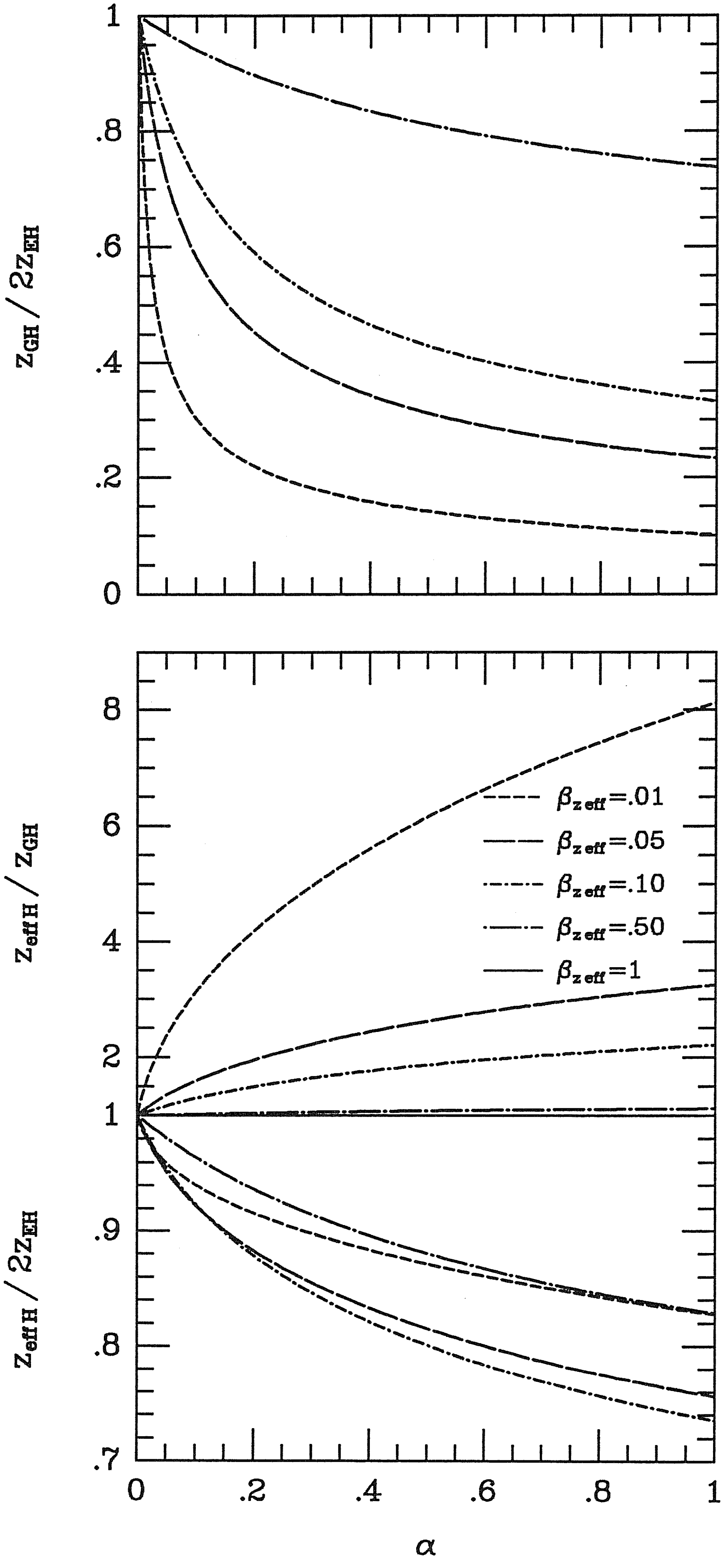
α= .01
α= .10
α= .50
α=1.00
$\beta_{z\,eff}^{5/3}$, $\beta_{z\,eff\ test}^{5/3}$
β_z
0
.2
.4
.6
.8
1
β_z=.01
β_z=.05
β_z=.10
β_z=.50
β_z=1
$\beta_{z\,eff}^{5/3}$, $\beta_{z\,eff\ test}^{5/3}$
α

Figure 8.14



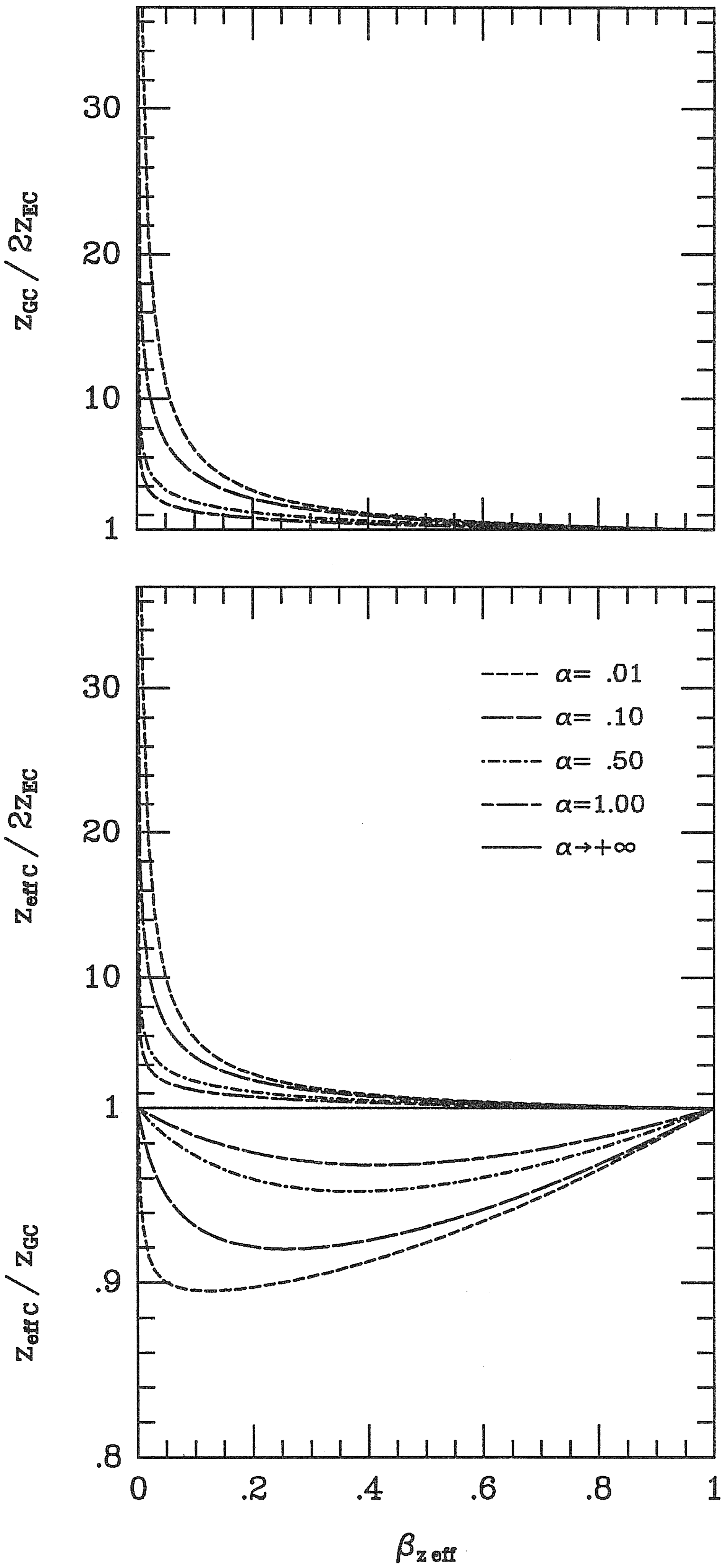
1
.8
.6
.4
.2
0
$z_{GH} / 2z_{EH}$
8
6
4
2
1
.9
.8
.7
$z_{eff\,H} / z_{GH}$
$z_{eff\,H} / 2z_{EH}$
α=0
α= .05
α= .10
α= .50
α=1.00
0
.2
.4
.6
.8
1
$\beta_{z\,eff}$

Figure 8.14



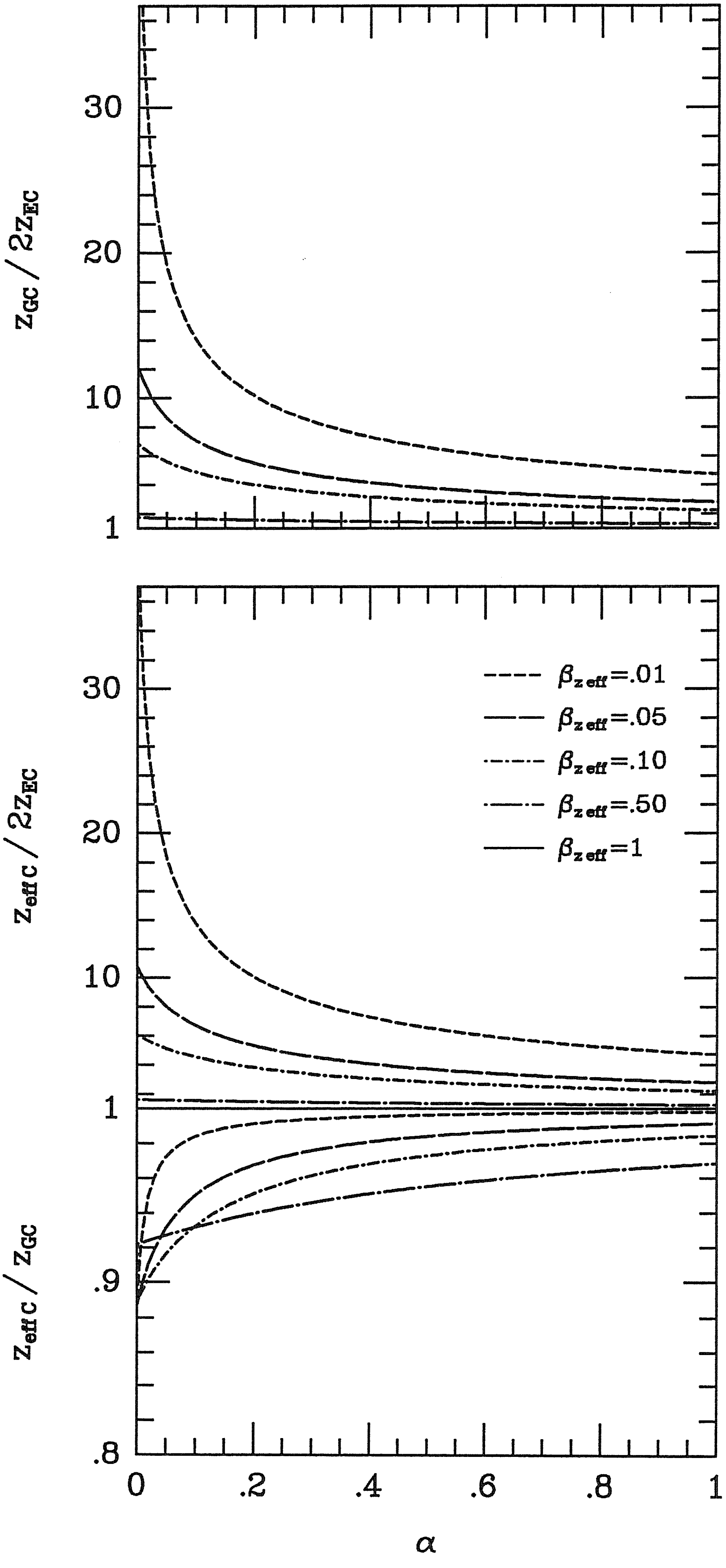
$z_{GH} / 2z_{EH}$
$z_{eff\,H} / z_{GH}$
$z_{eff\,H} / 2z_{EH}$
$\beta_{z\,eff}$=.01
$\beta_{z\,eff}$=.05
$\beta_{z\,eff}$=.10
$\beta_{z\,eff}$=.50
$\beta_{z\,eff}$=1
α

Figure 8.15



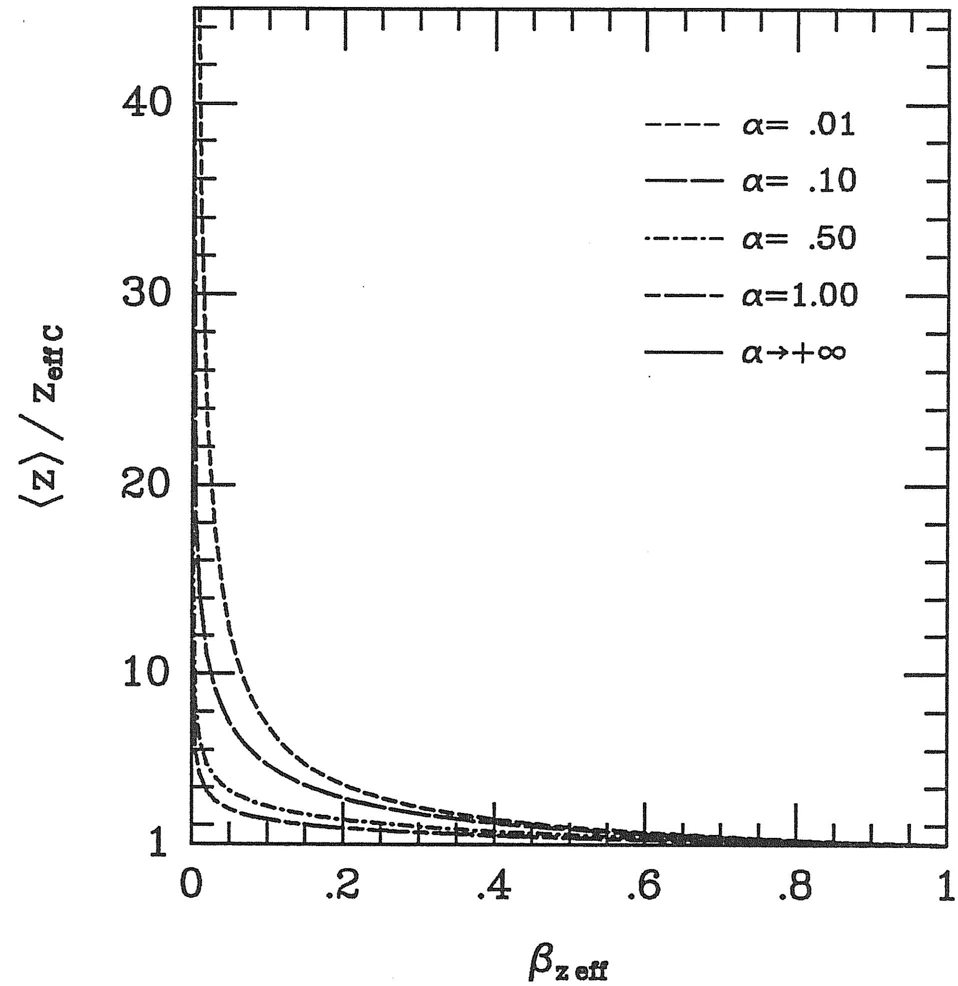
30
20
10
1
$z_{GC} / 2z_{EC}$
30
20
10
1
.9
.8
$z_{eff\,C} / 2z_{EC}$
$z_{eff\,C} / z_{GC}$
α= .01
α= .10
α= .50
α=1.00
α→+∞
0
.2
.4
.6
.8
1
$\beta_{z\,eff}$

Figure 8.15

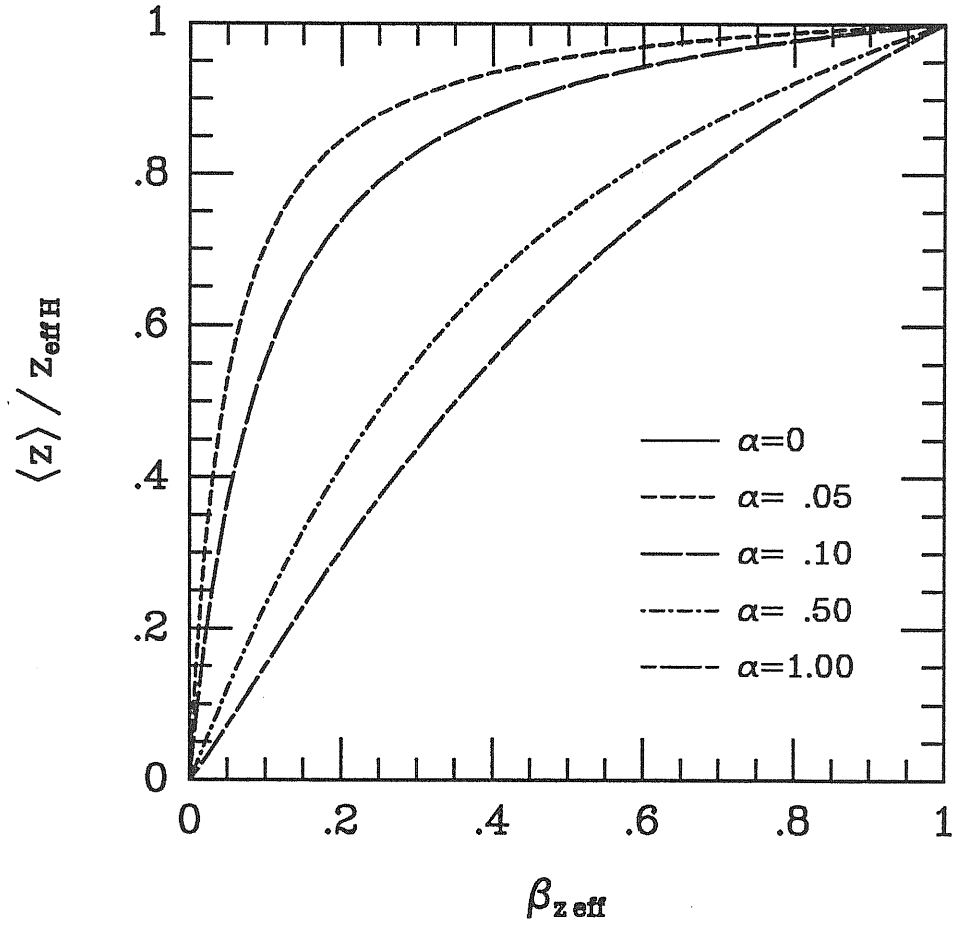


30
20
10
1
$z_{GC} / 2z_{EC}$
30
20
10
1
.9
.8
$z_{eff\,C} / 2z_{EC}$
$z_{eff\,C} / z_{GC}$
$\beta_{z\,eff}$=.01
$\beta_{z\,eff}$=.05
$\beta_{z\,eff}$=.10
$\beta_{z\,eff}$=.50
$\beta_{z\,eff}$=1
0
.2
.4
.6
.8
1
α

