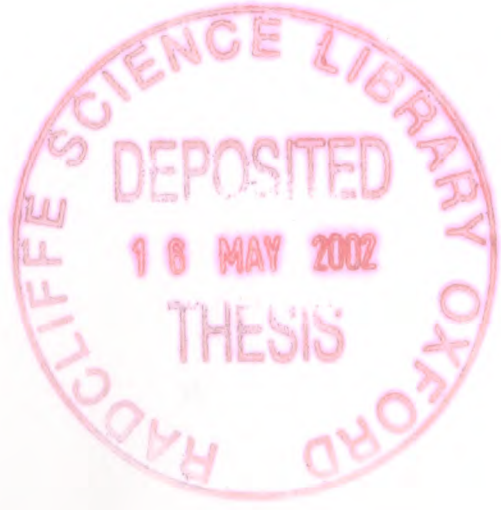
Simulations of Stars and Gas in Spiral Galaxies

Simon Lewis Berman
Merton College



Theoretical Physics
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy at the University of Oxford



Simulations of Stars and Gas in Spiral Galaxies

Simon Lewis Berman
Theoretical Physics, Department of Physics, 1 Keble Road, Oxford, OX1 3NP

A thesis submitted for the degree of Doctor of Philosophy
January 2002

Abstract

A new code, DUAL, is developed to model the 2D dynamics of stars and gas in spiral galaxies. DUAL combines grid-based hydrodynamics and N-body techniques.

A triaxial model of the bulge of M31 is constructed from its surface brightness profile and shown, using hydrodynamic simulations, to reproduce observations of molecular gas kinematics along the line of nodes of the disk. The bulge model rotates fast, with a pattern speed of 54 km s^{-1} kpc^{-1}, and a ratio of bulge semi-major axis to corotation radius $\mathcal{R} = 1.2$. The B band mass-to-light ratio is 6.5, the semi-major axis is 3.5 kpc, the bulge mass is 2.3×10^{10} $M_\odot$ and the angle between the projected minor axis of the bulge and the line of nodes of the disk is 15°.

A thick gas disk (HWHM of 200 → 500 pc) is added to account for CO observations parallel to the disk line of nodes. The thickness is minimized if the gas velocity drops by 20% per kpc above the plane. By combining reasonable values of the disk mass-to-light ratio or dynamical friction arguments with the bulge model, M31 is shown to have a maximal disk with a halo mass fraction within 3.5 kpc of just a few percent. In contrast, the Cold Dark Matter theory predicts a fraction of 22% to 30%.

The differences between optical and infrared morphologies of spiral galaxies are reproduced using hydrodynamic and N-body simulations of gas and stars. Gaseous spirals have tighter pitch angles than stellar spirals. This phenomenon is attributed to a phase shift between the stellar density and the potential. Gaseous images are more asymmetric, less smooth and more likely to have multiple arms. Morphological decoupling increases as the stellar arm-interarm contrast and Q parameter fall.

The flocculence of a galaxy is quantified by decomposing the galaxy images into logarithmic spirals and defining a parameter closely related to the uniformity of the resulting 2D Fourier spectrum.

I conclude that a significant amount of morphological decoupling in spiral galaxies is due to the difference in the dynamics of stars and gas, rather than dust, star formation or galaxy interactions.

Acknowledgments

I am extremely grateful to my supervisor, James Binney, for providing me with a path and helping me learn how to follow it. As a result of many discussions, I have realized the importance of rigorous and clear thinking, of relating scientific theory to reality, and of maintaining an overall vision.

Life in Theoretical Physics would not have been as fun and as bearable without the presence of a number of fine people. Wyn Evans provided much useful advice about the realities of astronomy and the role of graduate students within it. Alex was a constant source of support, if only to empathize with the gruelling task of writing a hydro code. On many occasions, he provided a sympathetic ear and a helpful response to my tales of bugs and compiler errors. And it was great to be part of each other's wedding celebrations. Much needed computing advice on getting to grips with UNIX, LaTeX and IDL was provided by Mark, and welcome hydrodynamic support, career advice (and Polish cabbage) came from Witold. Many, many hours of coffee time were enjoyed in the company of Ralf, Graeme and Stuart - may we have many more - and the comrades already mentioned. Tet, Serge, Jonty, Vasily and Ralf all made Room 2.9 a fun place to be and Walter and Prasenjit were both sources of much wisdom.

Jake, Heski and Paul made Shabbat in Oxford a real pleasure. Jake and Heski also, along with Oli, Dave and Gordon, kept me in shape on the squash court. One day I'll beat you guys! It has been really touching that so many friends and family have come to visit me and Juliet in Oxford. I hope they enjoyed being here as much as we enjoyed seeing them. In particular, my parents (and in-laws) have been a source of great strength - I hope that this thesis gives them, and Nana, much naches.

Thanks are due to Jerry Sellwood for providing me with the N-body code which, in an amended form, became half of the DUAL code. Thanks also to Dick van Albada and Carsten Arnholm who wrote the FS2 algorithm at the heart of the hydrodynamics and the code enabling C++functions to read Fortran, respectively. Ivanio Puerari generously supplied me with 2DFFT, the code which generates the Fourier spectra of Chapter 5. In the same chapter, David Block, also generously, allowed me to use his B and K' band data for NGC 2997, 3223 and 5085 and Walter Dehnen supplied the code to generate the stellar distribution function, for which I am very grateful. In Chapter 4, the calculations of the dark matter fraction in M31 utilize a code written by Julio Navarro and the CO data were collected by Loinard et al. (1995). Many further thanks are due to Laurent Loinard for stimulating discussions concerning the work in Section 4.4.2, which we have now submitted to MNRAS.

Financial support from Merton College and Theoretical Physics has enabled me to benefit from attending and presenting at a number of conferences, and for that I am truly grateful.

It has been a very difficult journey from arriving in Oxford to finishing this thesis. Although there have been some very good times, there have been many occasions when I did not think that I would get to the end. Through all of this, Juliet has put up with, supported, encouraged and inspired me, enabling me to complete the journey and be proud of it. On top of that, she has done a brilliant job in proof reading the entire text, although I remain responsible for any lingering errors. Juliet is both my wife and my best friend and I dedicate the contents of this work to her.

Contents

1 Introduction **1**
- 1.1 Spiral galaxies . . . 2
- 1.2 Bars and bulges . . . 2
  - 1.2.1 Triaxial bulges . . . 2
  - 1.2.2 Pattern speeds . . . 4
- 1.3 Dark matter . . . 5
- 1.4 The Andromeda galaxy . . . 7
  - 1.4.1 Observations . . . 7
  - 1.4.2 Mass models of the bulge of M31 . . . 7
- 1.5 Dual morphology spiral galaxies . . . 9
  - 1.5.1 Observations . . . 9
  - 1.5.2 Dust penetrated classification . . . 10
  - 1.5.3 Methods of modelling stars and gas in spiral galaxies . . . 10
- 1.6 Layout of the thesis . . . 12

2 Theory **13**
- 2.1 Hydrodynamics . . . 13
  - 2.1.1 The ISM as a 2D isothermal fluid . . . 13
  - 2.1.2 The Euler equations, Riemann problems and shocks . . . 13
  - 2.1.3 Computational fluid dynamics . . . 19
  - 2.1.4 The FS2 algorithm . . . 25
- 2.2 Gas dynamics in barred galaxies . . . 28
  - 2.2.1 Regular orbits . . . 28
  - 2.2.2 Previous simulations of gas dynamics in barred galaxies . . . 30
- 2.3 N-body theory . . . 32
  - 2.3.1 Stellar dynamics as an N-body problem . . . 32
  - 2.3.2 Softened gravity . . . 33
  - 2.3.3 Particle-mesh codes . . . 34
  - 2.3.4 Previous N-body simulations of spiral galaxies . . . 36
- 2.4 Galactic stability . . . 37

3 Description of the DUAL code **39**
- 3.1 Conceptual overview . . . 39
- 3.2 Structural details . . . 40
  - 3.2.1 The source code files for DUAL . . . 40
  - 3.2.2 Using C++ . . . 40
  - 3.2.3 Installation and the makefile . . . 41
  - 3.2.4 Miscellaneous functions . . . 42
- 3.3 Hydrodynamics . . . 43
  - 3.3.1 The dynamic variables - class `Node` . . . 43
  - 3.3.2 Managing the grid, the nodes and the time - class `Grid` . . . 44
  - 3.3.3 The FS2 algorithm - class `FS2` . . . 45
  - 3.3.4 A hydrodynamic sweep - class `Sweep` . . . 47
  - 3.3.5 Defining the gravitational potential - class `TwoDimPot` . . . 49