Figure 8.16

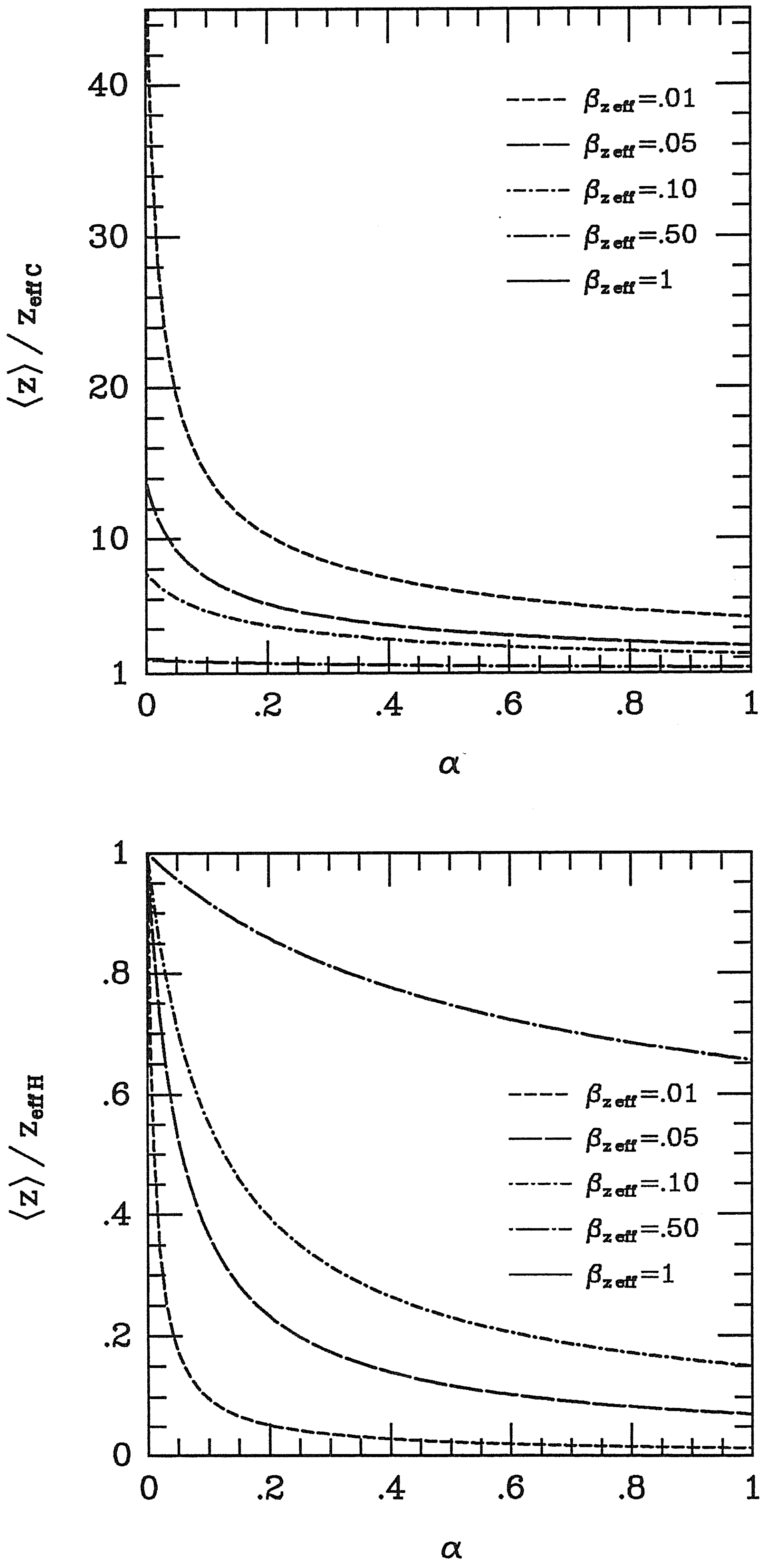


40
30
20
10
1
0
.2
.4
.6
.8
1
⟨z⟩ / z_eff C
β_z eff
α= .01
α= .10
α= .50
α=1.00
α→+∞



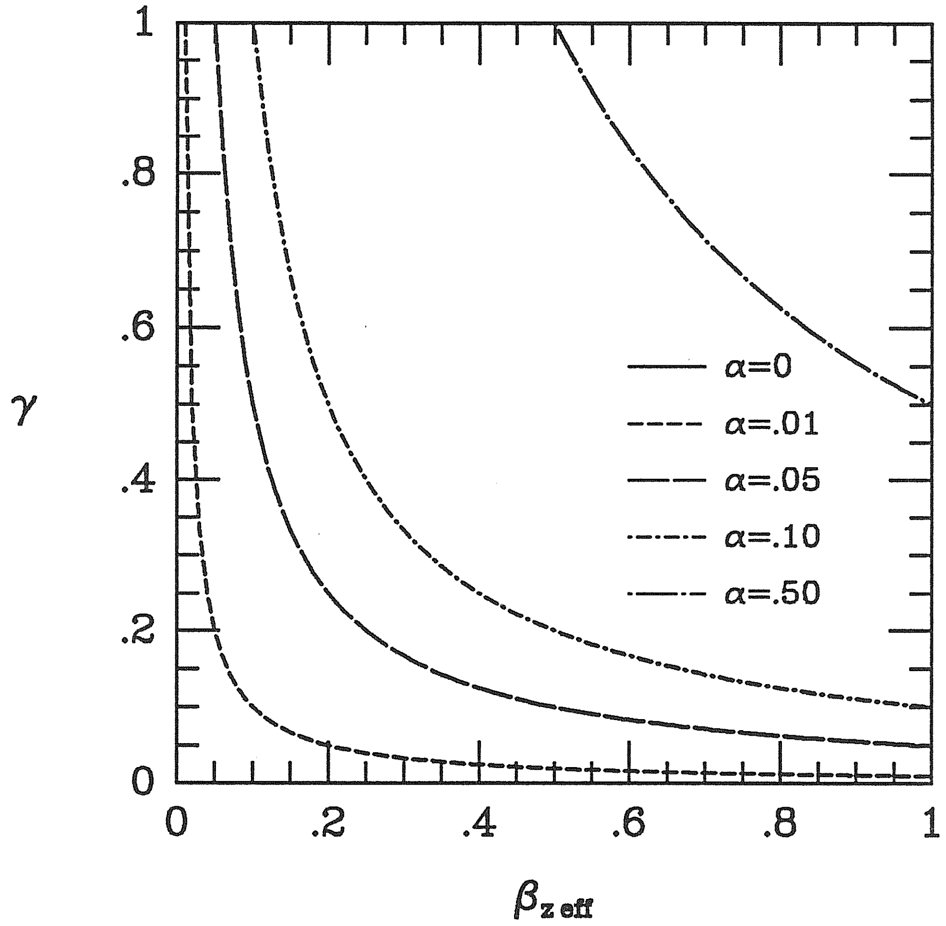
1
.8
.6
.4
.2
0
0
.2
.4
.6
.8
1
⟨z⟩ / z_eff H
β_z eff
α=0
α= .05
α= .10
α= .50
α=1.00

Figure 8.16

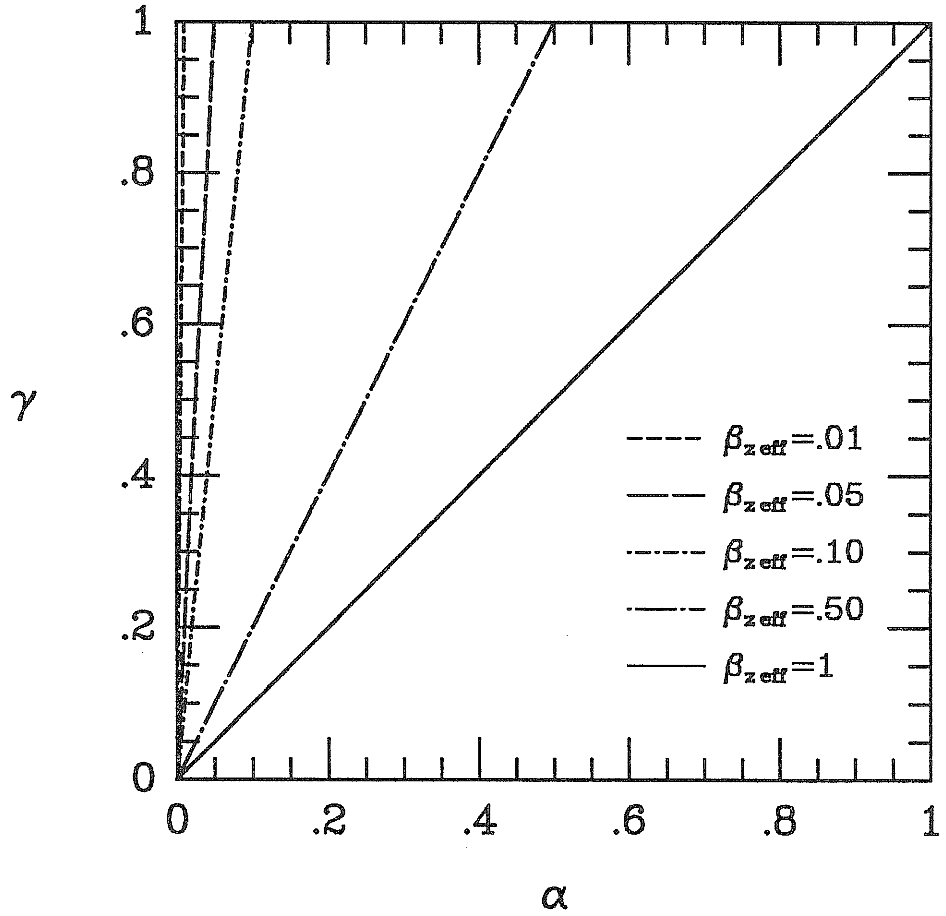


$\beta_{z\,eff}$=.01
$\beta_{z\,eff}$=.05
$\beta_{z\,eff}$=.10
$\beta_{z\,eff}$=.50
$\beta_{z\,eff}=1$
$\langle z \rangle / z_{eff\,C}$
α
$\beta_{z\,eff}$=.01
$\beta_{z\,eff}$=.05
$\beta_{z\,eff}$=.10
$\beta_{z\,eff}$=.50
$\beta_{z\,eff}=1$
$\langle z \rangle / z_{eff\,H}$
α

Figure 8.17



1
.8
.6
.4
.2
0
0
.2
.4
.6
.8
1
γ
$\beta_{z\,eff}$
α=0
α=.01
α=.05
α=.10
α=.50



1
.8
.6
.4
.2
0
0
.2
.4
.6
.8
1
γ
α
$\beta_{z\,eff}$=.01
$\beta_{z\,eff}$=.05
$\beta_{z\,eff}$=.10
$\beta_{z\,eff}$=.50
$\beta_{z\,eff}$=1

Figure 8.18

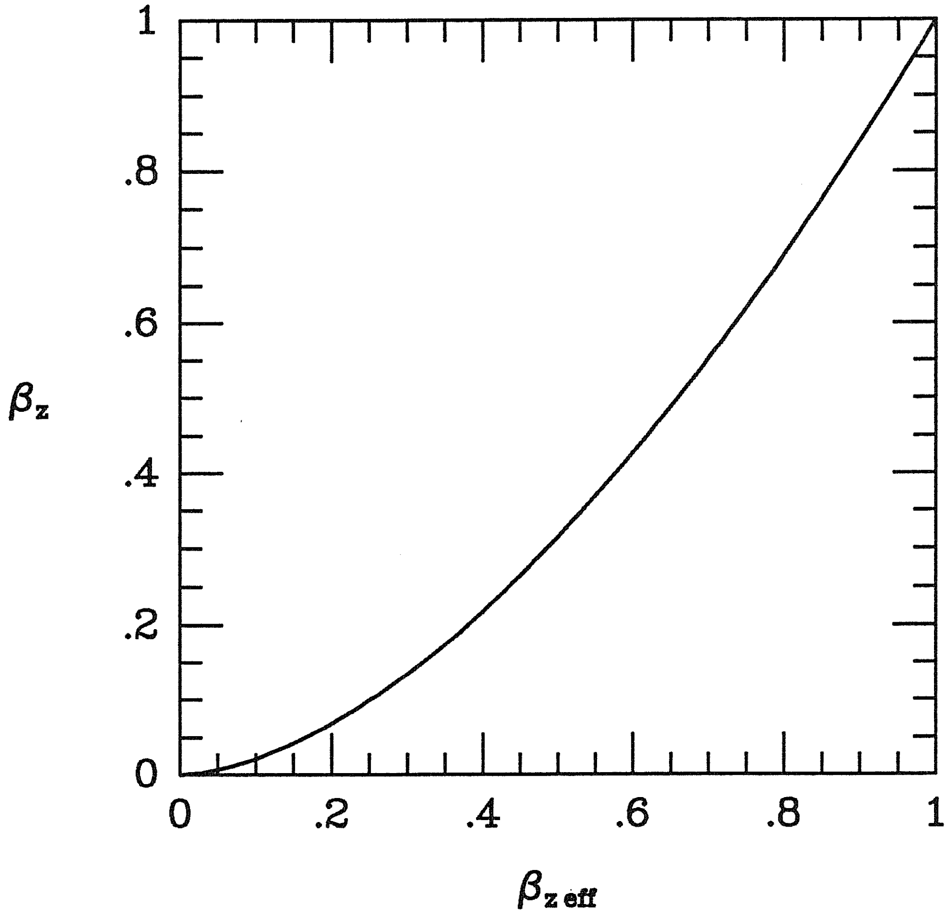
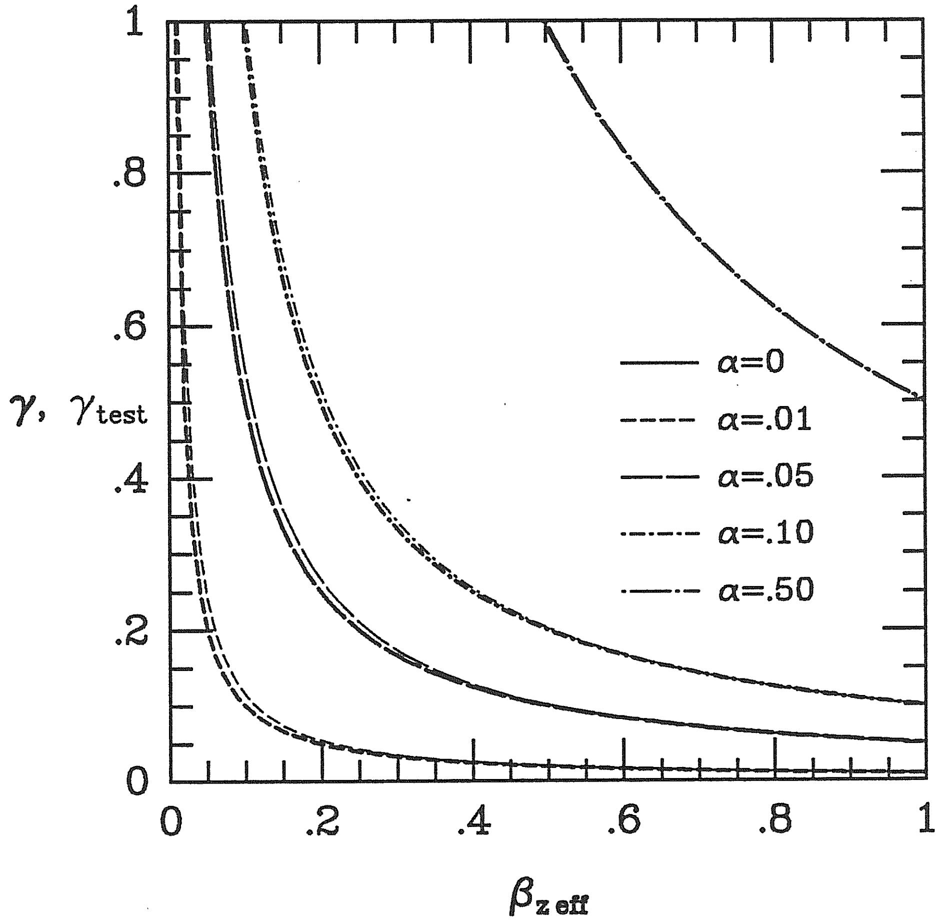
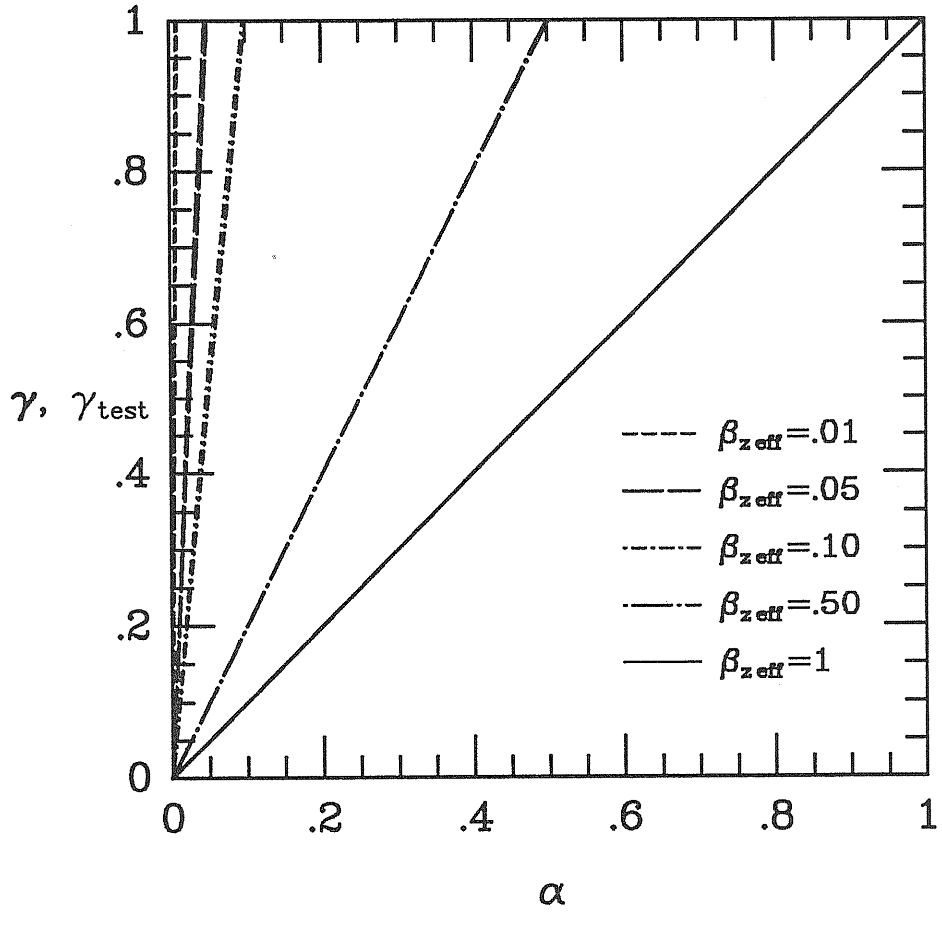


Figure 8.19

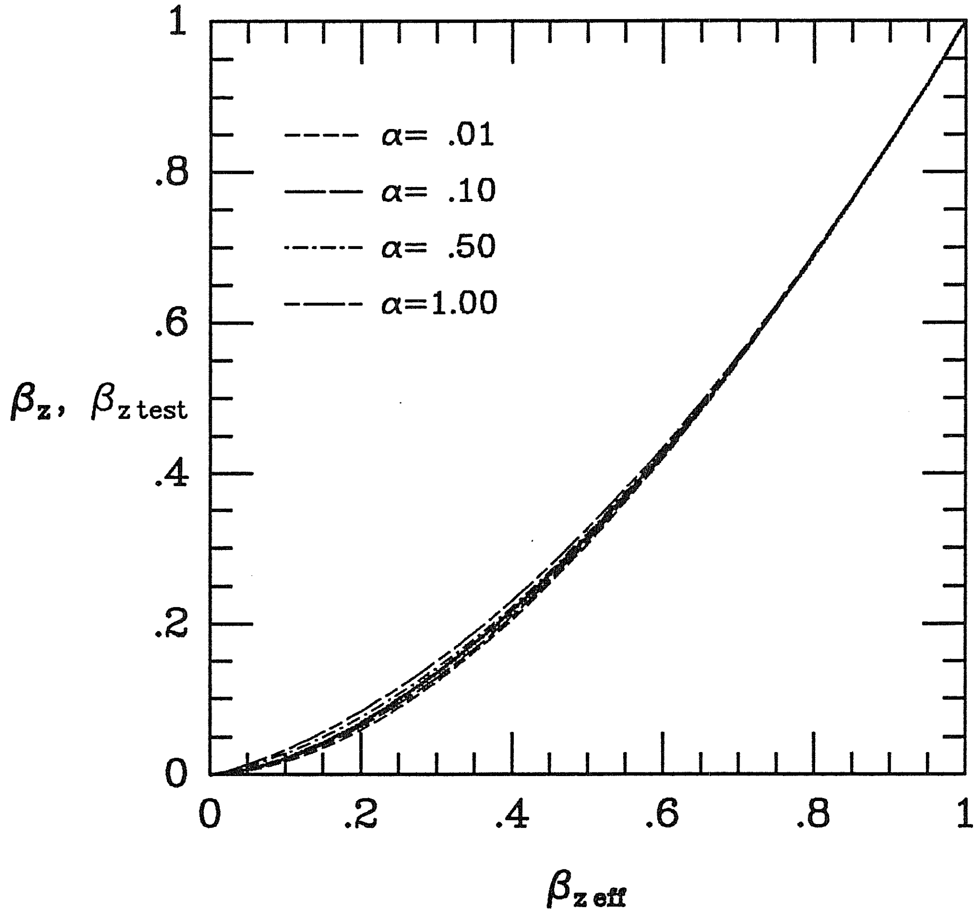


1
.8
.6
.4
.2
0
γ, γ_{test}
0
.2
.4
.6
.8
1
$\beta_{z\,eff}$
α=0
α=.01
α=.05
α=.10
α=.50

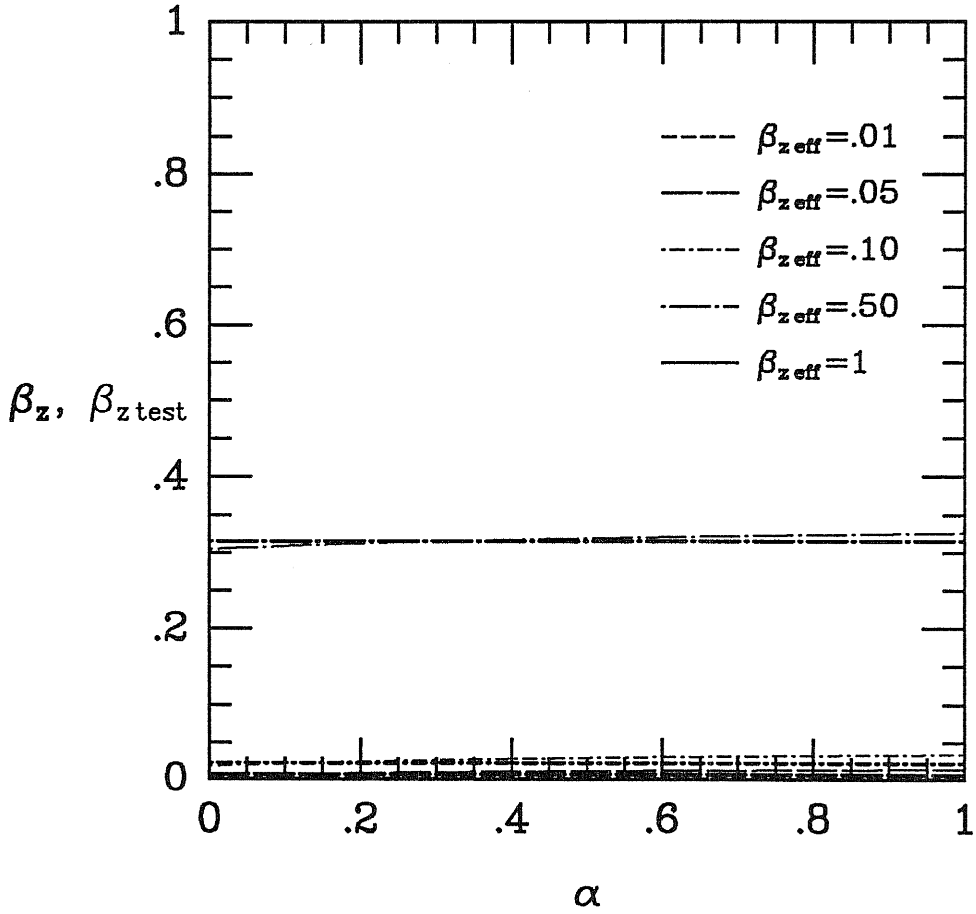


1
.8
.6
.4
.2
0
γ, γ_{test}
0
.2
.4
.6
.8
1
α
$\beta_{z\,eff}$=.01
$\beta_{z\,eff}$=.05
$\beta_{z\,eff}$=.10
$\beta_{z\,eff}$=.50
$\beta_{z\,eff}$=1

Figure 8.20

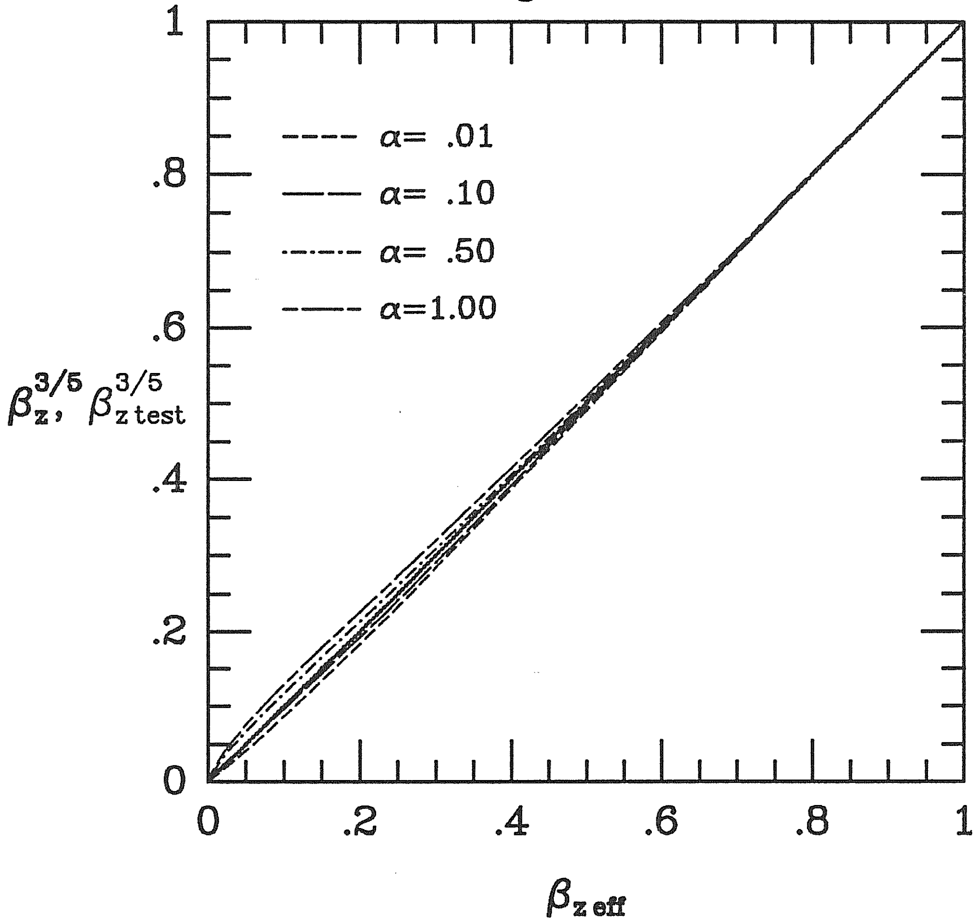


1
.8
.6
.4
.2
0
0
.2
.4
.6
.8
1
α= .01
α= .10
α= .50
α=1.00
βz, βz test
βz eff

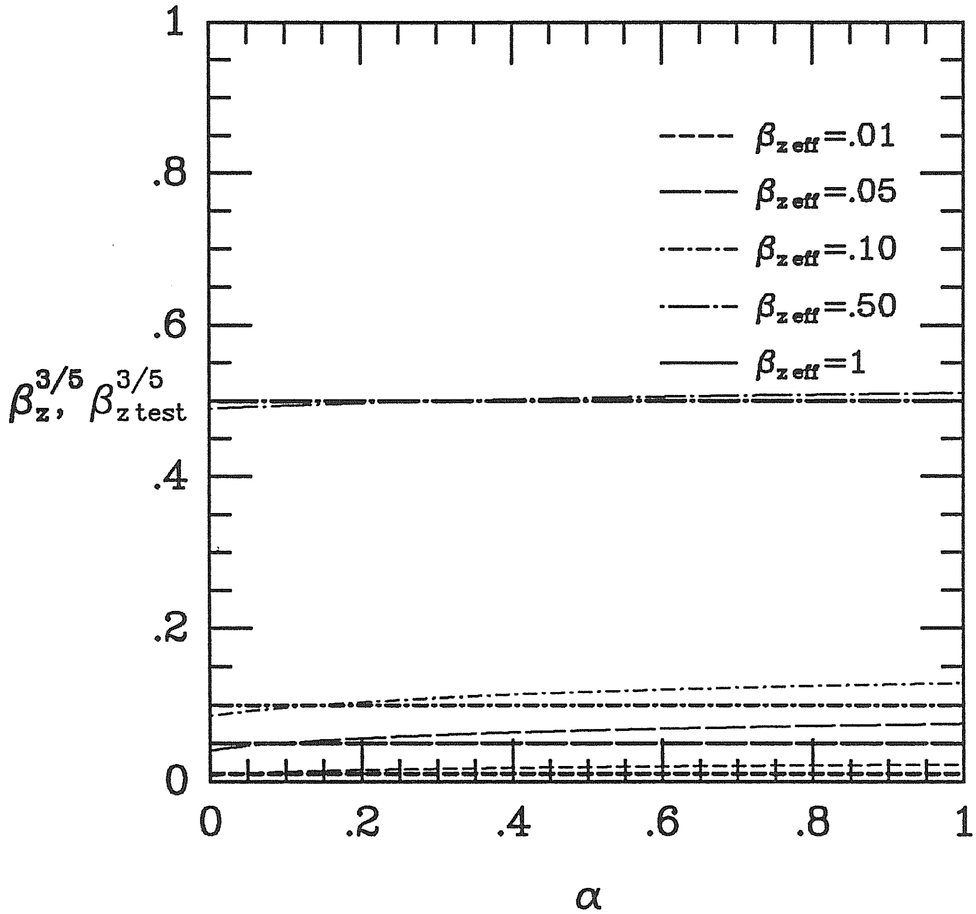


1
.8
.6
.4
.2
0
0
.2
.4
.6
.8
1
βz eff=.01
βz eff=.05
βz eff=.10
βz eff=.50
βz eff=1
βz, βz test
α

Figure 8.20



1
.8
.6
.4
.2
0
0
.2
.4
.6
.8
1
α= .01
α= .10
α= .50
α=1.00
$\beta_z^{3/5}, \beta_{z\,test}^{3/5}$
$\beta_{z\,eff}$



1
.8
.6
.4
.2
0
0
.2
.4
.6
.8
1
$\beta_{z\,eff}$=.01
$\beta_{z\,eff}$=.05
$\beta_{z\,eff}$=.10
$\beta_{z\,eff}$=.50
$\beta_{z\,eff}$=1
$\beta_z^{3/5}, \beta_{z\,test}^{3/5}$
α

Chapter 9

The Cold Interstellar Gas and the Stability of Galactic Discs: Local Finite-Thickness Effects

Summary

The results obtained in Chapter 8 as regards the vertical structure at equilibrium of two-component galactic discs are used to investigate their local linear stability properties. Under reasonable assumptions finite-thickness corrections to the local dispersion relation can be expressed in terms of two reduction factors lowering the response of the two components or, equivalently, their equilibrium surface densities. Different ansatz for such reduction factors, justified by extending the analysis performed by Vandervoort (1970a) for one-component purely stellar discs, are compared by studying the corresponding two-fluid marginal stability curves in standard star-dominated and peculiar gas-dominated regimes. It is found that the stabilizing role of finite-thickness effects can partially counterbalance the destabilizing role of the cold interstellar gas in linear regimes.

9.1 Introduction

Several attempts have already been made to estimate finite-thickness corrections to the local dispersion relation in one-component galactic discs (e.g., Sweet and McGregor 1964; Toomre 1964, 1974; Goldreich and Lynden-Bell

1965a,b; Vandervoort 1970a; Genkin and Safronov 1975; Bertin and Casertano 1982; Yue 1982a,b,c; Balbus and Cowie 1985; Morozov and Khoperskov 1986; Peng 1988; Fridman 1989; see also Fridman and Polyachenko 1984 for a review). While it is generally agreed that the form of the dispersion relation remains the same provided the unperturbed surface density is multiplied by a suitable reduction factor, different estimates of such reduction factor have been given by the various authors.

The most reliable and complete analysis is that performed by Vandervoort (1970a), which is local in the galactic plane and global perpendicularly to it. The reduction factor, found by solving an eigenvalue problem, has been shown to be very well approximated by the simple expression

$$\mathfrak{T} = \frac{1}{1 + |k|\langle z\rangle}, \quad |k|\langle z\rangle = O(1), \tag{9.1}$$

where k is the local radial wave-number of the perturbation and $\langle z\rangle$ is the thickness-scale of the galactic disc. This estimate should be compared to that naively obtained by Toomre (1964):

$$\mathfrak{T} = \frac{1 - e^{-|k|\langle z\rangle}}{|k|\langle z\rangle}, \quad |k|\langle z\rangle \ll 1. \tag{9.2}$$

The aim of our calculations is just to extend the rigorous partially global analysis carried out by Vandervoort (1970a) in such a way as to include the cold interstellar gas as well.

The investigation performed by Yue (1982a,b,c) is also of considerable interest, because it represents an attempt to take finite-thickness effects into account at a fully global level. As a compromise of this extension, the k-dependent reduction factor has been approximated by replacing the wave-number with its characteristic local value corresponding to the maximum of the marginal stability curve.

As regards more realistic models of galactic discs in which more (than one) components are present, no so rigorous partially global analysis has been performed (e.g., Shu 1968; see also: Lin 1970, Lin and Shu 1971; Vandervoort 1970b; Nakamura 1978; Jog and Solomon 1984a,b). Among these attempts the contribution of Shu (1968) is surely the most important one, while Vandervoort (1970b) refers only to a particular continuous model of stellar populations without performing a proper stability analysis.

The basis of our investigation, as already mentioned, is the stability analysis performed by Vandervoort (1970a) in the case of one-component purely stellar discs. The method we have employed is in fact a straightforward extension of that developed by Vandervoort (1970a) to the case in which two components are present (the consideration of a fluid component does not give rise to difficulties), but the resulting analysis is indeed much more complicated than in the one-component case (Romeo 1987, 1990a,b). For this reason we shall briefly discuss the one-component case before.

9.2 One-Component Case

9.2.1 Derivation of the Reduction Factor

The working assumptions made by Vandervoort (1970a) are the same as those used in the local kinetic formulation of the spiral structure theory (see Section 6.3), except that concerning the infinitesimal thickness of the models.