3.3.6 Including source terms - class `Source` . . . 50
3.3.7 Initializing the gas density and velocity - class `Setup` . . . 52
3.3.8 Ensuring that variables are dimensionless - class `Dim` . . . 53
3.3.9 The `main` programme . . . 53
3.4 The N-body solver . . . 56
3.4.1 Overview . . . 56
3.4.2 Dimensions and time step . . . 56
3.4.3 Selected routines and data structures . . . 57
3.4.4 Amendments to the N-body solver . . . 58
3.4.5 The C++ Fortran interface . . . 59
3.4.6 The `main` programme . . . 60
3.5 Combining hydrodynamic and N-body solvers . . . 61
3.5.1 Hydrodynamic functions which access N-body routines . . . 61
3.5.2 The `main` programme . . . 63
3.6 Particular implementations of DUAL . . . 66
3.6.1 Modelling the Andromeda galaxy . . . 66
3.6.2 Modelling stellar and gaseous morphology . . . 67
3.7 Tests and performance . . . 67
3.7.1 Axisymmetry test . . . 67
3.7.2 Shocktube tests . . . 67
3.7.3 Oscar and parallelization . . . 69

4 Hydrodynamic simulations of the bulge of M31 72
4.1 Observational motivation . . . 72
4.2 Modelling M31 . . . 75
4.2.1 Triaxial bulge . . . 75
4.2.2 Axisymmetric component . . . 78
4.2.3 Hydrodynamic parameters . . . 80
4.3 Simulation results . . . 81
4.3.1 Selection of models . . . 81
4.3.2 Gas flow in selected models . . . 82
4.4 Comparison with observations . . . 85
4.4.1 Observations along the line of nodes of the disk . . . 85
4.4.2 Observations away from the line of nodes of the disk . . . 88
4.4.3 Best model dust map . . . 94
4.5 The dark halo of Andromeda . . . 96
4.5.1 Predictions of Cold Dark Matter . . . 97
4.5.2 Implications of fast bars . . . 98
4.6 Conclusions . . . 99

5 Simulations of dual morphology in spiral galaxies 102
5.1 Observational and theoretical motivation . . . 102
5.1.1 Summary of observations . . . 102
5.1.2 Do optical and infrared images trace separate populations? . . . 103
5.1.3 Dual morphologies within the modal framework . . . 106
5.2 Modelling dual morphology galaxies . . . 108

5.2.1 The stellar disk . . . 109
5.2.2 The dark halo . . . 110
5.2.3 Stellar distribution function . . . 111
5.2.4 Stability and rotation . . . 112
5.2.5 The gaseous disk . . . 113
5.2.6 Simulation run time . . . 114
5.3 Analysis techniques . . . 115
5.3.1 Animating the results . . . 115
5.3.2 Harmonic analysis . . . 115
5.3.3 Flocculence . . . 121
5.3.4 Asymmetry . . . 124
5.3.5 Radial Fourier analysis . . . 124
5.3.6 Arm-interarm contrast . . . 124
5.3.7 Monitoring $Q(t)$ . . . 125
5.3.8 Pattern speed and resonances . . . 126
5.4 Simulation results . . . 127
5.4.1 Computational checks . . . 127
5.4.2 Time evolution of simulations . . . 128
5.4.3 Do these models reproduce the observations? . . . 130
5.4.4 Can these models predict galactic structure parameters? . . . 134
5.4.5 Comparison with observations . . . 135
5.5 Discussion . . . 136
5.5.1 The pitch angles of gaseous and stellar spirals . . . 137
5.5.2 The flocculence of gaseous and stellar spirals . . . 138
5.5.3 Intrinsic characteristics of galaxies . . . 139
5.6 Conclusions . . . 140

6 Conclusion **142**

A The $\mathcal{X}$ sweep of the FS2 algorithm **157**
A.1 Overview . . . 157
A.2 Evolving the cell centred values . . . 157
A.3 Reconstructing the intra-cell data . . . 158
A.4 Calculating the cell boundary values . . . 159
A.5 Determining the flux splitting function . . . 159

B Description of models of the bulge of M31 **162**

C Description of dual morphology models **174**

List of Tables

3.1 The directory structure for the DUAL source code. 40

4.1 Observed parameters of M31 73
4.2 Effect of changing ϕ on the bulge model 77
4.3 Hydrodynamic and computational parameters 81
4.4 The bulge model of maximum likelihood 82

5.1 Values of canonical parameters in dual morphology models 128

B.1 Description of all bulge models 162

C.1 Analysis of real galaxy images 174
C.2 Analysis of all simulations 191

List of Figures

2.1 Shock creation through wave steepening 15
2.2 Analytic result for 1D isothermal shocktube 16
2.3 Finite difference approximations to the first derivative of $f(x)$ 20
2.4 Discretization of the computational domain onto a grid 21
2.5 Splitting of the flux component within a cell 23
2.6 x_1 and x_2 orbits in a barred potential 29
2.7 Relationship between the softening length and the resulting Plummer length . . . 35
2.8 The leap frog time integration scheme. 37

3.1 A schematic of the amended FS2 algorithm 46
3.2 Calling tree for `SweepGrid` functions 48
3.3 Calling tree for `main` hydrodynamics programme 55
3.4 Calling tree for `main` N-body programme 62
3.5 Calling tree for `main` combined programme 64
3.6 Numerical result for 1D isothermal shocktube test 68
3.7 Numerical result for oblique 1D isothermal shocktube test 68
3.8 Fork-join parallelization model 69
3.9 Performance of DUAL with number of processors 71

4.1 Isophotal map for M31 74
4.2 Position-velocity diagram along the line of nodes of M31 74
4.3 Orientation of the bulge of M31 77
4.4 Circular and angular speeds of the model along the major and minor axes 79
4.5 Velocity and density fields for the best fit model 83
4.6 Comparison of different models' velocity fields 84
4.7 Observed and modelled velocities along the line of nodes of M31 86
4.8 Confidence regions for $\mathcal{R} = 1.0$ and 1.2 87
4.9 Confidence regions for $\mathcal{R} = 1.4$ 87
4.10 Observations of CO velocities parallel to the line of nodes of the disk 89
4.11 Bulge model including a corotating thick gas disk 92
4.12 Bulge model including a corotating thick gas disk with a vertical velocity gradient 93
4.13 Velocity field for the model of the bulge of M31 94
4.14 Absorption map for the model of the bulge of M31 95
4.15 Comparison of the central region of the absorption map with 7.7 μm observations 95

5.1 K' and B band observations of NGC 5195 103
5.2 K' and B band observations of NGC 309 104
5.3 K' and B band observations of NGC 5055 104
5.4 Classifying spiral galaxies on the basis of their intrinsic characteristics 108
5.5 Radial profile of the halo mass fraction for four different disk mass fractions . . . 111
5.6 Analytic and particle radial Q profile 113
5.7 Comparison of model and observed rotation curves 114
5.8 Example of a logarithmic spiral. 116
5.9 Stellar and gaseous density fields and harmonic spectra for a grand design spiral . 118
5.10 Stellar and gaseous density fields and harmonic spectra for a multi armed spiral . 119

5.11 Stellar and gaseous density fields and harmonic spectra for a flocculent spiral . . . 120
5.12 Harmonic analysis for test cases . . . 123
5.13 Radial variation of arm-interarm contrast . . . 125
5.14 Evolution of $Q(t)$ . . . 126
5.15 Position of all simulations within the parameter space . . . 127
5.16 Time evolution of density fields and Fourier spectra for `sta19` and `gas19` . . . 129
5.17 Pitch angles, flocculence and asymmetry for all simulations . . . 131
5.18 Correlation of time averaged arm-interarm contrast with gaseous flocculence and gaseous asymmetry . . . 134
5.19 Correlation of disk mass fraction with bar size and Q with stellar arm-interarm contrast . . . 135
5.20 Position of the $m = 2$ maxima for the grand design model . . . 138

B.1 Comparison with CO observations on major axis for all M31 bulge models . . . 166

C.1 Density fields and Fourier spectra for `sta13` and `gas13` . . . 175
C.2 Density fields and Fourier spectra for `sta16` and `gas16` . . . 176
C.3 Density fields and Fourier spectra for `sta17` and `gas17` . . . 177
C.4 Density fields and Fourier spectra for `sta19` and `gas19` . . . 178
C.5 Density fields and Fourier spectra for `sta21` and `gas21` . . . 179
C.6 Density fields and Fourier spectra for `sta26` and `gas26` . . . 180
C.7 Density fields and Fourier spectra for `sta28` and `gas28` . . . 181
C.8 Density fields and Fourier spectra for `sta30` and `gas30` . . . 182
C.9 Density fields and Fourier spectra for `sta36` and `gas36` . . . 183
C.10 Density fields and Fourier spectra for `sta38` and `gas38` . . . 184
C.11 Density fields and Fourier spectra for `sta40` and `gas40` . . . 185
C.12 Density fields and Fourier spectra for `sta42` and `gas42` . . . 186
C.13 Density fields and Fourier spectra for `sta43` and `gas43` . . . 187
C.14 Density fields and Fourier spectra for `sta44` and `gas44` . . . 188
C.15 Density fields and Fourier spectra for `sta45` and `gas45` . . . 189
C.16 Density fields and Fourier spectra for `sta46` and `gas46` . . . 190

Chapter 1

Introduction

Galaxies are the building blocks of the universe and spiral galaxies are the good looking ones. From grand design to flocculent spirals their majestic beauty is inspirational.

Galactic beauty, however, is only skin deep. There is much more to spiral galaxies than is apparent from simple observations in optical light. Optical observations are dominated by young stars and gas which may constitute only 5% of the dynamical mass of a galaxy. The majority of the baryonic mass of a galaxy is revealed in infrared light, but even then half of the galaxy's total mass may still be hidden in the form of the mysterious dark matter.

The aim of this thesis is to address these issues by investigating whether mass follows light in spiral galaxies. As such, a hydrodynamics and N-body code capable of modelling the 2D dynamics of gas and stars in a spiral galaxy is developed. The code will be used to shed light on those areas of astronomy which are unlit in the optical regime. In particular,

- to constrain the shape and dark matter content of the centre of M31 by reproducing observations of molecular gas kinematics,
- to predict intrinsic characteristics of spiral galaxies by reproducing observations of their dual morphology in the optical and infrared regimes.

In this Chapter, I review our current knowledge of spiral galaxies paying particular attention to the features most relevant to the studies of M31 and dual morphologies. This review is necessarily brief and mainly historical in nature. In Section 1.1 I describe features common to most spiral galaxies. In Sections 1.2 and 1.3 I discuss those features most pertinent to this thesis, namely bar-like components and dark matter. Previous models of M31 are detailed and compared in Section 1.4, whilst observations of dual morphologies and methods of modelling the dynamics of both stars and gas are reviewed in Section 1.5. Section 1.6 outlines the content of the rest of this thesis.

1.1 Spiral galaxies

Spiral galaxies are a beautiful and heterogeneous subset of the galaxies in the universe. They range from giants to dwarfs and from tightly wound, bulge dominated Sa galaxies to the loose spirals of Sc galaxies with almost no bulge at all. They might contain breathtaking grand design two arm spiral structures like M51, or ragged feathery flocculent structures like M63. And most of these spiral galaxies have a number of morphological features in common: the disk and the eponymous spiral arms, a central bulge, a dark halo and possibly a bar.

Current thinking suggests that our own Galaxy is an S(B)bc galaxy. This means that it is probably barred, has a small bulge, fairly open spiral arms and a significant quantity of gas by mass. Its disk consists of thin and thick components. The thin disk has a scale height of 280 pc and consists of young metal rich stars following an exponential surface brightness distribution $I(R) = I_0 \exp(-R/R_d)$, where $R_d = 2.8$ kpc is the disk scale length (Spergel et al. 1996). The thick disk has a scale height of ~ 1 kpc and consists of older and more metal poor stars (Binney and Merrifield 1998).

The interstellar medium (ISM) is composed of gas, dust, molecular clouds, cosmic rays and magnetic fields making a very complicated physical system. The gas can be divided into a number of components: hot ionized (10^6 K, 70% of ISM by volume), warm ionized (10^4 K), warm atomic (10^3 K, 20% of ISM by volume), cool atomic clouds (10^2 K) and cool molecular clouds ($10 - 30$ K). The clouds make up most of the mass of the ISM. The ISM itself is only a small fraction of the whole galaxy by mass but forms an essential component in terms of the evolution of a galaxy, being the birthplace and deathbed of its constituent stars and regulating their dynamics during their lifetime. 21 cm observations of atomic hydrogen show that gaseous disks are much more extended than stellar disks with a scale length more than three times as large.

1.2 Bars and bulges

1.2.1 Triaxial bulges

For many years, there has been ample evidence that bulges of barred spiral galaxies (SB bulges) are triaxial (Kormendy 1982). A spiral galaxy is not able to have an oblate spheroidal bulge if

the position angles of the major axes of the bulge, bar and disk are all different, unless the bulge is tipped with respect to the plane of the disk. Given that observations of edge-on galaxies show that bulges are rather flat (Kormendy 1981), it is likely that the shortest axis is perpendicular to the plane of the disk.

Using stellar kinematic data, Kormendy (1982) provide evidence that many SB bulges (in barred galaxies) are rotationally flattened and consistent with being oblate spheroids. A similar conclusion is reached by Kormendy and Illingworth (1982) for SA bulges (in unbarred galaxies). There is, however, photometric and kinematic evidence that bulges in unbarred galaxies are triaxial. In a similar way to barred galaxies, a misalignment between the apparent major axes of the bulge and the disk is an indicator of triaxiality, as is a twisting of the inner bulge isophotes with respect to those of the disk. These effects have been seen by Bertola et al. (1988) in NGC 2784 and by Zaritsky and Lo (1986) in a sample of 10 unbarred galaxies.

Stronger evidence for triaxiality comes from combining photometry with kinematic data. In a non-axisymmetric potential, the shape of the rotation curve depends on the position of the line of sight with respect to the major axis of the non-axisymmetric component. Anomalously high central gas velocities are indicative of triaxiality. By comparing the theoretical gas velocity field derived from a triaxial potential with the observed kinematic data, one can determine whether the assumption of triaxiality is justified. In this way, triaxiality has been demonstrated for NGC 4845 (Gerhard et al. 1989) and for our Galaxy (Gerhard and Vietri 1986).

It is important to know whether triaxiality is common in SA bulges, since any asymmetries in the underlying potential of the inner galaxy will have a large effect on the surface density distribution of interstellar gas. In a triaxial potential, the gas gets concentrated around features associated with resonances, such as rings, dust lanes, and nuclear spirals. These features may also provide the means of transporting gas into galaxy cores. Hence, triaxial bulges could provide clues to the fueling of central starbursts and active galactic nuclei. Furthermore, large scale movement of gas within a spiral galaxy could lead to changes in the overall morphology. In particular, a build up of gas in the galactic centre can result in either the creation of a bar or in its destruction.

1.2.2 Pattern speeds

The pattern of a bar or triaxial bulge rotates as a solid body in contrast to the differential rotation of the disk. Given the observational nature of astronomy and the vast time scales involved, it is not possible to directly observe the pattern speed. But it is a very important quantity to understand. A triaxial bar or bulge is often the largest contribution to the non-axisymmetric potential of a galaxy. Knowing the pattern speed enables models to be built of the stellar and gaseous dynamics of a galaxy. Furthermore, the pattern speed can be used to pin down the extent of any dark halo that may be present via the techniques of dynamical friction and velocity field modelling.

Although the pattern speeds of bars and bulges have proved to be difficult to determine for individual galaxies, Kormendy (1981) reviewed evidence from N-body models which showed that bars rotate rapidly. Athanassoula (1992b) performed extensive hydrodynamic simulations showing that the observed morphology of dust lanes in barred galaxies could only be reproduced if the ratio of the bar's corotation radius to its length was $\mathcal{R} = 1.2 \pm 0.2$ i.e. bars are rapid rotators.

Pattern speeds have been calculated for a number of individual galaxies, generally for those with available 2D kinematic or photometric data. The Tremaine-Weinberg method (Tremaine and Weinberg 1984) derives the pattern speed of a flat galactic disk directly from surface brightness and radial velocity observations of a tracer which is assumed to obey the continuity equation i.e. the tracer must be neither created nor destroyed. Amongst others who have used this method, Merrifield and Kuijken (1995) find that $\mathcal{R} = 1.4 \pm 0.3$ for NGC 936, Gerssen et al. (1999) find that $\mathcal{R} = 1.15^{+0.38}_{-0.23}$ for NGC 4596, Debattista and Sellwood (2000) find that $\mathcal{R} = 0.9 \pm 0.1$ for NGC 7079 and Debattista et al. (2001) find that $\mathcal{R} = 0.91 \pm 0.12$ for NGC 1023.