The method employed to solve the Vlasov equation when z-motions are taken into account is based on the existence of an adiabatic invariant J_z, whose approximate constancy characterizes the vertical motion of disc stars. This is a characteristic of highly flattened galactic discs, where the frequency of the oscillation in the z-direction is large compared to the frequency of the epicyclic motion in the symmetry plane, which in turn is generally of the same order as the pattern frequency of spiral waves. To the non-trivial lowest order of approximation, assuming an unperturbed distribution function of modified Schwarzschild type, it is found that the perturbation induced in the volume density is

$$\begin{aligned} \rho_1 \;=\; & -\frac{\rho}{2\pi G\rho_0\Delta^2}\,\frac{1}{\sqrt{2\pi c_z^2}}\int_{-\infty}^{+\infty}(\Phi_1-\langle\Phi_1\rangle)\,\exp\left(-\frac{v_z^2}{2c_z^2}\right)dv_z \\ & -\frac{\rho|k|}{4\pi G\rho_0\Delta}\,\frac{1}{\sqrt{2\pi c_z^2}}\int_{-\infty}^{+\infty}\mathcal{D}\langle\Phi_1\rangle\,\exp\left(-\frac{v_z^2}{2c_z^2}\right)dv_z\,, \end{aligned} \qquad (9.3)$$

where we have adopted the same notations as in Sections 6.3 and 8.2, and $\langle\ldots\rangle$ denotes the average over the angles conjugated to the action J_z. Moreover,

$$\mathcal{D}\equiv\frac{4\pi G\rho_0\Delta|k|}{\kappa^2-(\omega-m\Omega)^2}\,F_{\mathrm{kin}\,\nu}(x_{\mathrm{kin}})=\frac{2\pi G\sigma|k|}{\kappa^2-(\omega-m\Omega)^2}\,F_{\mathrm{kin}\,\nu}(x_{\mathrm{kin}})\,; \qquad (9.4)$$

$$F_{\mathrm{kin}\,\nu}(x_{\mathrm{kin}})\equiv\frac{1-\nu^2}{x_{\mathrm{kin}}}\left[1-\frac{\nu e^{-x_{\mathrm{kin}}}}{2\,\sin(\nu\pi)}\int_{-\pi}^{+\pi}e^{-x_{\mathrm{kin}}\cos p}\cos(\nu p)\,dp\right]; \qquad (9.5)$$

$$\nu\equiv\frac{\omega-m\Omega}{\kappa}\,,\qquad x_{\mathrm{kin}}\equiv k^2\,\frac{c_r^2}{\kappa^2}\,. \qquad (9.6)$$

In these formulae, $\mathcal{D}=1$ is just the uncorrected local dispersion relation, $F_{\mathrm{kin}\,\nu}$ is a reduction factor which lowers the response of high radial velocity dispersion stars, and ν is the dimensionless Doppler-shifted frequency of the spiral wave.

In considering the Poisson equation, we shall assume that the following ordering holds:

$$\frac{\Delta}{\langle r\rangle} \ll 1\,, \quad \frac{m}{|k|r} \sim \frac{c_r}{r\kappa} \ll 1\,, \quad |k|\Delta = O(1)\,. \tag{9.7}$$

Self-consistency requires

$$\begin{aligned} \frac{\partial^2 \Phi_1}{\partial z^2} - k^2\Phi_1 &= \frac{2\rho}{\rho_0\Delta^2}\left[\frac{1}{\sqrt{2\pi c_z^2}}\int_{-\infty}^{+\infty}\langle\Phi_1\rangle \exp\left(-\frac{v_z^2}{2c_z^2}\right)dv_z - \Phi_1\right] \\ &\quad - \frac{\Lambda\rho}{\rho_0\Delta^2}\frac{1}{\sqrt{2\pi c_z^2}}\int_{-\infty}^{+\infty}\langle\Phi_1\rangle \exp\left(-\frac{v_z^2}{2c_z^2}\right)dv_z\,, \\ &\quad \text{where} \quad \Lambda \equiv |k|\Delta\mathcal{D}\,, \end{aligned} \tag{9.8}$$

to the same order of approximation. This wave equation is to be solved with the boundary conditions

$$\lim_{|z|\to+\infty} \Phi_1(r,\theta,z;t) = 0\,. \tag{9.9}$$

This complicated eigenvalue problem for Λ can be reduced to a simpler eigenvalue problem, by observing that the effective perturbation induced in the volume density

$$\rho_1^* \equiv -\frac{\rho|k|}{4\pi G\rho_0\Delta}\mathcal{D}\Phi_1 \tag{9.10}$$

gives rise to the same perturbation induced in the surface density as ρ_1 to the order of approximation to which we are working. Thus, a reasonable approximation to the eigenvalue problem for Λ can be obtained by replacing ρ_1 with ρ_1^*:

$$\frac{\partial^2\Phi_1}{\partial z^2} - k^2\Phi_1 = -\frac{\Lambda}{\Delta^2}\operatorname{sech}^2\left(\frac{z}{\Delta}\right)\Phi_1\,, \quad \lim_{|z|\to+\infty}\Phi_1(r,\theta,z;t) = 0\,, \tag{9.11}$$

which is a Schrödinger-type wave equation. Note, however, that this eigenvalue problem differs from the usual quantum-mechanical problem, for it consists in fixing the energy of the particle (in a bound state) and in seeking the depths of the potential well for which that energy is allowed. Nevertheless, the solution can be obtained along a line similar to the quantum-mechanical case (see, e.g., Landau and Lifshitz 1977).

The corresponding quantum condition is (cf. Vandervoort 1970a)

$$\Lambda_n = (n + |k|\Delta)(n + |k|\Delta + 1) \quad (n \in \mathbb{N})\,. \tag{9.12}$$

Note that only the lowest eigenvalue is physically relevant, because all the others do not vanish as $|k|\Delta \to 0$ in such a way as to recover the equation valid for infinitesimally thin discs, which is just equation (9.11) with the right-hand side replaced by the term $4\pi G\sigma_1\,\delta(z)$. Therefore, imposing the restricted quantum condition

$$\Lambda_0 = |k|\Delta\,(1+|k|\Delta) \quad \Longrightarrow \quad \mathcal{D}_0 = 1+|k|\Delta \tag{9.13}$$

and taking equation (9.4) into account, we find that finite-thickness corrections to the local dispersion relation correspond to a weakening in the response of disc stars, which can equivalently be thought of as a lowering in their equilibrium surface density by the reduction factor

$$\mathfrak{T} \equiv \frac{\sigma_{\mathrm{red}}}{\sigma} = \frac{1}{1+|k|\Delta} = \frac{1}{1+|k|\langle z\rangle}\,. \tag{9.14}$$

Therefore, the reduced surface density σ_{red} and not σ is relevant to the local stability of one-component purely stellar discs. From the form of the reduction factor it is apparent that such a simple result does no longer hold when a global (in the galactic plane) analysis is performed.

9.2.2 Local Stability

The analysis performed by Vandervoort (1970a), strictly speaking, refers to a single stellar component described in the framework of a kinetic approach. However, to the order of approximation to which we are working, equation (9.10) still holds in the case of a single fluid component and the same results presented above apply, provided the kinetic reduction factor $F_{\mathrm{kin}\,\nu}$ defined in equation (9.5) is replaced by the fluid reduction factor

$$F_{\mathrm{fl}\,\nu}(x_{\mathrm{fl}}) \equiv \frac{1}{1+(x_{\mathrm{fl}}/1-\nu^2)}\,, \tag{9.15}$$

whose dimensionless argument involves the same scaling defined in equation (9.6) with the radial velocity dispersion c_r replaced by the equivalent planar acoustic speed c:

$$x_{\mathrm{fl}} \equiv k^2\,\frac{c^2}{\kappa^2}\,. \tag{9.16}$$

Taking this fact into account, in what follows we shall adopt the usual fluid description, which is more convenient for investigating the stability properties of galactic discs analytically and reproduces the main results almost completely except near the Lindblad resonances.

Taking equations (8.16) and (8.17) into account, the reduction factor $\mathfrak{T}$ derived in the previous subsection can conveniently be rewritten in the dimensionless form

$$\mathfrak{T}^{-1} = 1+\frac{Q^2}{2\overline{\lambda}}\,\delta\,, \tag{9.17}$$

where we have adopted the following scaling and parametrization:

$$\bar{\lambda} \equiv \frac{k_{\mathrm{T}}}{|k|}\,, \quad \text{where} \quad k_{\mathrm{T}} \equiv \frac{\kappa^2}{2\pi G\sigma}\,; \tag{9.18}$$

$$\delta \equiv \frac{c_z^2}{c^2} \quad (0 < \delta < 1)\,; \tag{9.19}$$

$$Q \equiv \frac{c\kappa}{\pi G\sigma}\,, \tag{9.20}$$

Q being the well-known local stability parameter. The case $\delta = 0$ represents the limit of an infinitesimally thin system. The marginal stability curve in the $(\bar{\lambda}, Q^2)$ plane can easily be derived from its uncorrected expression, by applying the reduction factor to each dimensionless quantity involving the surface density:

$$\frac{Q^2}{\mathcal{T}^2} = 4\frac{\bar{\lambda}}{\mathcal{T}}\left(1 - \frac{\bar{\lambda}}{\mathcal{T}}\right) \quad \Longrightarrow$$

$$Q^2 = \left(\frac{\bar{\lambda}}{\delta}\right) \cdot \left[-(2\bar{\lambda}\delta + 1) + \sqrt{4\bar{\lambda}^2\delta^2 - 4\bar{\lambda}\delta + (1 + 8\delta)}\right]\,, \tag{9.21}$$

which we consider in the range $0 \le \bar{\lambda} \le 1$.

Alternatively, the reduction factor can conveniently be rewritten in another dimensionless form:

$$\mathcal{T}^{-1} = 1 + \frac{\zeta}{\bar{\lambda}}\,, \tag{9.22}$$

where we have adopted the same scaling defined in equation (9.18) and a new parametrization involving

$$\zeta \equiv k_{\mathrm{T}}\langle z\rangle\,. \tag{9.23}$$

The case $\zeta = 0$ represents the limit of an infinitesimally thin system. Proceeding along the same line as before, we derive the following expression for the marginal stability curve:

$$Q^2 = \frac{4\bar{\lambda}^2}{\bar{\lambda} + \zeta}\left[(1 - \zeta) - \bar{\lambda}\right]\,, \tag{9.24}$$

which we consider in the range $0 \le \bar{\lambda} \le 1 - \zeta$.

The astrophysical relevance of the two parametrizations introduced above and their mathematical implications will be clarified in Subsection 9.3.2 in the more general case of two-component galactic discs, to which our analysis is mainly devoted. For the moment it is sufficient to note that these two parametrizations are so related:

$$\zeta = \frac{1}{2}Q^2\delta\,, \tag{9.25}$$

as can be established comparing equations (9.17) and (9.22). As we shall explain in more detail in Subsection 9.3.2, for a given δ only the prescription

$$\zeta = \bar{\zeta} \equiv \frac{1}{2}\,\bar{Q}^2\delta \tag{9.26}$$

involving the maximum of the marginal stability curve $\bar{Q}^2 = \bar{Q}^2(\delta)$ is physically meaningful in a local approach, because it refers to situations characterized by the same critical level of stability. The marginal stability curves derived in the context of the two parametrizations (9.19) and (9.23) are compared according to the prescription (9.26) in Figures 9.1–9.2, respectively. Their different qualitative behaviours can be ascribed to the different physical meaning of the two parametrizations, as will be discussed extensively in the two-component case.

9.3 Two-Component Case

9.3.1 Ansatz for the Reduction Factors

When two stellar populations are considered instead of one, the method employed by Vandervoort (1970a) to solve the Vlasov equation when z-motions are taken into account applies separately to each component. As we have stressed in Subsection 9.2.2, the same relation between the perturbations induced in the volume density and in the potential can be obtained in the framework of a fluid approach to the non-trivial lowest order of approximation, provided a proper reinterpretation of the relevant quantities is given (cf. Shu 1968). Therefore, assuming that the ordering

$$\frac{\langle z\rangle}{\langle r\rangle} \ll 1\,, \quad \frac{m}{|k|r} \sim \frac{c_{r\mathrm{H}}}{r\kappa} \ll 1\,, \quad |k|\langle z\rangle = O(1) \tag{9.27}$$

holds, we find that the wave equation (9.11) is replaced by

$$\frac{\partial^2\Phi_1}{\partial z^2} - k^2\Phi_1 = -\left(\frac{\Lambda_\mathrm{H}}{z^2_{\mathrm{eff\,H}}}\,\frac{\rho_\mathrm{H}}{\rho_{0\mathrm{H}}} + \frac{\Lambda_\mathrm{C}}{z^2_{\mathrm{eff\,C}}}\,\frac{\rho_\mathrm{C}}{\rho_{0\mathrm{C}}}\right)\Phi_1\,,$$

$$\lim_{|z|\to+\infty}\Phi_1(r,\theta,z;t) = 0\,, \quad \text{where} \quad \Lambda_\mathrm{i} \equiv |k| z_{\mathrm{eff\,i}}\mathcal{D}_\mathrm{i}\,. \tag{9.28}$$

In analogy with the one-component case we have defined

$$\mathcal{D}_\mathrm{i} \equiv \frac{4\pi G\rho_{0\mathrm{i}}z_{\mathrm{eff\,i}}|k|}{\kappa^2 - (\omega - m\Omega)^2}\,F_{\mathrm{i}\nu}(x_\mathrm{i}) = \frac{2\pi G\sigma_\mathrm{i}|k|}{\kappa^2 - (\omega - m\Omega)^2}\,F_{\mathrm{i}\nu}(x_\mathrm{i})\,, \tag{9.29}$$

where $\mathcal{D}_\mathrm{H} + \mathcal{D}_\mathrm{C} = 1$ is just the uncorrected local dispersion relation and the reduction factors $F_{\mathrm{i}\nu}(x_\mathrm{i}) = F_{\mathrm{kin\,i}\nu}(x_{\mathrm{kin\,i}}), F_{\mathrm{fl\,i}\nu}(x_{\mathrm{fl\,i}})$ are given by equations

(9.5), (9.6) and (9.15), (9.16) respectively, depending on the kind of approach employed to describe each component.

In contrast to the one-component case which is exactly soluble, for arbitrary values of the local parameter $\beta_z \equiv c_{zC}^2/c_{zH}^2$ or, equivalently, $\beta_{z\,\mathrm{eff}} \equiv z_{\mathrm{eff\,C}}/z_{\mathrm{eff\,H}}$ the wave equation (9.28) cannot be reduced to a Fuchsian differential equation or to other well-known classes of differential equations whose solutions are expressible in terms of elementary or special functions, even in the gaussian and exponential asymptotic limits (see, e.g., Abramowitz and Stegun 1970; Bender and Orszag 1978; Erdélyi 1956, 1981; Gradshteyn and Ryzhik 1980; Morse and Feshbach 1953; Smirnov 1964).

In order to overcome such difficulties connected with the solution of this **Double-eigenvalue problem** for Λ_{H} and Λ_{C}, we have developed a perturbative method (Romeo 1987) similar but not identical to that employed in quantum mechanics to solve the time-independent Schrödinger equation when the potential is of the form $U(z) = V(z) + W(z)$, $W(z)$ representing a small perturbation to $V(z)$ (see, e.g., Bender and Orszag 1978). Differences between our perturbative method and that used in quantum mechanics arise because, in quantum-mechanical language, the energy of the particle is kept fixed and *two* eigenvalues are associated with a *double*-term potential. This method has further been investigated and refined by the author, but the results are still in a preliminary form. Note that a WKBJ approach leading to a quantum condition of Bohr-Sommerfeld type cannot be employed here because it fails for small values of the quantum number (recall that we are interested in $n = 0$), as can be deduced by a comparison with the one-component case.

For what follows it is sufficient to make some general remarks concerning the wave equation (9.28), without going into the details of the perturbative method.

The quantum condition of this double-eigenvalue problem, which determines the corrected local dispersion relation, is expected to be of the form

$$\mathcal{Q}(\Lambda_{\mathrm{i}}, |k| z_{\mathrm{eff\,j}}, \alpha; n) = 0 \,. \tag{9.30}$$

While it seems reasonable that also in this two-component case only the lowest eigenvalues ($n = 0$) are physically relevant, it cannot be expected "a priori" that such relation can be reduced to the particular form

$$\mathcal{D}_{\mathrm{H}}\, \mathfrak{T}_{\mathrm{H}}(|k| z_{\mathrm{eff\,i}}, \alpha) + \mathcal{D}_{\mathrm{C}}\, \mathfrak{T}_{\mathrm{C}}(|k| z_{\mathrm{eff\,i}}, \alpha) = 1 \tag{9.31}$$

in analogy with the one-component case, where $\mathfrak{T}_{\mathrm{i}}(|k| z_{\mathrm{eff\,j}}, \alpha)$ are the reduction factors of the two components. If this is not the case, finite-thickness corrections to the local dispersion relation cannot simply be expressed in terms of two reduction factors, one for each component. In other words, the local dispersion relation cannot be reduced to the form obtained in the case of infinitesimally thin discs by a suitable scaling of the surface densities of the two components.

It is interesting to note in this context that the condition for the occurrence of a restricted quantum condition giving rise to a local dispersion relation of the "reduced" form (9.31) can be identified with the convergence criterion of the perturbative method, which is expected to take place only in particular asymptotic regimes of the local parameters α, β_z or equivalently $\beta_{z\,\mathrm{eff}}$ (Romeo 1987, and unpublished results; cf. Shu 1968).

From the considerations made above it follows that, in contrast to the one-component case, finite-thickness corrections to the local dispersion relation do not generally reduce to a simple scaling of the surface densities of the two components. The use of two corresponding reduction factors can reasonably be justified only when the two components are not strongly coupled. For the moment we are not able to express this statement in a precise mathematical form. In most situations of astrophysical interest, such as in the case in which the two components are identified with the stars of the active disc and the cold interstellar gas, this condition is rather general and is fulfilled throughout the galactic disc, except possibly in the outermost parts where the mass density of the cold interstellar gas becomes comparable to that of low-velocity dispersion stars. It is to such cases that our analysis is devoted.

We shall now make some simple ansatz concerning the form of the two reduction factors, which for the moment can only be justified at an intuitive level. A comparison between the wave equations (9.11) and (9.28) derived in the one-component and in the two-component cases, respectively, shows that Δ and $z_{\mathrm{eff\,i}}$ play in a sense a similar role: they are the effective thickness-scales which allow to express the local dispersion relations in terms of the unperturbed surface densities, introduced in place of the unperturbed volume densities in the plane. On the other hand, the different z-dependence of the equilibrium volume densities in the two cases is a source of dissimilarity in such a role.

This argument suggests that when the contribution of the last effect can be neglected, or in other words when the two components are not strongly coupled (cf. the non-restrictive condition mentioned above), the corresponding reduction factors should fairly well be approximated by the Effective ansatz

$$\mathfrak{T}_{\mathrm{i}} \equiv \frac{\sigma_{\mathrm{red\,i}}}{\sigma_{\mathrm{i}}} \simeq \mathfrak{T}_{\mathrm{eff\,i}} \equiv \frac{1}{1+|k|z_{\mathrm{eff\,i}}} \quad [\,|k|z_{\mathrm{eff\,i}} = O(1)\,]\,, \tag{9.32}$$

where we have indicated the formal ordering which we expect to be required by the underlying approximation.