Hydrodynamical simulations have been performed for certain galaxies in which the gas response to a background stellar potential is matched to observed gas kinematics. Since the pattern speed is used as an input parameter of the stellar potential, it is possible to determine how fast the bars or bulges in these galaxies rotate. Simulations that have been run include Lindblad and Kristen (1996) and Lindblad et al. (1996) who found that $\mathcal{R} = 1.3$ for NGC 1300 and $\mathcal{R} = 1.3$ for NGC 1365 respectively and Weiner and Sellwood (1999) and Weiner et al. (2001) who calculated that the bars of both NGC 4123 and our Galaxy are fast rotators with $\mathcal{R} = 1.2$ and $\mathcal{R} = 1.4$

respectively.

Our Galaxy has also been modelled by Dehnen (1999b) using stellar velocities derived from Hipparchos data. He showed that the bi-modal distribution of stellar velocities in the solar neighbourhood can be accounted for by the Galactic bar if the Outer Lindblad Resonance is located about one kpc inside the solar radius. Combining this with other observational constraints implies that the Galactic bar is fast, with $\mathcal{R} = 1.25 \pm 0.2$.

1.3 Dark matter

The mass of the Galaxy as measured by the kinematics of bound dwarf galaxy satellites and globular clusters is $2 \times 10^{12} M_{\odot}$ (Wilkinson and Evans 1999). This is much larger than the total of the luminous matter in the disk, bulge and spheroid, which gives a mass of $9 \times 10^{10} M_{\odot}$. Hence, within the bounds of the Galaxy we are lead to one of two conclusions: either Newtonian gravity is in need of modification or there exists a large mass of dark matter which only makes its presence felt via gravitational interactions.

Modified Newtonian dynamics or MOND (Milgrom 1983) has been successful in reproducing rotation curves and other galactic scale observations by amending Newtonian gravity in the limit of low accelerations. But MOND does not provide a complete description of N-body dynamics, thus limiting its applicability to cosmology and galaxy formation studies.

Disregarding MOND, there is strong evidence for the presence of dark matter halos in other galaxies. van Albada et al. (1985) present observations of H I for NGC 3198. They find that the rotation curve is roughly flat out to at least 11 disk scale lengths, by which point the rotation curve for an exponential disk alone would have fallen to just half of its maximum. For the highest disk mass consistent with a halo density that decreases monotonically with galactic radius (the maximum disk solution), at 11 disk scale lengths the halo mass would be four times that of the disk.

Many cosmological N-body simulations have shown that cold dark matter (CDM) – collisionless particles interacting only through gravity – can be used as a successful description of large scale structure in the universe. CDM has problems on galactic scales, however, in providing a solution

to the missing mass problem. In particular, simulated dark halo density profiles (e.g. NFW profiles of Navarro et al. (1996)) are found to be much steeper than implied by the central parts of the rotation curves of spiral galaxies, there are many more satellites found around CDM halos than near to our own Galaxy, and simulated disks are generally much smaller than real galaxies disks.

There have been many attempts to alter or add to the basic CDM scenario to ensure that the idea of dark matter becomes consistent with observations. I mention only two. Self-interacting dark matter (Spergel and Steinhardt 2000) may remove the inner cusps of dark halos and the over abundance of satellite galaxies but introduce overly spherical halo cores. Early galactic winds could eject low angular momentum baryons from halos (Binney et al. 2001), removing inner cusps and increasing the size of galactic disks. It remains to be seen how successful these ideas will be in solving the problems experienced by models of dark matter on galactic scales.

These problems have been heightened by recent calculations of the optical depth to microlensing which imply that all of the matter in the direction of the Galactic bulge must be capable of microlensing (Binney et al. 2000; Binney and Evans 2001), ruling out the existence of particle dark matter in the central Galaxy and the presence of a dark halo with a cuspy density profile.

Weiner and Sellwood (1999) performed hydrodynamic simulations of the barred galaxy NGC 4123 to determine the relative contributions of the disk and dark halo to the galaxy's rotation curve. Taking the input potential as a combination of the deprojected I band image of the disk and an analytic halo, they were only able to match the the observed gas flows and shocks if the disk mass to light ratio was 80 - 100% of its largest possible value: a maximum disk.

Further evidence for barred spiral galaxies having minimal dark halos in their central regions was provided by Debattista and Sellwood (2000) who showed that dynamical friction between a stellar bar and a dark halo implies that galaxies with fast bars must have low density halos. Since almost every bar pattern speed which has been calculated has yielded a fast bar (see Section 1.2), we must conclude that barred galaxies have minimal halos.

The situation is not so clear cut for unbarred galaxies. It might seem counter intuitive for barred galaxies to be disk dominated and unbarred galaxies to be halo dominated in their inner parts. However, the simulations of Kranz et al. (2001) find that a significant halo ($\frac{1}{3}$ of the total

mass inside one K' band exponential scale length) is needed to reproduce the gas velocity field of NGC 4254. Although, they may not be justified in assuming a fixed pattern speed for the unbarred galaxy, their result indicates that the dark matter content of unbarred galaxies warrants further investigation.

1.4 The Andromeda galaxy

1.4.1 Observations

The Andromeda galaxy, or M31, has been studied for over a thousand years. It is one of the few galaxies which can be seen with the naked eye and was mentioned in the *Book of the Fixed Stars* (Abd al-Rahman b. Umar al-Sufi 964). As the closest large spiral to our own Galaxy, Andromeda provides us with an ideal laboratory for probing the structure and dynamics of spiral galaxies. Its large inclination angle (77°) means that although it is difficult to determine accurate positional data on its apparent minor axis, more information is available on the kinematics of the galaxy.

Due to its close proximity, high resolution photometric observations of both stars and gas have long been possible. Maps in U, B and V bands were obtained by de Vaucouleurs (1958) and updated versions in U, B, V and R bands by Walterbos and Kennicutt (1988).

There is both photometric and kinematic evidence for triaxiality in M31. Lindblad (1956) noted that the apparent major axes of bulge and disk are misaligned and that the central isophotes are twisted with respect to the isophotes of the outer disk. He attributed this to the presence of a triaxial bulge rotating in the plane of the disk.

In addition, anomalously high velocities in the central 10′ of M31 have been observed in Hα (Pellet 1976), atomic hydrogen (Whitehurst and Roberts 1972; Brinks and Burton 1984) and CO molecular gas (Stark 1985; Loinard et al. 1995).

1.4.2 Mass models of the bulge of M31

In order to account for the observations described above, a number of models have been proposed for the central region of M31 and its bulge in particular. Ruiz (1976) constructed distribution functions with two integrals for a two component model in which the nucleus and inner bulge

(out to 6′) are represented by concentric oblate spheroids, tilted with respect to the disk. The models were able to match the surface brightness profile and produce an inner peak in the rotation curve but were significantly at odds with kinematic observations, in particular the bulge velocity dispersions and the magnitude of the rotation curve.

Stark (1977) assumed that the bulge rotates in the plane of the disk. He showed that the combination of this assumption, the observed surface brightness profile and the twisted isophotes that were noted by Lindblad, implies that the bulge is one of a one parameter family of triaxial models. The parameter, ϕ, is the angle between the minor axis of the bulge and the line of nodes of the disk.

Following this, Simien et al. (1979) built a dynamical model based on an oblate spheroidal bulge, using the surface brightness profile, the stellar rotation curve and the stellar velocity dispersion as constraints. Their model predicts a reasonable B band mass-to-light ratio of 6 and a nearly isotropic velocity dispersion. By their own admission, however, Stark's model provides a better fit to the photometry, implying that a triaxial model may be necessary.

Kent (1989) improved on the above dynamical models by using newer data sets, adding a nucleus, disk and constant density dark halo and modelling the bulge as an oblate spheroid with variable ellipticity. Kent believed that observations were not good enough to detect the consequences of the almost certainly triaxial nature of the bulge. The best fitting model predicts a similar bulge mass-to-light ratio to that of Simien et al. (1979), but was unable to account for the non-circular gas streaming velocities.

By calculating closed, non-intersecting, periodic orbits in the plane of the disk and arguing that interstellar gas largely follows these orbits, Stark and Binney (1994) attempted to account for the non-circular gas motions and showed that the anomalously high central gas velocities mentioned above would occur in a simple, analytic, rotating barred potential.

These models are improved upon further through the investigation detailed in Chapter 5.