Other reasonable ansatz involving the asymptotic thickness-scales of the two components and characterized by a lower level of accuracy can heuristically be justified. We shall refer to them as the Gaussian ansatz

$$\mathfrak{T}_{\rm i} \approx \mathfrak{T}_{\rm Gi} \equiv \frac{1}{1+|k|z_{\rm Gi}} \quad [\,|k|z_{\rm GC} \gg 1\,], \tag{9.33}$$

the Exponential ansatz

$$\mathfrak{T}_{\rm i} \approx \mathfrak{T}_{\rm Ei} \equiv \frac{1}{1+2|k|z_{\rm Ei}} \quad [\,2|k|z_{\rm EH} \ll 1\,], \tag{9.34}$$

and the Mixed ansatz

$$\mathfrak{T}_{\rm H} \approx \mathfrak{T}_{\rm GH} \equiv \frac{1}{1+|k|z_{\rm GH}} \quad [\,|k|z_{\rm GH} \gg 1\,],$$

$$\mathfrak{T}_{\rm C} \approx \mathfrak{T}_{\rm EC} \equiv \frac{1}{1+2|k|z_{\rm EC}} \quad [\,2|k|z_{\rm EC} \ll 1\,]. \tag{9.35}$$

Attention must be paid not to confuse these ansatz with the homonymous approximations introduced in Subsection 8.3.3, both involving the asymptotic thickness-scales of the two components but in conceptually different contexts. Note that the orderings specified in equations (9.27) and (9.32) are to be understood in the maximal sense (cf. maximal orderings in asymptotic perturbation expansions), and hence do not exclude the more particular orderings relevant to equations (9.33)–(9.35).

As mentioned above, the effective ansatz should provide more *accurate* estimates than the asymptotic ansatz (the level of accuracy is comparable in regimes of astrophysical interest), and in any case has a wider range of applicability. A similar level of accuracy is accomplished when the system approaches the limiting cases in which only one component is present or two components with the same vertical velocity dispersions can be identified, apart from the obvious infinitesimally thin case. These asymptotic regimes are correctly described and the corresponding limiting cases exactly reproduced by all the ansatz introduced above. Moreover, asymptotically our ansatz match the reduction factors derived by Shu (1968) in such particular regimes.

Finally, note that when a single equivalent component is taken to be representative of the whole active disc, the Global Effective ansatz

$$\mathfrak{T}_{\rm i} \sim \mathfrak{T}_{\rm glob} \equiv \frac{1}{1+|k|\langle z\rangle} \quad [\,|k|\langle z\rangle = O(1)\,] \tag{9.36}$$

represents the correct prescription for estimating the corresponding reduction factor. One should be careful, anyway, not to identify this ansatz with the reduction factor derived by Vandervoort (1970a): $\langle z\rangle$ has a different conceptual meaning in the two cases [cf. equations (8.17) and (8.41), respectively]. We argue that our estimate is more *accurate* than that provided by Vandervoort (1970a), just because it refers to the whole system and thus contains implicitly also the contribution of the cold interstellar gas.

9.3.2 Local Stability

The qualitative discussion given in the previous subsection will now be quantified, by studying in the framework of a fluid approach the marginal stability curves corresponding to the ansatz introduced above. In the same spirit as in Subsection 9.2.2 we shall introduce two parametrizations, the use of which is suggested by their astrophysical relevance and/or mathematical convenience, as will be explained in detail later on.

Let us first consider the two-component extension of the parametrization (9.19). The reduction factors of the two components can be expressed in the following general dimensionless form:

$$\mathfrak{T}_{\mathrm{i}}^{-1} = 1 + \frac{Q_{\mathrm{H}}^2}{2\bar{\lambda}}\,\mathfrak{A}_{\mathrm{i}}\,, \tag{9.37}$$

where the functions $\mathfrak{A}_{\mathrm{i}}$ depend on the ansatz employed:

$$\mathfrak{A}_{\mathrm{eff\,H}} \equiv \frac{\delta_{\mathrm{H}}}{1+\alpha}\,\frac{z_{\mathrm{eff\,H}}}{2z_{\mathrm{EH}}}\,, \qquad \mathfrak{A}_{\mathrm{eff\,C}} \equiv \frac{\beta\delta_{\mathrm{C}}}{1+\alpha}\,\frac{z_{\mathrm{eff\,C}}}{2z_{\mathrm{EC}}}\,; \tag{9.38}$$

$$\mathfrak{A}_{\mathrm{GH}} \equiv \frac{\delta_{\mathrm{H}}}{1+\alpha}\,\frac{z_{\mathrm{GH}}}{2z_{\mathrm{EH}}}\,, \qquad \mathfrak{A}_{\mathrm{GC}} \equiv \frac{\beta\delta_{\mathrm{C}}}{1+\alpha}\,\frac{z_{\mathrm{GC}}}{2z_{\mathrm{EC}}}\,; \tag{9.39}$$

$$\mathfrak{A}_{\mathrm{EH}} \equiv \frac{\delta_{\mathrm{H}}}{1+\alpha}\,, \qquad \mathfrak{A}_{\mathrm{EC}} \equiv \frac{\beta\delta_{\mathrm{C}}}{1+\alpha}\,; \tag{9.40}$$

in the same subscript notations used for the reduction factors. In these formulae we have adopted the same scaling and basic parametrization defined in equations (7.7) and (7.5), (7.6) respectively, and the additional Velocity-Dispersion parametrization

$$\delta_{\mathrm{i}} \equiv \frac{c_{z\mathrm{i}}^2}{c_{\mathrm{i}}^2} \quad (0 < \beta\delta_{\mathrm{C}} < \delta_{\mathrm{H}} < 1,\ \ 0 < \delta_{\mathrm{C}} \leq 1)\,, \tag{9.41}$$

based on the parametrization involving α, β_z introduced in Subsection 8.3.3 and so related to it:

$$\beta_z = \beta\,\frac{\delta_{\mathrm{C}}}{\delta_{\mathrm{H}}}\,. \tag{9.42}$$

The case $\delta_{\mathrm{i}} = 0$ represents the limit of an infinitesimally thin system. When the two components are identified with the stars of the active disc and the cold interstellar gas ($\delta_{\mathrm{C}} = 1$, due to its collisional nature), the cases $\delta_{\mathrm{H}} = \beta$ (i.e., $\beta_z = 1$) and $\delta_{\mathrm{H}} = 1$ correspond to a totally ineffective vertical heating and to an isotropic heating, respectively.

The marginal stability curve in the $(\bar{\lambda}, Q_{\mathrm{H}}^2)$ plane can be derived from its uncorrected expression along a line similar to the one-component case:

$$\beta\,\frac{Q_{\mathrm{H}}^4}{\mathfrak{T}_{\mathrm{H}}^4} + 4\,\frac{\bar{\lambda}}{\mathfrak{T}_{\mathrm{H}}}\,\frac{Q_{\mathrm{H}}^2}{\mathfrak{T}_{\mathrm{H}}^2}\left[\frac{\bar{\lambda}}{\mathfrak{T}_{\mathrm{H}}}(1+\beta) - \left(\alpha\,\frac{\mathfrak{T}_{\mathrm{C}}}{\mathfrak{T}_{\mathrm{H}}} + \beta\right)\right] + 16\,\frac{\bar{\lambda}^3}{\mathfrak{T}_{\mathrm{H}}^3}\left[\frac{\bar{\lambda}}{\mathfrak{T}_{\mathrm{H}}} - \left(1 + \alpha\,\frac{\mathfrak{T}_{\mathrm{C}}}{\mathfrak{T}_{\mathrm{H}}}\right)\right] = 0$$

$$\Downarrow$$

$$\begin{aligned}
&\beta \mathfrak{A}_{\mathrm{H}} \mathfrak{A}_{\mathrm{C}} Q_{\mathrm{H}}^{8} \\
&+ 2\bar{\lambda} Q_{\mathrm{H}}^{6} \Big[2\bar{\lambda} (1+\beta) \mathfrak{A}_{\mathrm{H}} \mathfrak{A}_{\mathrm{C}} + \beta (\mathfrak{A}_{\mathrm{H}} + \mathfrak{A}_{\mathrm{C}}) \Big] \\
&+ 4\bar{\lambda}^{2} Q_{\mathrm{H}}^{4} \Big\{ 4\bar{\lambda}^{2} \mathfrak{A}_{\mathrm{H}} \mathfrak{A}_{\mathrm{C}} + 2\bar{\lambda} (1+\beta) (\mathfrak{A}_{\mathrm{H}} + \mathfrak{A}_{\mathrm{C}}) + \Big[\beta - 2 (\alpha \mathfrak{A}_{\mathrm{H}} + \beta \mathfrak{A}_{\mathrm{C}}) \Big] \Big\} \\
&+ 16\bar{\lambda}^{3} Q_{\mathrm{H}}^{2} \Big\{ 2\bar{\lambda}^{2} (\mathfrak{A}_{\mathrm{H}} + \mathfrak{A}_{\mathrm{C}}) + \bar{\lambda} \Big[(1+\beta) - 2 (\alpha \mathfrak{A}_{\mathrm{H}} + \mathfrak{A}_{\mathrm{C}}) \Big] - (\alpha + \beta) \Big\} \\
&+ 64\bar{\lambda}^{5} \Big[\bar{\lambda} - (1+\alpha) \Big] = 0\,, \qquad (9.43)
\end{aligned}$$

which we consider in the range $0 \le \bar{\lambda} \le 1+\alpha$. From this relation a local stability criterion can be stated in analogy with the case of one-component systems. A function $\bar{Q}^2 = \bar{Q}^2(\alpha, \beta, \delta_{\mathrm{i}})$, which reduces to unity when $\alpha = 0$ and $\delta_{\mathrm{i}} = 0$, can be defined in such a way that when $Q_{\mathrm{H}}^2 > \bar{Q}^2$ the system is locally stable at all wavelengths. For $Q_{\mathrm{H}}^2 < \bar{Q}^2$ we expect the system to be locally unstable in ranges of wavelengths defined by the marginal stability condition.

Let us now consider the two-component extension of the parametrization (9.23). The reduction factors of the two components will be evaluated according to the Effective ansatz alone, which is expected to provide more accurate estimates than the asymptotic ansatz. They can be expressed in the following dimensionless form:

$$\mathfrak{T}_{\mathrm{eff\,H}}^{-1} = 1 + \frac{\zeta_{\mathrm{H}}}{\bar{\lambda}}\,, \quad \mathfrak{T}_{\mathrm{eff\,C}}^{-1} = 1 + \frac{\zeta_{\mathrm{H}}}{\bar{\lambda}} \beta_{z\,\mathrm{eff}}\,, \qquad (9.44)$$

where we have adopted the same scaling and basic parametrization used above, and the additional Wave-Number parametrization

$$\zeta_{\mathrm{H}} \equiv k_{\mathrm{H}} z_{\mathrm{eff\,H}}\,, \quad \beta_{z\,\mathrm{eff}} \equiv \frac{z_{\mathrm{eff\,C}}}{z_{\mathrm{eff\,H}}} \quad (0 < \beta_{z\,\mathrm{eff}} < 1)\,, \qquad (9.45)$$

based on the parametrization involving α, $\beta_{z\,\mathrm{eff}}$ introduced in Subsection 8.3.3 and similar to that adopted by Shu (1968). The case $\zeta_{\mathrm{H}} = 0$ represents the limit of an infinitesimally thin system, and the case $\beta_{z\,\mathrm{eff}} = 1$ represents the limit of a system in which the two components have the same scale-height (and hence the same vertical velocity dispersions).

Proceeding along the same line as before, we derive the following expression for the marginal stability curve:

$$\begin{aligned}
&\beta Q_{\mathrm{H}}^{4} \Big[\bar{\lambda}^{2} + \bar{\lambda} \zeta_{\mathrm{H}} (1 + \beta_{z\,\mathrm{eff}}) + \zeta_{\mathrm{H}}^{2} \beta_{z\,\mathrm{eff}} \Big] \\
&+ 4\bar{\lambda}^{2} Q_{\mathrm{H}}^{2} \Big\{ \bar{\lambda}^{2} (1+\beta) - \bar{\lambda} \Big[(\alpha+\beta) - (1+\beta) \zeta_{\mathrm{H}} (1 + \beta_{z\,\mathrm{eff}}) \Big] \\
&\qquad + \zeta_{\mathrm{H}} \Big[(1+\beta) \zeta_{\mathrm{H}} \beta_{z\,\mathrm{eff}} - (\alpha + \beta \beta_{z\,\mathrm{eff}}) \Big] \Big\}
\end{aligned}$$

$$+ 16\bar{\lambda}^4\Big\{\bar{\lambda}^2 - \bar{\lambda}\Big[(1+\alpha) - \zeta_{\rm H}\,(1+\beta_{z\,\rm eff})\Big] + \zeta_{\rm H}\Big[\zeta_{\rm H}\beta_{z\,\rm eff} - (\alpha+\beta_{z\,\rm eff})\Big]\Big\} = 0\,, \qquad (9.46)$$

which we consider in the range $0 \le \bar{\lambda} \le \bar{\lambda}_{\maltese}$, where the upper zero is given by

$$\begin{aligned} \bar{\lambda}_{\maltese} &= \frac{1}{2}\Big\{(1+\alpha) - \zeta_{\rm H}\,(1+\beta_{z\,\rm eff}) \\ &\quad + \sqrt{\Big[(1+\alpha) - \zeta_{\rm H}\,(1+\beta_{z\,\rm eff})\Big]^2 - 4\zeta_{\rm H}\Big[\zeta_{\rm H}\beta_{z\,\rm eff} - (\alpha+\beta_{z\,\rm eff})\Big]}\,\Big\} \\ &< 1+\alpha\,. \end{aligned} \qquad (9.47)$$

Note that in the context of this parametrization the marginal stability curve can degenerate into the origin of the $(\bar{\lambda}, Q_{\rm H}^2)$ plane in the critical regimes of the local parameters $\zeta_{\rm H}$ and $\beta_{z\,\rm eff}$ corresponding to the vanishing of its upper zero. The condition which should be satisfied for avoiding this singular behaviour of the marginal stability curve can thus be expressed by imposing the positivity of its upper zero:

$$\bar{\lambda}_{\maltese} > 0 \quad\Longrightarrow\quad \zeta_{\rm H} < 1 + \frac{\alpha}{\beta_{z\,\rm eff}}\,. \qquad (9.48)$$

However, this is not a physical limitation for $\zeta_{\rm H}$ because it corresponds to the obvious conditions $\delta_{\rm i} < +\infty$, as can be deduced comparing the dimensionless forms of the reduction factors of the two components evaluated according to the effective ansatz in the context of these two parametrizations. Moreover, values of $\zeta_{\rm H}$ sufficiently close to such a critical upper bound are not consistent with the working assumption $|k|\langle z\rangle = O(1)$ in the range of wavelengths relevant to the marginal stability curve. Taking these considerations into account, the same local stability criterion stated in the context of the velocity-dispersion parametrization applies, with $\bar{Q}^2 = \bar{Q}^2(\alpha, \beta, \zeta_{\rm H}, \beta_{z\,\rm eff})$.

We shall now discuss the advantages and drawbacks inherent in these two parametrizations.

A considerable advantage of the wave-number parametrization lies in the fact that the equation for the marginal stability curve can be solved analytically according to the standard technique employed in the case of quadratic algebraic equations. The equation for its stationary points can thus be given in a relatively simple explicit form, which turns out to be extremely useful for investigating the corrected two-phase region. Instead, when the velocity-dispersion parametrization is adopted, the stationariness condition can only be given in the implicit form of a system of two non-linear (quartic and cubic) algebraic equations, the alternative explicit relation being of no practical utility.

An advantage of the velocity-dispersion parametrization which should not be underestimated lies in the fact that, when the two components are identified with the stars of the active disc and the cold interstellar gas, the lower and the upper physical limitations corresponding to a totally ineffective vertical heating and to an isotropic heating of the stellar component respectively, as derived on observational grounds and suggested by stability considerations, can directly be included and the collisional nature of the gaseous component can trivially be taken into account. Instead, when the wave-number parametrization is adopted, the previous conditions do not admit any straightforward translation.

Even though the wave-number parametrization is not as physically transparent as the velocity-dispersion parametrization, its use can reasonably be justified in view of its considerable mathematical convenience. Therefore, in order to make up for such a drawback, it is of extreme interest to study the way in which the results obtained in the context of one of them can be translated into corresponding results for the other.

The first step consists in noting that these two parametrizations are so related:

$$\zeta_{\mathrm{H}} = \frac{1}{2} Q_{\mathrm{H}}^2 \frac{\delta_{\mathrm{H}}}{1+\alpha} \frac{z_{\mathrm{eff\,H}}}{2 z_{\mathrm{BH}}}, \tag{9.49}$$

as can be established comparing equations (9.37), (9.38) and (9.44).

Now the question arises which value of the local stability parameter should be used in this relation. If the two-component equilibrium model is specified (global stability analysis), there is no ambiguity: the relevant value of the local stability parameter is that corresponding to the point where all the other profiles are calculated. If the two-component equilibrium model is not specified (local stability analysis), and if we want to compare the marginal stability curves derived in the context of these two parametrizations, a natural and physically meaningful choice is represented by the Global prescription

$$\zeta_{\mathrm{H}} = \bar{\zeta}_{\mathrm{H}} \equiv \frac{1}{2} \bar{Q}^2 \frac{\delta_{\mathrm{H}}}{1+\alpha} \frac{z_{\mathrm{eff\,H}}}{2 z_{\mathrm{BH}}} \tag{9.50}$$

involving the global maximum of the marginal stability curve, which refers to situations characterized by the same critical level of stability. Note, however, that this is not the only reasonable choice if peculiar gas-dominated regimes are involved: depending on the various situations, it could be of more interest to consider one of the other stationary points of the marginal stability curve.

If the global prescription is used to relate these two parametrizations, the corresponding marginal stability curves are characterized by a global maximum having the same location and height. On the other hand, any other point $(\bar{\lambda}, Q_{\mathrm{H}}^2)$ belonging to the marginal stability curve derived in the context of the wave-number parametrization can be viewed as belonging to the marginal stability curve derived in the context of the velocity-dispersion parametrization for correspondingly larger values of the local parameters $\delta_{\rm i}$:

$$\delta_{\rm i}' = \bar{\delta}_{\rm i} \equiv \delta_{\rm i}\,\frac{\bar{Q}^2}{Q_{\rm H}^2} > \delta_{\rm i}\,, \tag{9.51}$$

as can be deduced comparing equations (9.49) and (9.50) once the monotonic behaviour of $\beta_{z\,\rm eff}$ with varying β_z at fixed α is taken into account. Unfortunately, these descriptive complications cannot be avoided.

As an immediate consequence of the considerations made above, it can be shown that the wave-number parametrization exhibits some unphysical features already suspected in the discussion concerning the upper limitation (9.48). Here we do not want to consider this problem in detail, but just to give a correct idea of the reason for such a singular behaviour.

For any given value of $\zeta_{\rm H}$, however small, the upper physical limitations $\delta_{\rm i} \leq 1$ can only be fulfilled in a part of the range of wavelengths relevant to the marginal stability curve, because as $\bar{\lambda}$ approaches the lower and upper zeros the corresponding $\delta_{\rm i}$ take arbitrarily large values [see in particular equation (9.51) and relative discussion]. Note in this context that, as regards the consistency with the working assumption expressed by the maximal ordering $|k|\langle z\rangle = O(1)$, the upper bounds $\delta_{\rm i} = 1$ roughly correspond to the onset of a "dangerous" situation at the global maximum of the marginal stability curve or at the gaseous peak, if present (the gaseous peak is always more "dangerous" than the stellar peak, even in the case in which it is less high, because it occurs at much shorter wavelengths). On the other hand, the lower physical limitations $\delta_{\rm H} > \beta\delta_{\rm C} > 0$ do not give rise to such difficulties when the global prescription is used.

From these considerations it follows that the marginal stability curve derived in the context of the wave-number parametrization does not provide a faithful representation of the local stability properties of galactic discs.

In presenting the results of the local stability analysis performed in this chapter, we have considered the standard star-dominated and the peculiar gas-dominated regimes already investigated in Chapter 7 in the less general context of infinitesimally thin two-component models of galactic discs. Moreover, other regimes of the relevant local parameters which are typical in the solar neighbourhood have been considered.

The marginal stability curves derived according to the effective ansatz in the context of the velocity-dispersion and of the wave-number parametrizations are compared in Figures 9.3–9.4, respectively. We have used the global prescription to relate these two parametrizations, and the fit approximation to express the equilibrium-related dimensionless quantities in terms of α and β_z. The non-perfect agreement between the critical levels of stability characterizing corresponding situations (in the sense of the global prescription) can be ascribed to two facts: the low accuracy of the fit approximation in peculiar gas-dominated regimes, and the evaluation of the input values for the wave-number parametrization only up to the second decimal digit. Apart from these numerical effects, the qualitatively different behaviour of corresponding marginal stability curves can be traced back to the fact that the wave-number parametrization contains the local value of the stability parameter implicitly, as already explained.

A comparison is also made between the gaussian, the exponential and the mixed ansatz in the context of the velocity-dispersion parametrization supplemented by the fit approximation (not necessary in the exponential case). The corresponding marginal stability curves are shown in Figures 9.5–9.7, respectively. In addition, a comparison is made between the gaussian and the mixed ansatz in the context of the velocity-dispersion parametrization supplemented by the (self-consistent) exponential approximation. The corresponding marginal stability curves are shown in Figures 9.8–9.9, respectively. The quantitative discrepancies found in Figures 9.5–9.9 can easily be understood by taking into account the decoupling of the two components in peculiar gas-dominated regimes, and noting that the chain of inequalities (8.46) between the relevant thickness-scales implies the following chain of inequalities between the reduction factors evaluated according to the various ansatz:

$$\mathfrak{T}_{\mathrm{EH}} < \mathfrak{T}_{\mathrm{eff\,H}} < \mathfrak{T}_{\mathrm{GH}} < \mathfrak{T}_{\mathrm{glob}} < \mathfrak{T}_{\mathrm{GC}} < \mathfrak{T}_{\mathrm{eff\,C}} < \mathfrak{T}_{\mathrm{EC}}\,. \tag{9.52}$$

As a result, when finite-thickness effects are taken into account, the system tends to be more stable. More precisely, the stabilizing role of such effects can partially counterbalance the destabilizing role of the cold interstellar gas in linear regimes. As a consequence of these competing roles, in some cases the properties of the marginal stability curve are qualitatively different from the infinitesimally thin case.

We shall now use some qualitative arguments, based on the first step of an iterative method for calculating Q_{H}^2 from its uncorrected value, to provide a simple justification of these results. Such arguments are only indicative, and may be quite inaccurate if the convergence of the iterative method is slow. For what follows it is sufficient to understand the way in which the points of the marginal stability curve relevant to the stability analysis are modified when finite-thickness effects are taken into account. However, it should be borne in mind that such simple arguments are indeed able to account for all the basic properties of the marginal stability curve. Specifically, we find for the Upper zero

$$\bar{\lambda}_{\maltese} \sim \mathfrak{T}_{\mathrm{H}} + \alpha\mathfrak{T}_{\mathrm{C}} \lesssim 1 + \alpha\,; \tag{9.53}$$

for the Stellar peak

$$\bar{\lambda} \sim \frac{1}{2}\mathfrak{T}_{\mathrm{H}} \lesssim \frac{1}{2}\,, \quad Q_{\mathrm{H}}^2 \sim \mathfrak{T}_{\mathrm{H}}^2 + 4\alpha\mathfrak{T}_{\mathrm{H}}\mathfrak{T}_{\mathrm{C}} \lesssim 1 + 4\alpha\,; \tag{9.54}$$

and for the Gaseous peak

$$\bar{\lambda} \sim \frac{1}{2}\,\alpha\mathfrak{T}_{\mathrm{C}} \lesssim \frac{1}{2}\,\alpha\,, \quad Q_{\mathrm{H}}^2 \sim \frac{\alpha^2}{\beta}\,\mathfrak{T}_{\mathrm{C}}^2 + 4\alpha\mathfrak{T}_{\mathrm{H}}\mathfrak{T}_{\mathrm{C}} \lesssim \frac{\alpha^2}{\beta} + 4\alpha\,. \tag{9.55}$$

In these formulae the reduction factors of the two components are intended to be calculated to the non-trivial lowest order of iteration. From these expressions it follows, in agreement with physical intuition, the *decoupling* of the two components in *peculiar* gas-dominated regimes, as expressed by the leading terms in the stellar and in the gaseous peaks.

Although such qualitative arguments account for the stabilization of these two peaks and their shift toward shorter wavelengths, a more sophisticated asymptotic expansion analysis is required to obtain more reliable estimates. Specifically, in the context of the wave-number parametrization, we find for the Upper zero

$$\bar{\lambda}_{\maltese} \approx 1 + (\alpha - \zeta_{\mathrm{H}})$$

$$[\,\alpha \ll 1;\ \beta = O(\alpha) \vee \beta = O(\alpha^2);\ \zeta_{\mathrm{H}}, \beta_{z\,\mathrm{eff}} = O(\alpha)\,]\,; \tag{9.56}$$

for the Stellar peak

$$\bar{\lambda} \approx \frac{1}{2}\,, \quad Q_{\mathrm{H}}^2 \approx 1 + 4\,(\alpha - \zeta_{\mathrm{H}})$$

$$[\,\alpha \ll 1;\ \beta = O(\alpha) \vee \beta = O(\alpha^2);\ \zeta_{\mathrm{H}}, \beta_{z\,\mathrm{eff}} = O(\alpha)\,]\,; \tag{9.57}$$

and for the Gaseous peak

$$\bar{\lambda} \approx \frac{1}{2}\,\alpha\,, \quad Q_{\mathrm{H}}^2 \approx \frac{\alpha^2}{\beta} + 4\alpha\left[1 - \left(\frac{2\zeta_{\mathrm{H}}}{\alpha + 2\zeta_{\mathrm{H}}} + \frac{\zeta_{\mathrm{H}}\beta_{z\,\mathrm{eff}}}{\beta}\right)\right]$$

$$[\,\alpha \ll 1;\ \beta = O(\alpha^2);\ \zeta_{\mathrm{H}}, \beta_{z\,\mathrm{eff}} = O(\alpha)\,]\,. \tag{9.58}$$

The orderings indicated below the corresponding asymptotic expressions are suggested by the analysis performed in the infinitesimally thin case, where they characterize the Two-Phase region of the (β, α) plane. Note in this context that the formal ordering involving the local parameters ζ_{H} and $\beta_{z\,\mathrm{eff}}$ is to be viewed as a maximal ordering, which at this stage we are not able to specify more precisely.

We shall now make some simple considerations concerning the astrophysical relevance of the two parametrizations introduced in this chapter. Even though the velocity-dispersion and wave-number parametrizations cannot easily be compared in a local stability analysis, when the two-component equilibrium galactic model is specified and a global stability analysis is performed, these two parametrizations turn out to be equivalent (mathematical complications apart). As suggested by observations, for our galaxy the proper input local parameters might be $\delta_{\mathrm{i}}(r)$ while for external galaxies $\zeta_{\mathrm{H}}(r)$ and $\beta_{z\,\mathrm{eff}}(r)$ might be more appropriate.

Thus, apparently, the situation seems to be more complicated for our galaxy: the local stability analysis which should be carried out to determine the global maximum $\bar{Q}^2(r)$ of the marginal stability curve in the context of the velocity-dispersion parametrization is considerably more difficult. If we want to bypass this difficulty adopting the wave-number parametrization, recall that relating the former to the latter is also not an easy task and, indeed, is equivalently difficult because it requires the knowledge of $\bar{Q}^2(r)$. The situation would certainly be more tractable, but still complicated, if the input profiles were $c_{\mathrm{zi}}(r)$, in which case it would be convenient to derive $z_{\mathrm{eff\,i}}(r)$ and to adopt the wave-number parametrization. In my opinion, since the vertical and planar heating mechanisms in galactic discs are almost decoupled as well as their relevant stability properties, the constancy of $\delta_{\mathrm{H}}(r)$ throughout the Galactic disc, commonly invoked to fit or deduce the relevant stellar velocity dispersion profiles, should be viewed as a convenient observational working assumption rather than a well-established observational constraint. Therefore, $\delta_{\mathrm{i}}(r)$ might not be so meaningful as input parameters even for our galaxy.

Also in the case of external galaxies there might be some interpretative complications in finding out whether the observed $z_{\mathrm{eff\,i}}(r)$ have the same physical meaning as those inferred on theoretical grounds: generally gaussian, exponential or sech-squared profiles are used to fit the observed brightness distribution of the stellar component, and a constant mass-to-light ratio is assumed to derive the corresponding volume mass density profiles. Consequently, these fitting analyses are not able to discriminate between the true $z_{\mathrm{eff\,H}}(r)$ and $z_{\mathrm{GH}}(r)$ or $z_{\mathrm{EH}}(r)$, so that we should carefully reinterpret all the related local parameters and derive their correct values. Alternatively, it could be wise to abandon the idea of performing a "strictly" correct analysis, and to be content with performing an "approximately" correct analysis in the spirit of the gaussian and (self-consistent) exponential approximations discussed in Subsection 8.3.3 and of the homonymous ansatz discussed in this subsection: the essential physics is not missed.

Other discussions of related topics can be found in Subsection 6.4.2 and in Bahcall (1984), Bahcall and Casertano (1984).

References

Abramowitz M. and Stegun I. A. (1970), *Handbook of Mathematical Functions*, Dover, New York.

Bahcall J. N. (1984), *Ap. J.* **276**, 156.

Bahcall J. N. and Casertano S. (1984), *Ap. J. Lett.* **284**, L35.

Balbus S. A. and Cowie L. L. (1985), *Ap. J.* **297**, 61.

Bender C. M. and Orszag S. A. (1978), *Advanced Mathematical Methods for Scientists and Engineers*, McGraw-Hill, New York.

Bertin G. and Casertano S. (1982), *A. Ap.* **106**, 274.

Erdélyi A. (1956), *Asymptotic Expansions*, Dover, New York.

Erdélyi A. (1981), *Higher Transcendental Functions*, 3 Vols., Krieger, Malabar.

Fridman A. M. (1989), in *Dynamics of Astrophysical Discs*, Cambridge University Press, Cambridge, Sellwood J. A. ed., p. 185 (Manchester, UK, 1988).

Fridman A. M. and Polyachenko V. L. (1984), *Physics of Gravitating Systems*, 2 Vols., Springer-Verlag, New York.

Genkin I. L. and Safronov V. S. (1975), *Sov. Astron.* **19**, 189.

Goldreich P. and Lynden-Bell D. (1965a), *M.N.R.A.S.* **130**, 97.

Goldreich P. and Lynden-Bell D. (1965b), *M.N.R.A.S.* **130**, 125.

Gradshteyn I. S. and Ryzhik I. M. (1980), *Table of Integrals, Series, and Products*, Academic Press, New York.

Jog C. J. and Solomon P. M. (1984a), *Ap. J.* **276**, 114.

Jog C. J. and Solomon P. M. (1984b), *Ap. J.* **276**, 127.

Landau L. D. and Lifshitz E. M. (1977), *Quantum Mechanics*, Pergamon Press, Oxford.

Lin C. C. (1970), in *Galactic Astronomy*, Vol. 2, Gordon and Breach, New York, Chiu H.-Y. and Muriel A. eds., p. 1 (Stony Brook, New York USA, 1968).

Lin C. C. and Shu F. H. (1971), in *Astrophysics and General Relativity*, Vol. 2, Gordon and Breach, New York, Chrétien M. et al. eds., p. 235 (Brandeis University, Waltham, Massachusetts USA, 1968).

Morozov A. G. and Khoperskov A. V. (1986), *Astrophys.* **24**, 266.

Morse P. M. and Feshbach H. (1953), *Methods of Theoretical Physics*, 2 Vols., McGraw-Hill, New York.

Nakamura T. (1978), *Prog. Theor. Phys.* **59**, 1129.

Peng Q.-H. (1988), *A. Ap.* **206**, 18.

Romeo A. B. (1987), *M. Phil. Thesis*, SISSA, Trieste, Italy.

Romeo A. B. (1990a), in *Chemical and Dynamical Evolution of Galaxies*, Giardini, Pisa, Ferrini F., Franco J. and Matteucci F. eds., p. 463 (Elba Island, Italy, 1989).

Romeo A. B. (1990b), "Self-Regulation and Secular Heating of Galactic Discs", to appear in *IAU Eur. Reg. Astron. Meeting* **12** (Davos, Switzerland, 1990).

Shu F. H. (1968), *Ph. D. Thesis*, Harvard University, Cambridge, Massachusetts USA.

Smirnov V. I. (1964), *A Course of Higher Mathematics*, Vols. 2–4, Pergamon Press, Oxford.

Sweet P. A. and McGregor D. D. (1964), *M.N.R.A.S.* **128**, 195.

Toomre A. (1964), *Ap. J.* **139**, 1217.

Toomre A. (1974), in *Highlights Astron.* **3**, Reidel, Dordrecht, Contopoulos G. ed., p. 457 (Sidney, Australia, 1973).

Vandervoort P. O. (1970a), *Ap. J.* **161**, 87.

Vandervoort P. O. (1970b), *Ap. J.* **162**, 453.
Yue Z. Y. (1982a), *Geophys. Astrophys. Fluid Dyn.* **20**, 1.
Yue Z. Y. (1982b), *Geophys. Astrophys. Fluid Dyn.* **20**, 47.
Yue Z. Y. (1982c), *Geophys. Astrophys. Fluid Dyn.* **20**, 69.

Figure Captions

Figure 9.1. One-fluid marginal stability curves in the $(\bar{\lambda}, Q^2)$ plane for some values of the local parameter δ. The case $\delta = 0$ represents the limit of an infinitesimally thin system.

Figure 9.2. Same as Figure 9.1 for the parametrization involving ζ. The input values of $\bar{\zeta}$ correspond to the input values of δ in Figure 9.1. They are calculated according to the prescription $\bar{\zeta} = \frac{1}{2}\bar{Q}^2\delta$ discussed in the text, which refers to situations characterized by the same critical level of stability, with an accuracy of two decimal digits.

Figure 9.3. Two-fluid marginal stability curves in the $(\bar{\lambda}, Q_{\mathrm{H}}^2)$ plane derived evaluating the reduction factors of the two components according to the effective ansatz for some values of the local parameters δ_{i}, α and fixed $\beta = .01$ (top), and for values of the same local parameters corresponding to typical lower and upper bounds in the solar neighbourhood (bottom). The fit approximation has been used to express the equilibrium-related dimensionless quantities in terms of α and β_z. The case $\delta_{\mathrm{i}} = 0$ represents the limit of an infinitesimally thin system. The cases $\delta_{\mathrm{H}} = \beta$ and $\delta_{\mathrm{H}} = 1$ correspond to a totally ineffective vertical heating and to an isotropic heating, respectively.

Figure 9.4. Same as Figure 9.3 for the wave-number parametrization. The input values of ζ_{H} and $\beta_{z\,\mathrm{eff}}$ correspond to the input values of δ_{i} in Figure 9.3. They are calculated according to the global prescription discussed in the text, which refers to situations characterized by the same critical level of stability, with an accuracy of two decimal digits.

Figure 9.5. Same as Figure 9.3 for the gaussian ansatz.

Figure 9.6. Same as Figure 9.3 for the exponential ansatz. The fit approximation has not been used in this case.

Figure 9.7. Same as Figure 9.3 for the mixed ansatz.

Figure 9.8. Same as Figure 9.5 for the (self-consistent) exponential approximation.

Figure 9.9. Same as Figure 9.7 for the (self-consistent) exponential approximation.