1.5 Dual morphology spiral galaxies

1.5.1 Observations

Spiral galaxies look different when observed in different wavebands. The differences in the morphology of the galaxy M51 when viewed in yellow and blue light were mentioned by Zwicky (1955) and Zwicky (1957):

> *While the two main spiral arms are smooth in terms of the yellow-green population they appear interrupted and they show some weird lateral deflections and outcroppings within the blue populations of the galaxy.*
> *... a normal spiral when seen in the light of its blue stars and a barred spiral in the light of its yellow-green stars.*

Hence, a single optical image was unable to describe the true morphology of a spiral galaxy.

Further evidence for the danger of equating optical images with the underlying mass distribution of spiral galaxies came from Strom et al. (1976) and Burstein and Rubin (1985). As a method of observing density waves and the underlying stellar disk, Strom et. al. suggested that galaxies should be imaged at near infrared wavelengths. Burstein and Rubin then showed that the overall mass distribution, as determined from the rotation curve, is uncorrelated with optical morphology or Hubble type.

With the availability of K' band imaging, it became possible to observe many galaxies in the optical and infrared to determine what, if any, morphological differences exist between the two wavebands. Block and Wainscoat (1991) showed that NGC 309 (Figure 5.2) exhibits dramatic dual morphologies, appearing as a multi armed unbarred spiral in the optical and a two armed barred galaxy in the infrared. Block et al. (1994) illustrates the lack of correlation between B band and K' band galaxy morphology with images of seven galaxies in both wavebands: NGC 309, 521, 1637, 2997, 4622, 4736 and 5195.

Rix and Rieke (1993) observed M51 in a multitude of wavelengths. Their K band image clearly shows the lack of flocculence and presence of a central bar pointed out by Zwicky. A number of optically flocculent galaxies were observed in the infrared by Thornley (1996): NGC 2403, 3521, 4414 and 5055, and by Block et al. (1996): NGC 2841. They found that the infrared images exhibited very weak, but smooth, stellar spiral arms that had not been seen in the optical.

Further galaxies were observed by Grosbol and Patsis (1998): NGC 3223, 5085, 5247, 5861 and 7083, and Seigar and James (1998): NGC 5478, UGC 6958 and UGC 8939. Grosbol and Patsis find K' band grand design spirals in all five galaxies, despite only two of them having optically grand design structure. The spiral arm pitch angles in their sample are shown to be tighter in bluer colours. Seigar and James, however, find the reverse: the B band spiral arms are looser. Both find it difficult to determine a galaxy's Hubble type from its K band morphology.

1.5.2 Dust penetrated classification

Given the radical differences in morphology between optical and infrared images of spiral galaxies, Block and Puerari (1999) define a new classification system based on a physical framework derived from infrared images of the underlying stellar distribution.

Their dust penetrated classification system categorizes galaxies on the basis of the pitch angle of their spiral arms and on their dominant Fourier component (the number of arms). Pitch angles are divided into three classes: α, β and γ corresponding to $4° < i < 14°$, $18° < i < 30°$ and $35° < i < 49°$, respectively, where i is the pitch angle of the spiral arms in the K' band. Buta and Block (2001) extend the system to include bar strength, measured as the ratio between the maximum tangential force and the mean axisymmetric radial force.

Block and Puerari (1999) demonstrate that galaxies of diverse Hubble classification can be represented by the same dust penetrated (DP) class. For instance, NGC 718 (SABa) and NGC 309 (SABc) exhibit very different optical morphology and yet are remarkably similar in the infrared and both placed in the β class. And although NGC 2841 and NGC 7083 are both Sb galaxies, they have α and γ classes, respectively.

1.5.3 Methods of modelling stars and gas in spiral galaxies

In Chapter 3, I develop a code to simultaneously model the dynamics of stars and gas in a spiral galaxy in order to address the morphological differences mentioned above. Hence, I present here a selection of previous methods of modelling stars and gas and highlight their conclusions regarding galaxy morphology. Of the many previous simulations, no refereed results have used both finite-difference hydrodynamic and N-body techniques. The theory behind, and examples of,

hydrodynamic and N-body simulations are described in Sections 2.1 and 2.3, respectively.

Miller et al. (1970) start with 115659 N-body gas particles and convert a fraction of them to stellar particles at every time step. 'Gas' particles remain dynamically cool by redistributing their velocities to reduce their velocity dispersion but conserve momentum. The gas component forms spiral-like structure whereas the stellar component heats up until it becomes pressure supported and loses almost all of its sub-structure.

A number of later simulations make use of the sticky particle technique of Schwarz (1981) in which gas is represented by particles which have no self gravity but collide inelastically if they get too close to each other. In particular, Carlberg and Freedman (1985) use 30000 stellar particles and 10000 sticky gaseous ones. Large scale spiral structure forms in both components. The narrower gaseous arms overlay the stellar spirals and the number of arms is roughly equal to the inverse of the active fraction of mass in the disk.

A combination of 4141 stellar particles and finite-difference hydrodynamics on a 50×50 grid is used by Schroeder and Comins (1986) and Schroeder (1989). The potential is calculated from the total stellar and gaseous density. Perhaps because of the few number of particles used, these simulations were unable to reproduce reasonable galaxy morphologies.

Thomasson et al. (1990) use 30000 stellar and 30000 sticky gaseous particles. A stellar cooling routine reduces high peculiar velocities to the average local velocity. Previous simulations were only able to generate long lived multiple armed structures. Here, grand design spiral structures persist for 3 rotation periods in both components due to the inclusion of an ILR inside a Q-barrier. The stellar arms are smoother and wider than those in the gas. These results are improved upon by Elmegreen and Thomasson (1993) using 100000 N-body particles and the same stellar cooling routine. By adding star formation and gas heating algorithms, grand design structure persists for over 6 rotations, with a third arm occasionally being visible. Flocculent gas spirals are observed in models with $Q_{\mathrm{star}} > 2.5$, and smaller pitch angles arise if shear rates are high in the disk.

Junqueira and Combes (1996) combine an N-body code with the Beam Scheme (Sanders and Prendergast 1974) and include the self gravity of the gas. They find that the fraction by mass of gas plays a critical role in determining the stability, and hence propensity to forming and dissolving

bars and spirals, of their simulations. A low gas mass fraction regulates the dynamical heating of the stars and prolongs the time for which bars and spirals can survive. Too much gas produces additional lumpiness which can scatter the stars and destroy large scale stellar structures.

Gerritsen and Icke (1997) add star formation algorithms to TREESPH, a 3D N-body tree code which has been combined with smooth particle hydrodynamics (Hernquist and Katz 1989). They do not assume that the gas is isothermal, but produce a two phase ISM. This results in a Schmidt power law dependence of the star formation rate on the gas density. Flocculent structures are found in the cold gas and young star components, with no structure apparent in the older stars. The cause of the flocculence is pinned on the effect of shear on gravitational clustering of the dissipative gas.

Lastly, the AAS bulletins of Zeltwanger (1999) and Comins and Zeltwanger (2000) describe a self-consistent code incorporating N-body techniques (100000 stellar plus 10000 sticky gas particles), self gravitating hydrodynamics, star formation, gas infall and supernovae. Persistent spiral structure is claimed for at least 44 rotation periods or 11 Gyr.

The above comparisons are made more difficult by the lack of a consistent definition of the onset and dissapearance of spiral structure in these simulations.

1.6 Layout of the thesis

This thesis aims to document the code which I have developed and the two topics which have been tackled with it. In Chapter 2, I describe some of the theory which is used later in the thesis, namely hydrodynamics, gas dynamics in barred galaxies, N-body theory and galactic stability. Chapter 3 provides documentation of the DUAL code and should be sufficient to enable use of the code by others for further research. The investigation into the shape, rotation speed and dark matter content of the central bulge region of M31 is described in Chapter 4. Chapter 5 provides details of my research into using simple simulations to reproduce observations of dual morphology of spiral galaxies in the optical and infrared regimes. Finally, Chapter 6 presents the conclusions of the above topics and suggestions for further research.

Chapter 2

Theory

2.1 Hydrodynamics

2.1.1 The ISM as a 2D isothermal fluid

Cowie (1980) argued that the cold dense ensemble of molecular clouds ($n \sim 50\ \mathrm{cm}^{-3}, T \sim 100\ \mathrm{K}$) which forms the bulk of the mass of the interstellar medium could be accurately represented by an isothermal fluid, if the time scale for cloud growth or cloud cooling is short compared to the dynamical time scale. In these cases, (generally found to be true), the system attains a state of quasi equilibrium in which the isothermal sound speed c is taken to be the mass averaged cloud velocity dispersion and the fluid surface density Σ is the average density of the surrounding clouds.

2.1.2 The Euler equations, Riemann problems and shocks

The Euler equations

By modelling the ISM as a 2D isothermal, continuous fluid, its dynamics can be followed using the isothermal version of the 2D, non-linear Euler equations of hydrodynamics.

These equations are derived from the conservation laws of mass and x and y momentum, plus the equation of state for a compressible frictionless isothermal fluid. For an ideal gas with constant temperature the equation of state defines the pressure as $P = \Sigma RT$, where T is the temperature and R is the gas constant. Since $c^2 = (\partial P/\partial \Sigma)_T$, this can be written as $P = \Sigma\, c^2$. The temperature is constant, and so no energy equation is needed.

The derivation of the Euler equations neglects viscous stresses and heat fluxes. However, body forces, for instance gravitational source terms, can be included on the right hand side. This gives

the 2D, conservative form of the Euler equations with source terms,

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{U})}{\partial x} + \frac{\partial \mathbf{G}(\mathbf{U})}{\partial y} = \mathbf{S}(\mathbf{U}),$$

$$\mathbf{U} = \begin{pmatrix} \Sigma \\ \Sigma u \\ \Sigma v \end{pmatrix}, \ \mathbf{F} = \begin{pmatrix} \Sigma u \\ \Sigma(u^2 + c^2) \\ \Sigma uv \end{pmatrix},$$

$$\mathbf{G} = \begin{pmatrix} \Sigma v \\ \Sigma uv \\ \Sigma(v^2 + c^2) \end{pmatrix}, \ \mathbf{S} = \begin{pmatrix} 0 \\ -\Sigma\phi_x \\ -\Sigma\phi_y \end{pmatrix}, \tag{2.1}$$

where u and v are the gas velocities, and ϕ_x and ϕ_y are the gravitational potential gradients in the x and y directions respectively. $\mathbf{U}$ is the vector of conserved quantities (so called because these quantities - mass and momentum - are associated with the conservation laws from which the Euler equations are derived), $\mathbf{F}(\mathbf{U})$ and $\mathbf{G}(\mathbf{U})$ are flux vectors in the x and y directions and the $\mathbf{S}(\mathbf{U})$ is a vector of source terms.

For later reference, it is also useful to define an amended source term for the Euler equations in a rotating frame,

$$\mathbf{S} = \begin{pmatrix} 0 \\ \Sigma(\Omega_p^2 x + 2\Omega_p v - \phi_x) \\ \Sigma(\Omega_p^2 y - 2\Omega_p u - \phi_y) \end{pmatrix}, \tag{2.2}$$

where Ω_p is the rotation speed of the frame.

Discontinuities and shocks

The differential form of the Euler equations assumes that the density and velocity are smooth and continuous. This is often not the case, either in the initial conditions (e.g. the Riemann problem below) or during the evolution of the solution (as shown in Figure 2.1). Hence, they must be

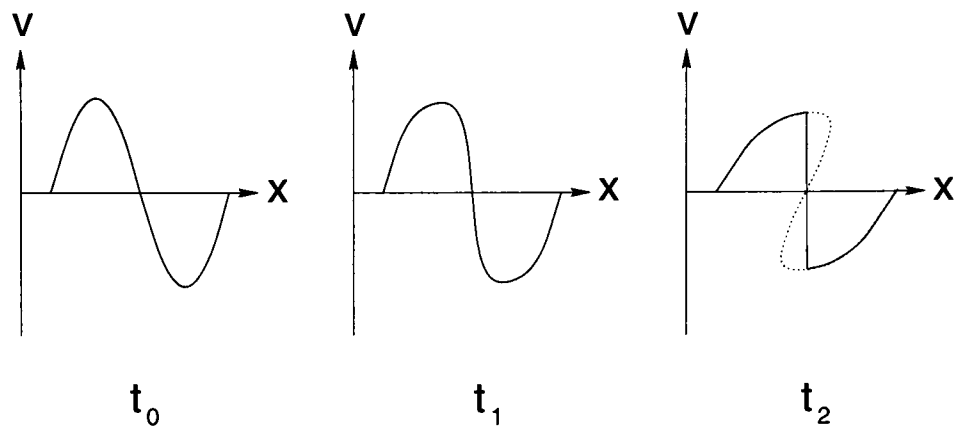


Figure 2.1: The three subpanels show subsequent times in the frame of the wave. If the rear of a wave propagates at a faster speed than its middle, the wave will steepen and, eventually, break. At this point the density and velocity try to become multi-valued, forming a shock wave or discontinuity.

supplemented by the Rankine-Hugoniot jump condition,

$$\mathbf{F}_1 - S\,\mathbf{U}_1 = \mathbf{F}_2 - S\,\mathbf{U}_2, \tag{2.3}$$

which is derived from the conservation of mass and momentum applied normal to any discontinuity. The subscripts 1 and 2 refer to variables before and after the shock respectively and S is the speed of the shock wave. In real gases, shock fronts develop that are not strict discontinuities and have a width due to their viscosity. This width is proportional to their mean free path, which for the interstellar medium is $\sim 6 \times 10^{-6}$ pc (Bowers and Deeming 1984), far smaller than the resolution of any galactic gasdynamical simulation.

The Riemann or shocktube problem

The Riemann problem is an initial value problem for a set of partial differential equations in which the dynamical variables are held constant in two regions separated by a discontinuity. For the Euler equations, the simplified 1D Riemann problem in which the initial velocity is zero everywhere is known as the shocktube problem and is physically equivalent to removing a diaphragm separating

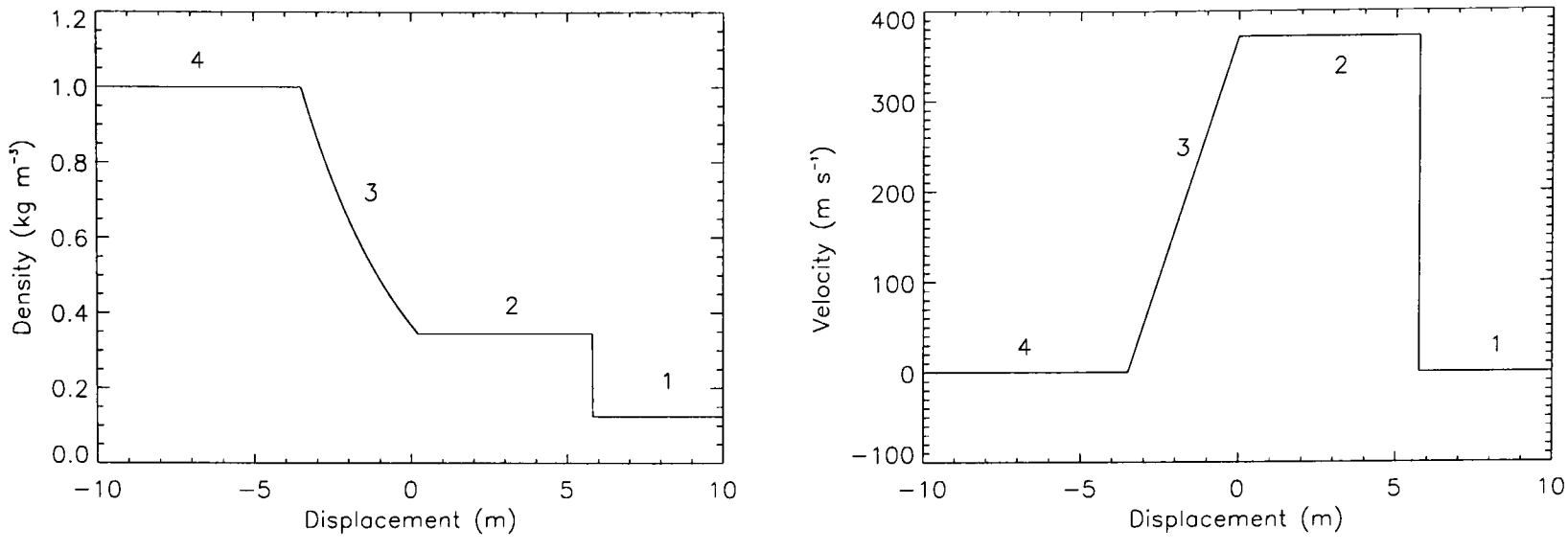


Figure 2.2: Analytic result for the density and velocity of a 1D isothermal shocktube after 0.01 s, with sound speed, $c = 350\ \mathrm{m\,s^{-1}}$. Regions 1 and 4 maintain the initial conditions, a shock wave exists between regions 1 and 2 and region 3 contains a rarefaction wave.

two stationary gases. For 1D isothermal gasdynamics the initial conditions are

$$\Sigma(x,0) = \begin{cases} \Sigma_1, & \text{if } x > 0 \\ \Sigma_4, & \text{if } x < 0 \end{cases}$$

$$u(x,0) = 0. \tag{2.4}$$

The Riemann problem is one of the few problems in gasdynamics with an analytic solution and so provides a very useful test of hydrodynamic algorithms and codes. Finite difference schemes usually approximate the true gas density and velocity fields within cells by a constant, linear or parabolic function whereas the cell boundary values are separated by discontinuities. Since the Riemann problem provides the solution at these discontinuities, it becomes integral to many numerical hydrodynamic schemes.