Figure 9.1

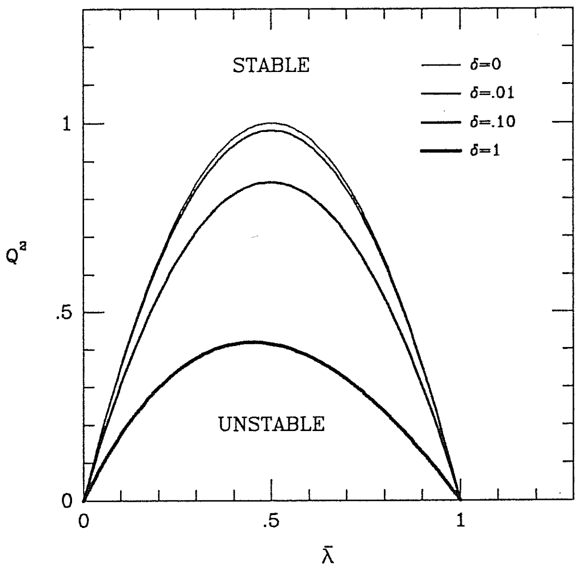


Figure 9.2

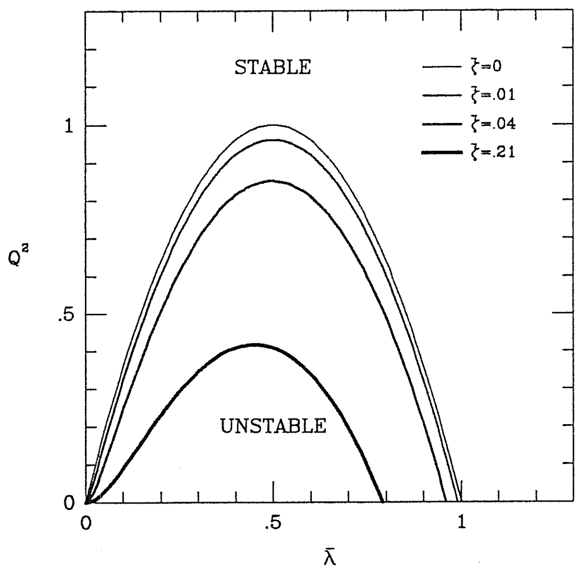
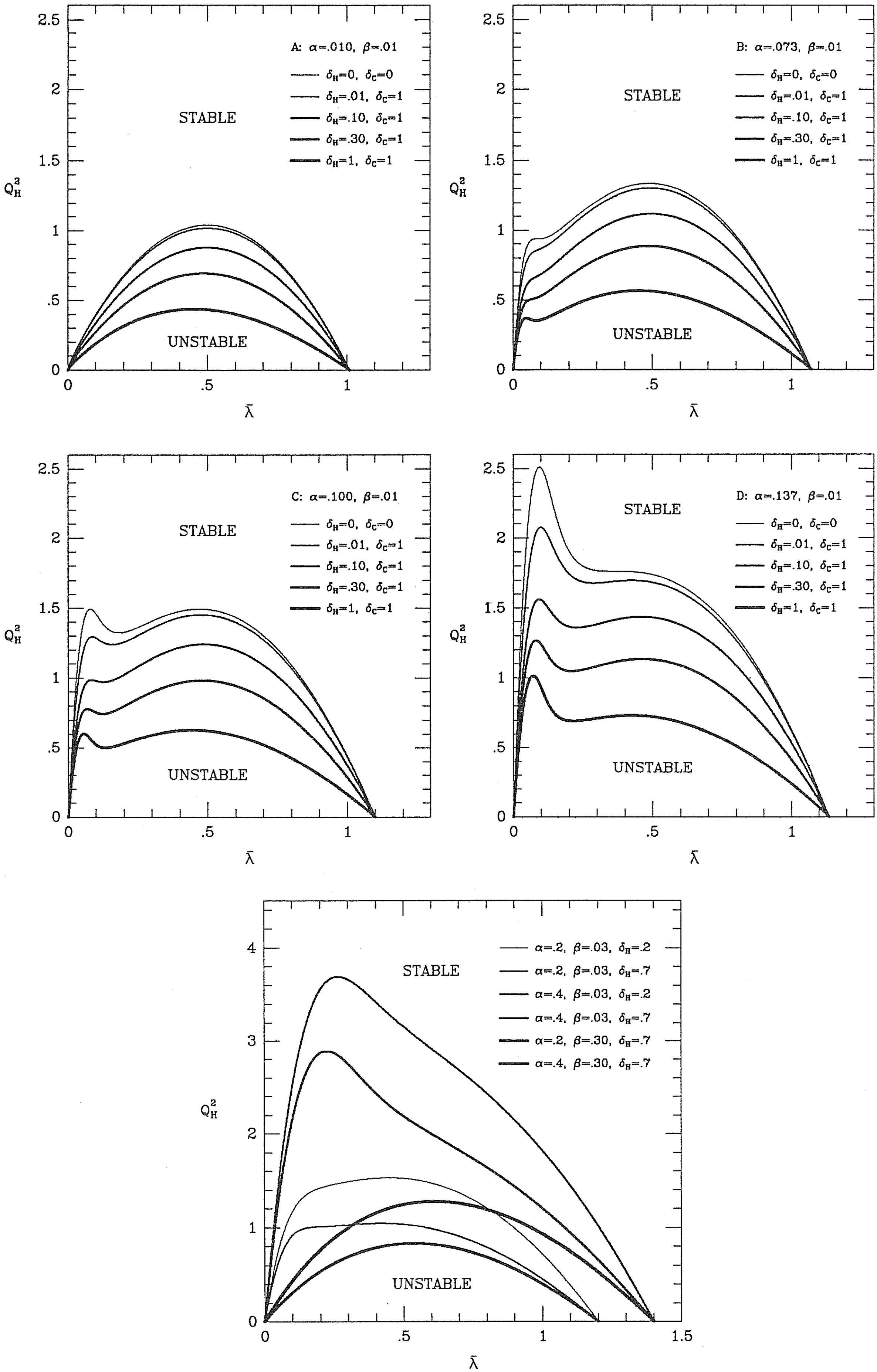
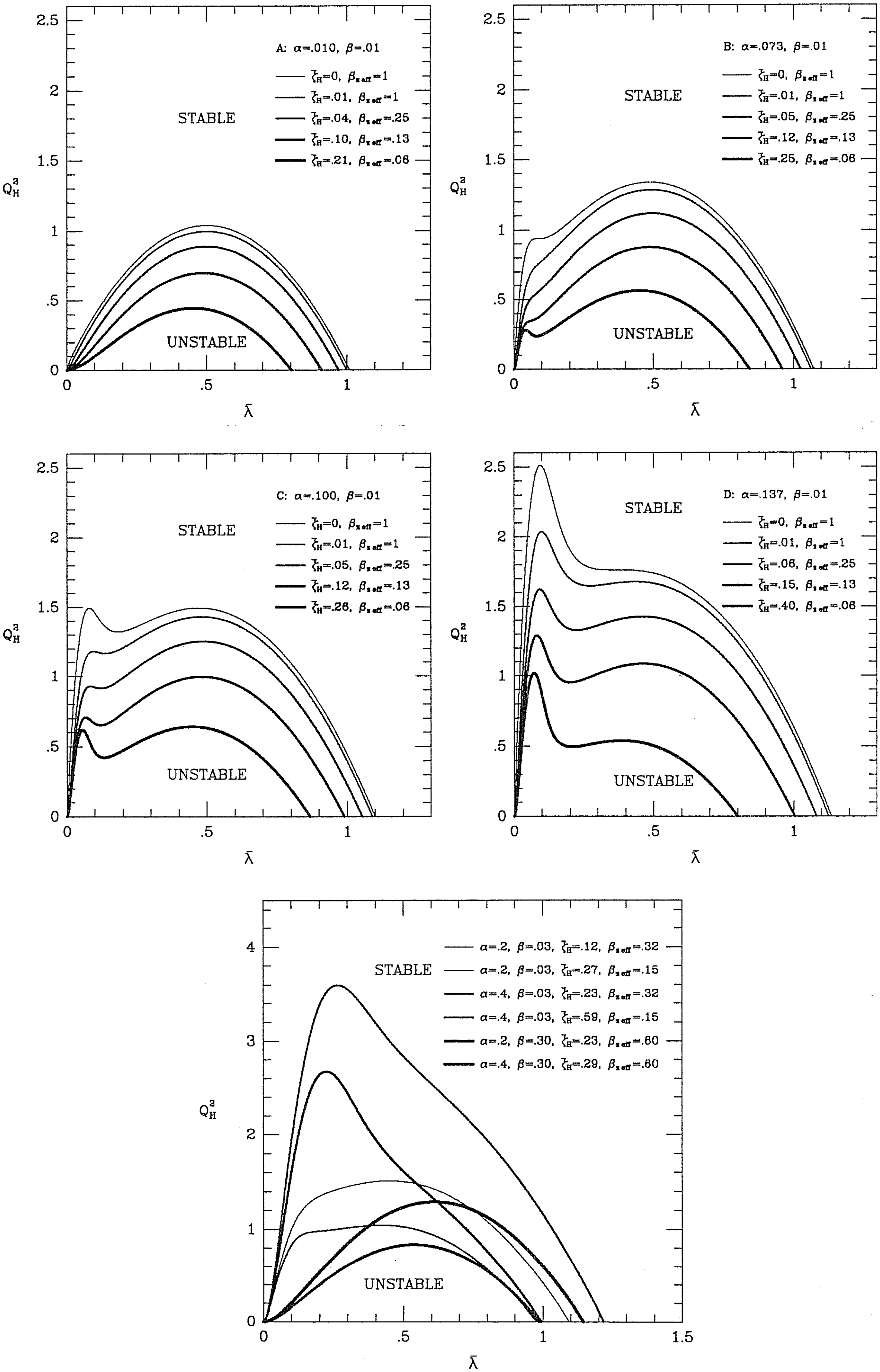


Figure 9.3



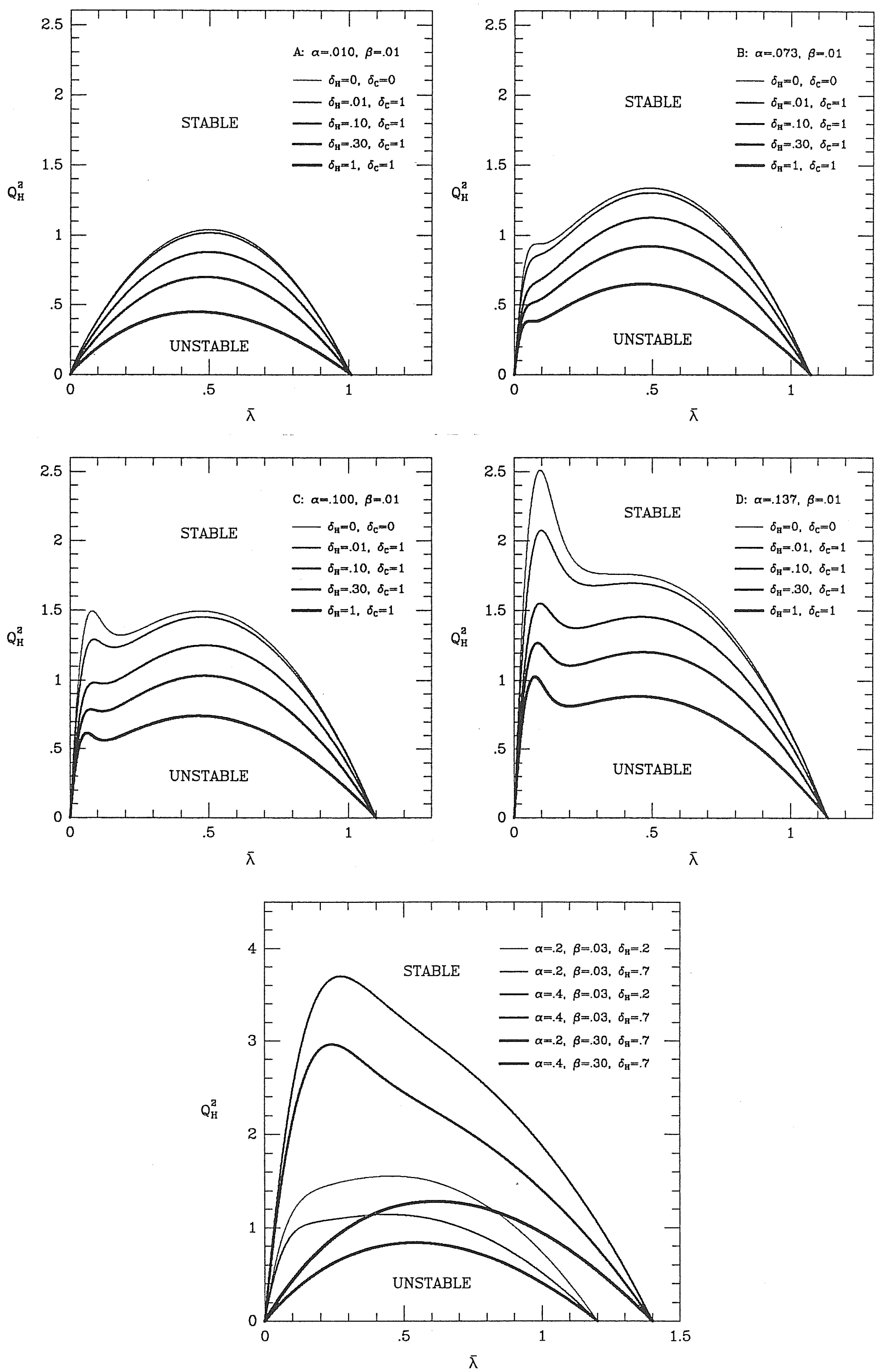
A: α=.010, β=.01
δH=0, δC=0
δH=.01, δC=1
δH=.10, δC=1
δH=.30, δC=1
δH=1, δC=1
STABLE
UNSTABLE
B: α=.073, β=.01
C: α=.100, β=.01
D: α=.137, β=.01
α=.2, β=.03, δH=.2
α=.2, β=.03, δH=.7
α=.4, β=.03, δH=.2
α=.4, β=.03, δH=.7
α=.2, β=.30, δH=.7
α=.4, β=.30, δH=.7
Q_H^2
$\bar{\lambda}$

Figure 9.4



A: α=.010, β=.01
ζH=0, βxeff=1
ζH=.01, βxeff=1
ζH=.04, βxeff=.25
ζH=.10, βxeff=.13
ζH=.21, βxeff=.06
STABLE
UNSTABLE
B: α=.073, β=.01
ζH=0, βxeff=1
ζH=.01, βxeff=1
ζH=.05, βxeff=.25
ζH=.12, βxeff=.13
ζH=.25, βxeff=.06
C: α=.100, β=.01
ζH=0, βxeff=1
ζH=.01, βxeff=1
ζH=.05, βxeff=.25
ζH=.12, βxeff=.13
ζH=.26, βxeff=.06
D: α=.137, β=.01
ζH=0, βxeff=1
ζH=.01, βxeff=1
ζH=.06, βxeff=.25
ζH=.15, βxeff=.13
ζH=.40, βxeff=.06
α=.2, β=.03, ζH=.12, βxeff=.32
α=.2, β=.03, ζH=.27, βxeff=.15
α=.4, β=.03, ζH=.23, βxeff=.32
α=.4, β=.03, ζH=.59, βxeff=.15
α=.2, β=.30, ζH=.23, βxeff=.60
α=.4, β=.30, ζH=.29, βxeff=.60
Q²H
λ̄

Figure 9.5



A: α=.010, β=.01
δH=0, δC=0
δH=.01, δC=1
δH=.10, δC=1
δH=.30, δC=1
δH=1, δC=1
STABLE
UNSTABLE
B: α=.073, β=.01
δH=0, δC=0
δH=.01, δC=1
δH=.10, δC=1
δH=.30, δC=1
δH=1, δC=1
STABLE
UNSTABLE
C: α=.100, β=.01
δH=0, δC=0
δH=.01, δC=1
δH=.10, δC=1
δH=.30, δC=1
δH=1, δC=1
STABLE
UNSTABLE
D: α=.137, β=.01
δH=0, δC=0
δH=.01, δC=1
δH=.10, δC=1
δH=.30, δC=1
δH=1, δC=1
STABLE
UNSTABLE
α=.2, β=.03, δH=.2
α=.2, β=.03, δH=.7
α=.4, β=.03, δH=.2
α=.4, β=.03, δH=.7
α=.2, β=.30, δH=.7
α=.4, β=.30, δH=.7
STABLE
UNSTABLE