The particular shocktube problem used in Chapter 3, where $\Sigma_4 = 1\ \mathrm{kg\,m^{-3}}$, $\Sigma_1 = 0.125\ \mathrm{kg\,m^{-3}}$ and $c = 350\ \mathrm{m\,s^{-1}}$, is illustrated in Figure 2.2 at a time of $t = 0.01$ s. It shows the salient features of shocktube solutions for the 1D isothermal Euler equations. Regions 1 and 4 remain in the same state as they were at $t = 0$. Between regions 1 and 2 is a shock wave, creating a discontinuity in both density and velocity whereas region 3 contains a rarefaction wave. Riemann or shocktube

problems are best solved by considering first the shock wave, then the rarefaction wave and then the system as a whole.

The shock wave In the frame of the shock, the shock speed is $S = 0$ and velocities are $\hat{u} = u - S$. The jump conditions of Equation 2.3 then give

$$\begin{aligned} \Sigma_2 \hat{u}_2 &= \Sigma_1 \hat{u}_1, \\ \Sigma_2(\hat{u}_2^2 + c^2) &= \Sigma_1(\hat{u}_1^2 + c^2). \end{aligned} \tag{2.5}$$

Assuming that the shock is moving to the right and hence $S > u_2 > u_1$, combining the above two equations leads to,

$$\begin{aligned} S &= u_1 + c\sqrt{\frac{\Sigma_2}{\Sigma_1}}, \\ u_2 &= u_1 - c\sqrt{\frac{\Sigma_2}{\Sigma_1}}\left(\frac{\Sigma_1}{\Sigma_2} - 1\right). \end{aligned} \tag{2.6}$$

The density Σ_2 remains to be found later.

The rarefaction wave To solve the shocktube problem in smooth regions, we need to know the characteristic variables (quantities which remain constant along characteristics) of the 1D isothermal Euler equations. Rewriting the equations as

$$\begin{aligned} \mathbf{U}_t + \mathbf{A}(\mathbf{U})\,\mathbf{U}_x &= \mathbf{0}, \\ \mathbf{U} = \begin{pmatrix} \Sigma \\ \Sigma u \end{pmatrix}, \qquad \mathbf{A} &= \frac{\partial \mathbf{F}}{\partial \mathbf{U}} = \begin{pmatrix} 0 & 1 \\ c^2 - u^2 & 2u \end{pmatrix}, \end{aligned} \tag{2.7}$$

we determine the eigenvalues of $\mathbf{A}$ to be $\lambda^{\pm} = u \pm c$ by setting the determinant $|\lambda\mathbf{I} - \mathbf{A}| = \mathbf{0}$. Physically, eigenvalues are equal to speeds at which information can propagate along a wave. The

right eigenvectors of $\mathbf{A}$, $\mathbf{K}$, are such that $\mathbf{AK} = \lambda\,\mathbf{K}$, which gives

$$\mathbf{K}^{\pm} = \frac{1}{2c}\begin{pmatrix} 1 \\ u \pm c \end{pmatrix}. \tag{2.8}$$

The characteristic form of the Euler equations is obtained by diagonalizing $\mathbf{A}$ into the matrix $\mathbf{\Lambda} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$, where $\mathbf{Q} = (\mathbf{K}^{+}, \mathbf{K}^{-})$. Thus Equation 2.7 is transformed into

$$\mathbf{W}_t + \mathbf{\Lambda}\mathbf{W}_x = \mathbf{0},$$
$$\mathbf{\Lambda} = \begin{pmatrix} \lambda^{-} & 0 \\ 0 & \lambda^{+} \end{pmatrix} = \begin{pmatrix} u-c & 0 \\ 0 & u+c \end{pmatrix}, \tag{2.9}$$

where the characteristic variables $\mathbf{W} = (W^{+}, W^{-})^{T}$ are related to the conserved quantities $\mathbf{U}$ by

$$d\mathbf{W} = \mathbf{Q}^{-1}\,d\mathbf{U} = \begin{pmatrix} c-u & 1 \\ c+u & -1 \end{pmatrix}\begin{pmatrix} d\Sigma \\ d(\Sigma u) \end{pmatrix} = \begin{pmatrix} c\,d\Sigma + \Sigma\,du \\ c\,d\Sigma - \Sigma\,du \end{pmatrix}. \tag{2.10}$$

Using the definition of the total differential, Equation 2.9 can be solved to show that

$$dW^{\pm} = 0 \quad \text{if} \quad \frac{dx}{dt} = \lambda^{\pm}. \tag{2.11}$$

Hence, Equation 2.10 can be integrated to give the characteristic variables: quantities that remain constant within a simple wave along the characteristics $dx/dt = \lambda^{\pm} = u \pm c$ (Landau and Lifshitz 1959). For the 1D isothermal Euler equations, $W^{\pm} = u \pm c\ln\Sigma$ (Laney 1998).

Along the left moving rarefaction wave $\lambda = \lambda_{-} = u - c$. Hence $\mathrm{d}x/\mathrm{d}t = u - c$ and

$$u_3(x,t) = x/t + c. \tag{2.12}$$

In both regions 3 and 4, rightward moving information has propagated along the same character-

istics. Hence, $u_3(x,t) + c\ln\Sigma_3(x,t) = u_4 + c\ln\Sigma_4$ or

$$\Sigma_3(x,t) = \Sigma_4 \exp\left(\frac{u_4 - u_3(x,t)}{c}\right). \tag{2.13}$$

The whole system The same rightward moving characteristics can be applied across regions 2 and 4 to give $u_2 + c\ln\Sigma_2 = u_4 + c\ln\Sigma_4$. Combining this with Equation 2.6 yields

$$\Sigma_2 = \Sigma_4 \exp\left[\frac{u_4 - u_1}{c} + \sqrt{\frac{\Sigma_2}{\Sigma_1}}\left(\frac{\Sigma_1}{\Sigma_2} - 1\right)\right], \tag{2.14}$$

which must be solved iteratively to complete the solution of the Riemann problem. The shocktube problem solution is found by substituting $u_1 = u_4 = 0$ in the above formulae. A more complete analysis of the Riemann problem can be found in Toro (1999).

2.1.3 Computational fluid dynamics

Discretization of partial differential equations

To solve the Euler equations numerically, as in a grid based hydrodynamics code, they must first be approximated and then discretized.

Figure 2.3 illustrates how a derivative of a function $f(x)$ at x_0 may be approximated using data at discrete points $x_0 - \Delta x, x_0$ and $x_0 + \Delta x$. This is equivalent to considering a Taylor expansion of $f(x)$ at x_0 and neglecting terms of second order or higher,

$$\begin{aligned} f(x_0 + \Delta x) &= f(x_0) + \Delta x f_x(x_0) + O(\Delta x^2), \\ f(x_0 - \Delta x) &= f(x_0) - \Delta x f_x(x_0) + O(\Delta x^2). \end{aligned} \tag{2.15}$$

Rearranging these expansions gives the following first order finite difference approximations

$$\begin{aligned} f_x(x_0) &= \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} + O(\Delta x), && \text{a forward difference,} \\ f_x(x_0) &= \frac{f(x_0) - f(x_0 - \Delta x)}{\Delta x} + O(\Delta x), && \text{a backward difference,} \end{aligned} \tag{2.16}$$