Figure 9.6

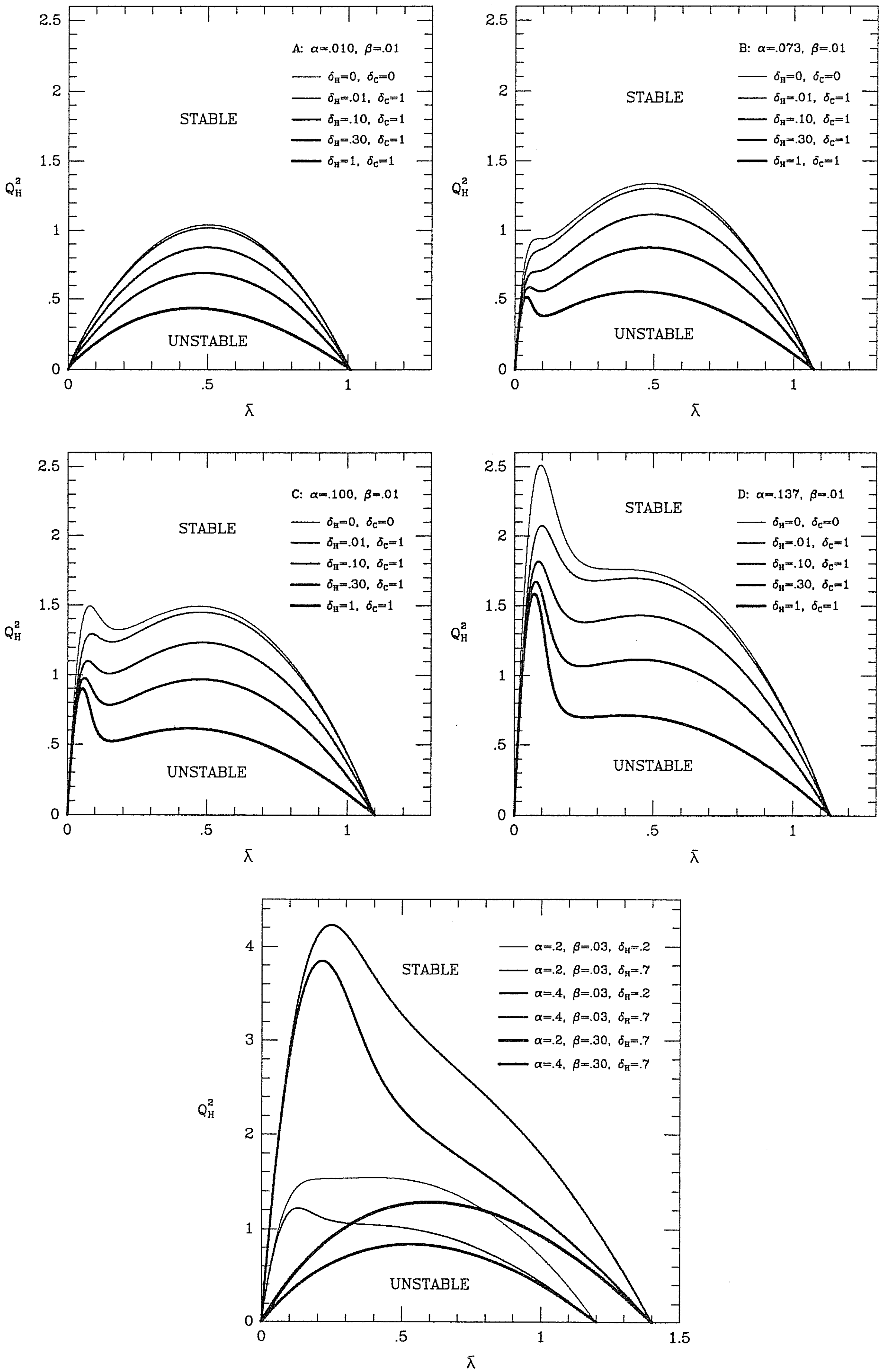
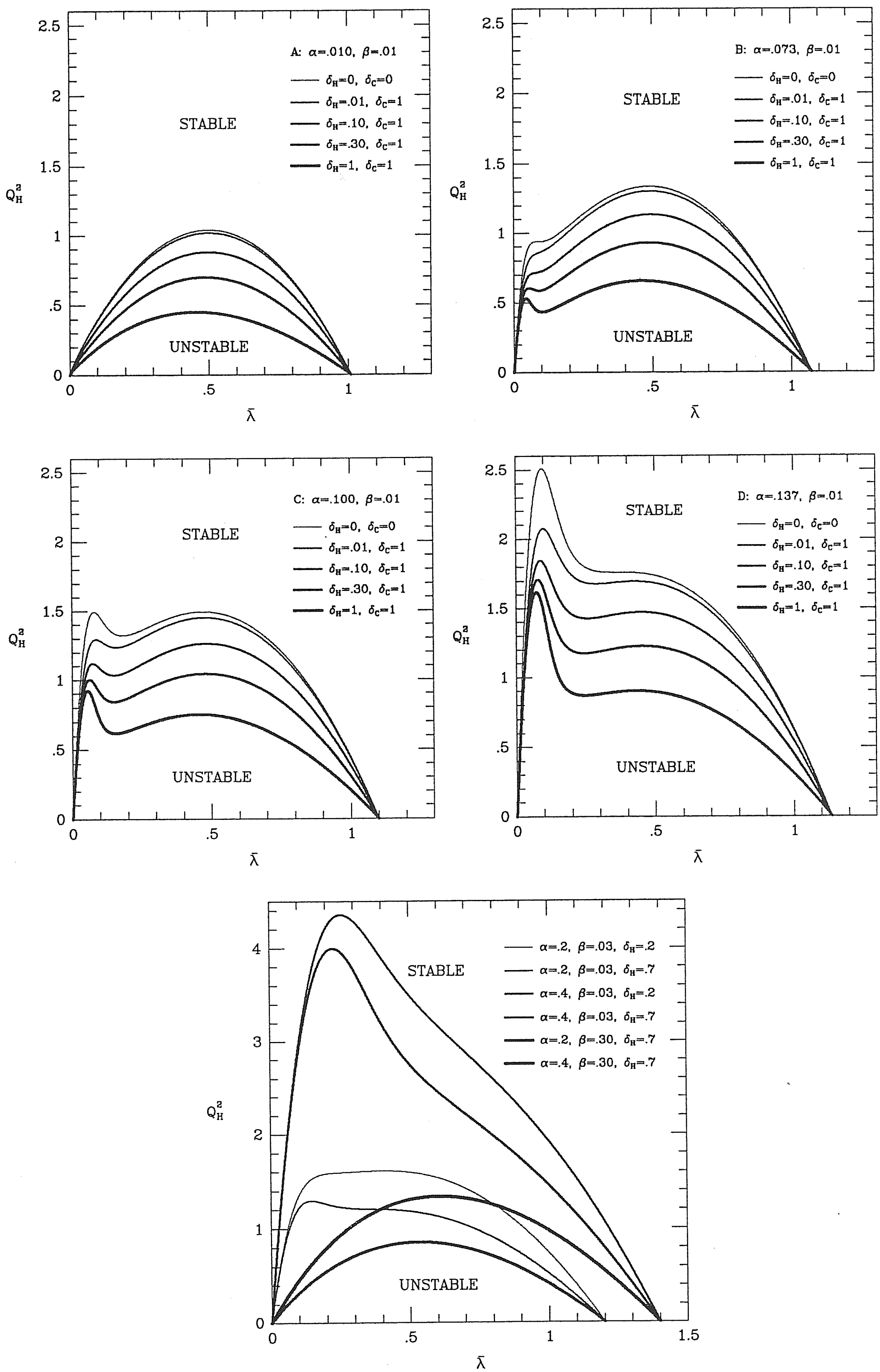

A: α=.010, β=.01
δH=0, δC=0
δH=.01, δC=1
δH=.10, δC=1
δH=.30, δC=1
δH=1, δC=1
STABLE
UNSTABLE
B: α=.073, β=.01
δH=0, δC=0
δH=.01, δC=1
δH=.10, δC=1
δH=.30, δC=1
δH=1, δC=1
STABLE
UNSTABLE
C: α=.100, β=.01
δH=0, δC=0
δH=.01, δC=1
δH=.10, δC=1
δH=.30, δC=1
δH=1, δC=1
STABLE
UNSTABLE
D: α=.137, β=.01
δH=0, δC=0
δH=.01, δC=1
δH=.10, δC=1
δH=.30, δC=1
δH=1, δC=1
STABLE
UNSTABLE
α=.2, β=.03, δH=.2
α=.2, β=.03, δH=.7
α=.4, β=.03, δH=.2
α=.4, β=.03, δH=.7
α=.2, β=.30, δH=.7
α=.4, β=.30, δH=.7
STABLE
UNSTABLE
Q²H
λ̄

Figure 9.7



A: α=.010, β=.01
δH=0, δC=0
δH=.01, δC=1
δH=.10, δC=1
δH=.30, δC=1
δH=1, δC=1
STABLE
UNSTABLE
B: α=.073, β=.01
C: α=.100, β=.01
D: α=.137, β=.01
α=.2, β=.03, δH=.2
α=.2, β=.03, δH=.7
α=.4, β=.03, δH=.2
α=.4, β=.03, δH=.7
α=.2, β=.30, δH=.7
α=.4, β=.30, δH=.7

Figure 9.8

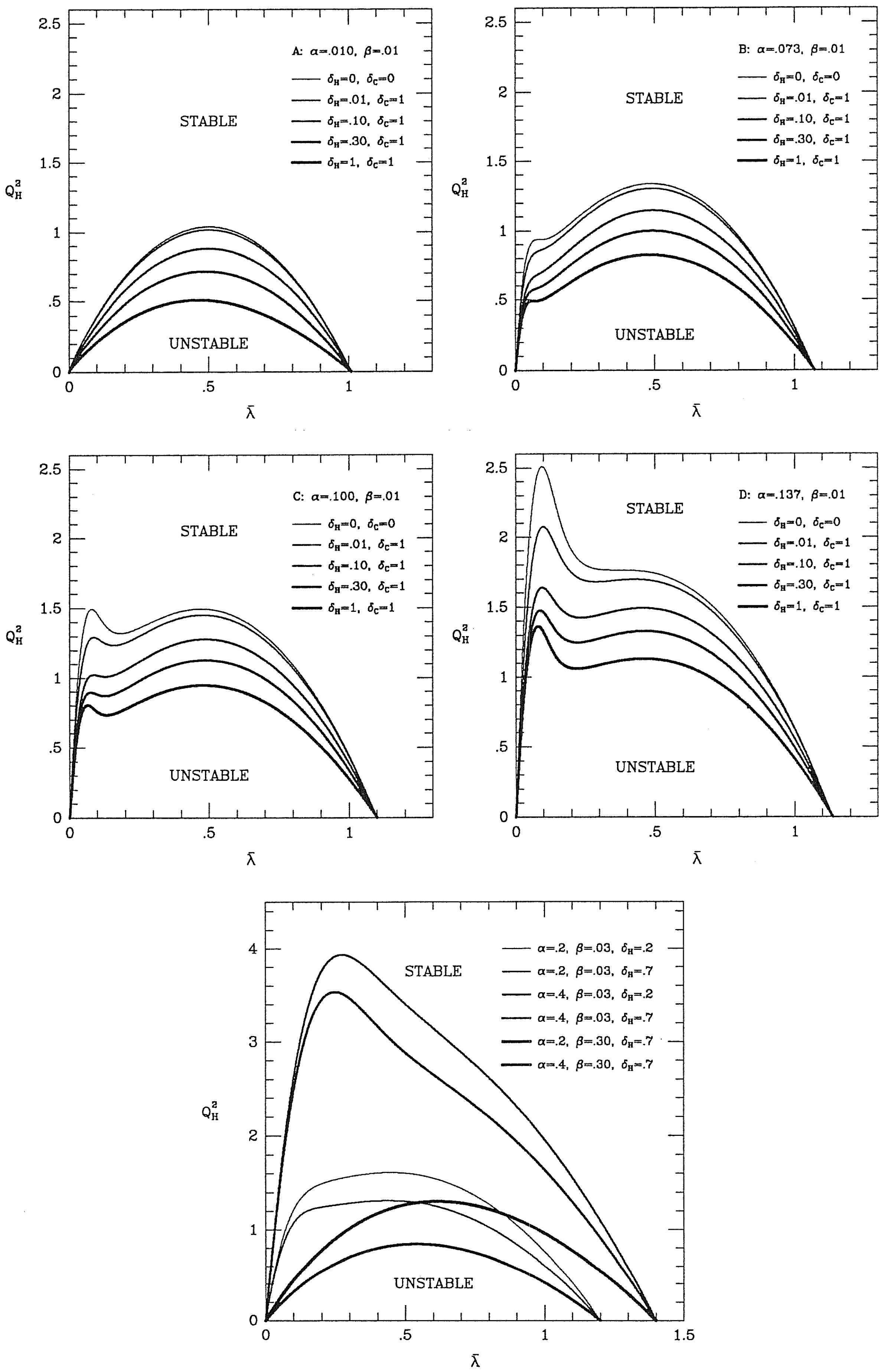
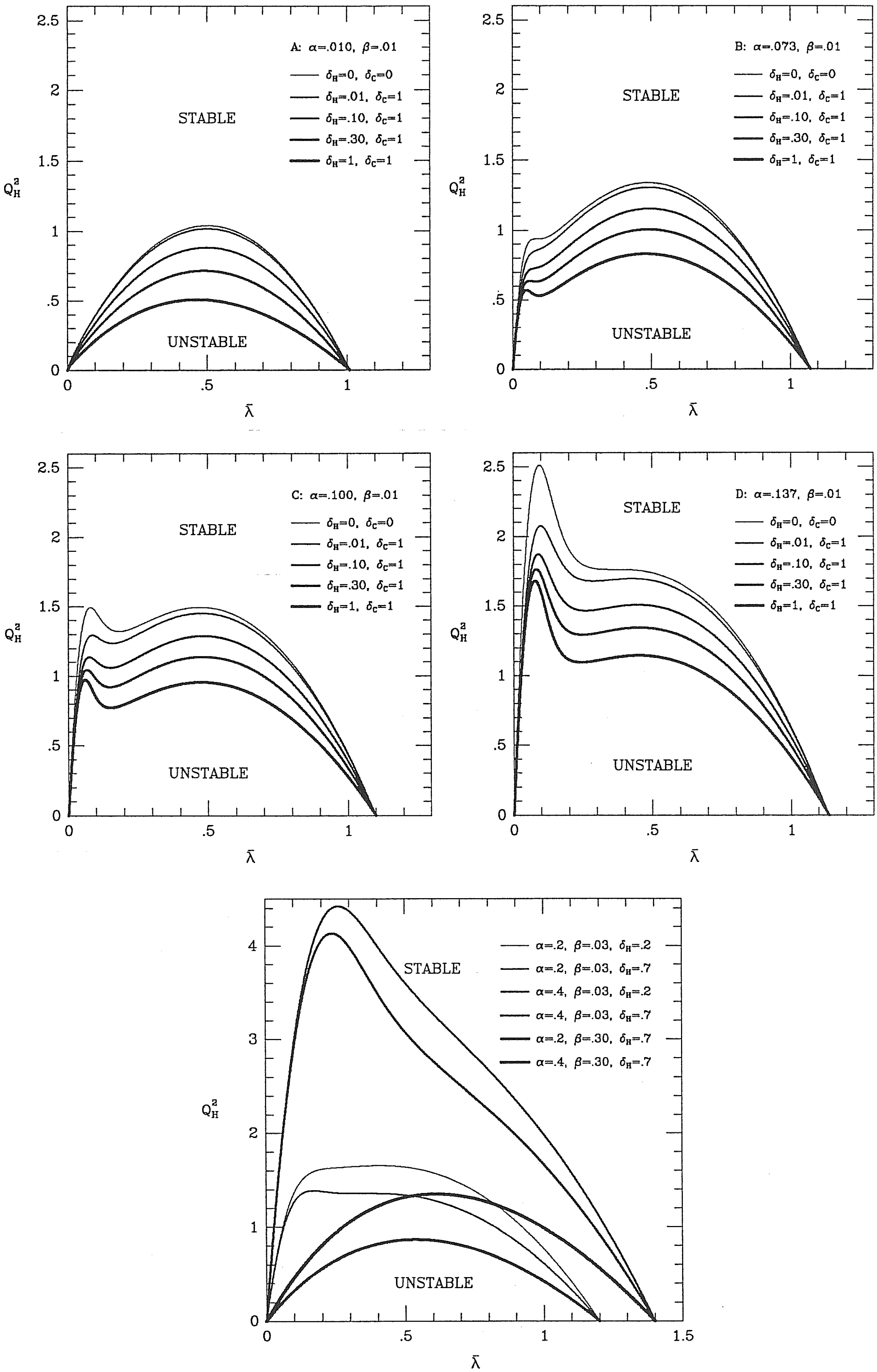

A: α=.010, β=.01
δH=0, δC=0
δH=.01, δC=1
δH=.10, δC=1
δH=.30, δC=1
δH=1, δC=1
STABLE
UNSTABLE
B: α=.073, β=.01
δH=0, δC=0
δH=.01, δC=1
δH=.10, δC=1
δH=.30, δC=1
δH=1, δC=1
STABLE
UNSTABLE
C: α=.100, β=.01
δH=0, δC=0
δH=.01, δC=1
δH=.10, δC=1
δH=.30, δC=1
δH=1, δC=1
STABLE
UNSTABLE
D: α=.137, β=.01
δH=0, δC=0
δH=.01, δC=1
δH=.10, δC=1
δH=.30, δC=1
δH=1, δC=1
STABLE
UNSTABLE
α=.2, β=.03, δH=.2
α=.2, β=.03, δH=.7
α=.4, β=.03, δH=.2
α=.4, β=.03, δH=.7
α=.2, β=.30, δH=.7
α=.4, β=.30, δH=.7
STABLE
UNSTABLE
Q_H^2
$\bar{\lambda}$

Figure 9.9


A: α=.010, β=.01
δH=0, δC=0
δH=.01, δC=1
δH=.10, δC=1
δH=.30, δC=1
δH=1, δC=1
STABLE
UNSTABLE
B: α=.073, β=.01
C: α=.100, β=.01
D: α=.137, β=.01
α=.2, β=.03, δH=.2
α=.2, β=.03, δH=.7
α=.4, β=.03, δH=.2
α=.4, β=.03, δH=.7
α=.2, β=.30, δH=.7
α=.4, β=.30, δH=.7
Q²H
λ̄
0
.5
1
1.5
2
2.5
3
4

Chapter 10

Concluding Remarks and Topics for Further Investigation

The analysis performed in this thesis represents the first step of a long-term project devoted to the study of self-regulation mechanisms, which are expected to drive the dynamical evolution of galactic discs and, in particular, to be responsible for their secular heating. A deep understanding of the problem which we have tackled can only be attained by a *scrupulous* investigation into the stability properties of galactic discs, taking into account effects which are generally neglected for mathematical convenience but which we have shown to be important or even *crucial*.

The analysis carried out in this thesis can surely be improved in the following two ways.

The good absolute accuracy of the fit approximation introduced in Chapter 8 is satisfactory only in standard star-dominated regimes. In peculiar gas-dominated regimes a more satisfactory relative accuracy is required. An *exact* determination of the local parameters α and $\beta_{z\,\mathrm{eff}}$ would thus be welcome.

Work is in progress for deriving more *accurate* estimates of the reduction factors of the two components, by employing a perturbative method to solve the double-eigenvalue problem for finite-thickness effects discussed in Chapter 9.

As a practical application of the analysis carried out in this thesis, the profiles of the local stability parameter Q_{H} for two-component models of the Galaxy will be derived in the framework of *global* spiral mode calculations and compared to those inferred on observational grounds. In particular, this extension might require a deep knowledge of the properties of the marginal stability curve and of the behaviour of its global maximum in the *peculiar* gas-dominated regimes characterizing the two-phase region, which at present is not available.

Only after that stage we could reasonably proceed to consider specific *local* models of self-regulation and to formulate a *global* quasi-linear theory of spiral structure, crowning our dreams.

Finally, the analysis carried out in this thesis to evaluate finite-thickness effects in two-component galactic discs is restricted to perturbations which propagate in the symmetry plane without affecting the vertical motion of the stars. The role of the hose-pipe instability associated with bending (even) waves should be estimated as well. An instability of this kind can in fact enhance the vertical component of the stellar velocity dispersion by collective effects (see, e.g., Kulsrud, Mark and Caruso 1972; Shlosman and Begelman 1989), and provide a welcome source of *vertical* heating, as suggested in Chapter 6.

References

Kulsrud R. M., Mark J. W.-K. and Caruso A. (1972), in *IAU Coll.* **10**, Reidel, Dordrecht, Lecar M. ed., p. 180 (Cambridge, UK, 1970).

Shlosman I. and Begelman M. C. (1989), *Ap. J.* **341**, 685.