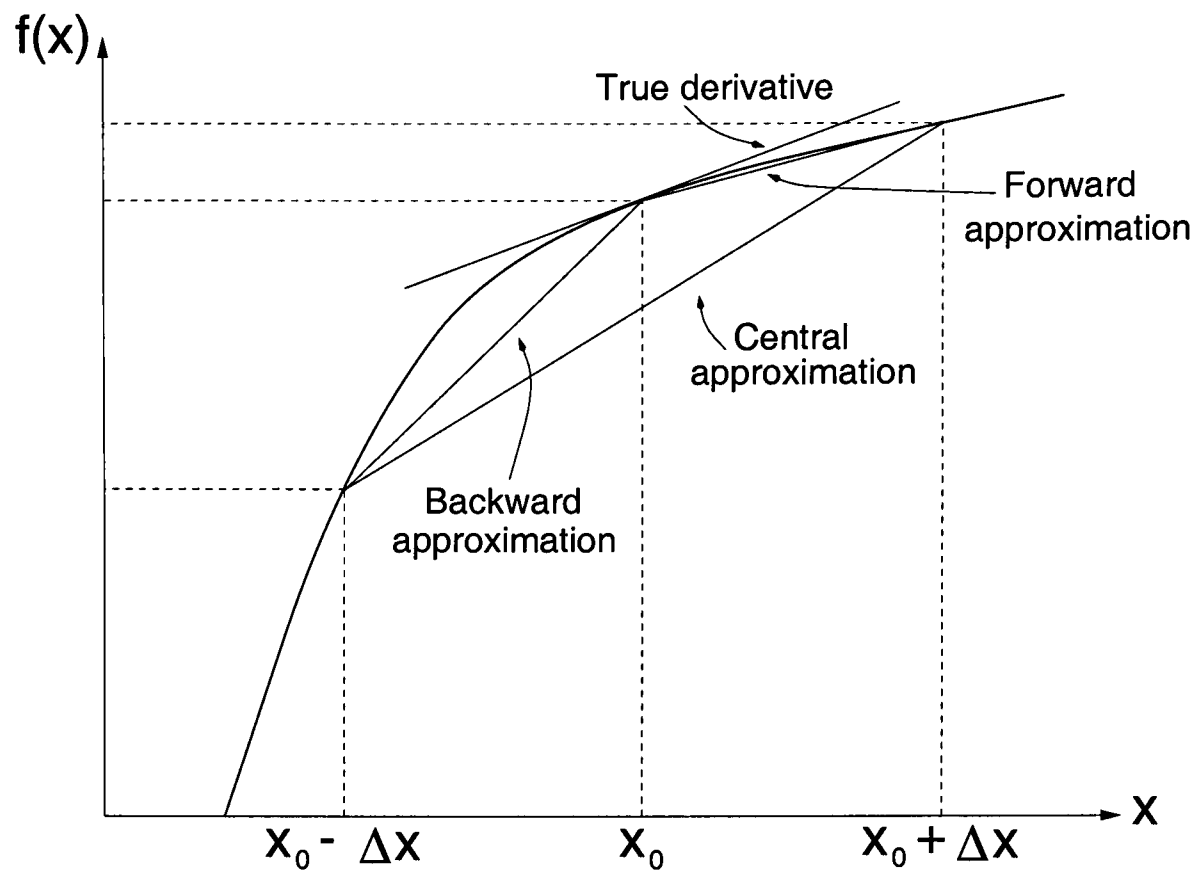


Figure 2.3: Forward, backward and central finite difference approximations to first derivative of $f(x)$ at point x_0.

whereas subtracting one from the other gives the second order approximation

$$f_x(x_0) = \frac{f(x_0 + \Delta x) - f(x_0 - \Delta x)}{2\Delta x} + O(\Delta x^2), \qquad \text{a central difference.} \tag{2.17}$$

The computational domain must also be discretized in both space and time, as shown in Figure 2.4 for the one dimensional case. Dotted lines mark the centres of cells whereas dashed lines mark their edges. The cells are positioned at $(x_i, t^n) = (i\Delta x, n\Delta t)$, where $i = 0, 1, \ldots, I-1$, $t = 0, 1, \ldots, N-1$, I is the number of cells and N is the number of time slices. When functions, for instance $u(x,t)$, are discretized, their values on the grid are referred to as $u_i^n = u(x_i, t^n)$. Solving a partial differential equation, e.g.

$$u_t + a u_x = 0, \tag{2.18}$$

on the grid entails calculating the quantities u_i^{n+1} from the values of u_i^n which would be given as initial conditions to the problem. The boundaries to the computational domain exist at $x = 0$ and $x = I\Delta x$, and the values of u_0^n and u_{I-1}^n on the boundaries are determined by the boundary conditions that would go with Equation 2.18. Equation 2.18 is discretized by applying the ap-

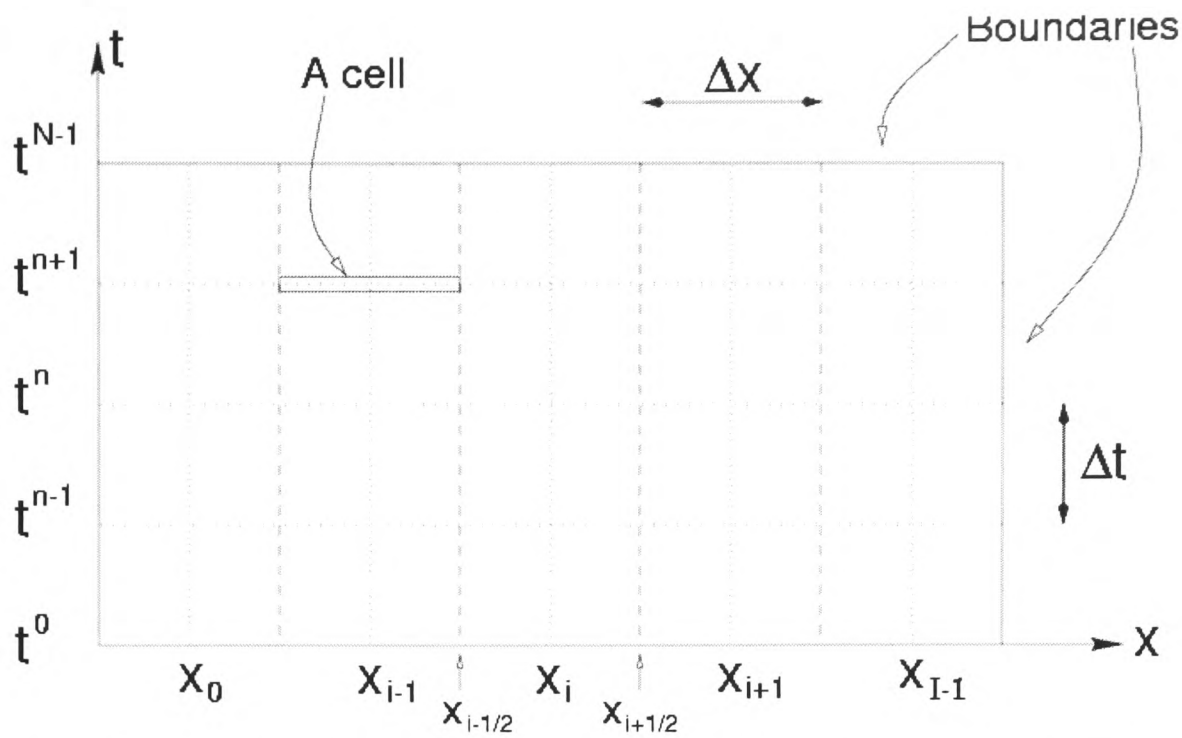


Figure 2.4: Discretization of the computational domain onto a grid. The dotted lines at $x_0, x_{i-1}, x_i, x_{i+1}$ and x_{I-1} mark the centres of their respective cells, whereas dashed lines at $x_{i+1/2}$ and $x_{i-1/2}$ are on cell boundaries. The highlighted cell is positioned at (x_{i-1}, t^{n+1}).

proximations of Equation 2.17 to the grid. If u_t is approximated by a forward difference and u_x by a central difference, the PDE becomes the finite-difference equation

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a\left(\frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x}\right) = 0, \tag{2.19}$$

which can be solved for the one unknown u_i^{n+1} to give the explicit Forward Time Central Space (FTCS) scheme

$$u_i^{n+1} = u_i^n - \frac{C}{2}(u_{i+1}^n - u_{i-1}^n), \tag{2.20}$$

where

$$C = \frac{a}{\Delta x/\Delta t} \tag{2.21}$$

is the dimensionless Courant number. This is the ratio of the maximum wave speed a (which will vary with the particular PDE and numerical scheme used) to the grid speed $\Delta x/\Delta t$.

Stability Unfortunately, the scheme of Equation 2.20 is useless since it is completely unstable. This can be shown by assuming a solution $u_i^n = A^n e^{i\theta \mathrm{j}}$, where $\mathrm{j} = \sqrt{-1}$. For stability, one requires that the amplitude $||A|| \leq 1$ but for Equation 2.20, $||A||^2 = 1 + C^2 \sin^2\theta$ and is therefore unstable for all values of C except 0.

However, using the same solution and assuming a positive sound speed a, a similar scheme based on the forward instead of central difference,

$$u_i^{n+1} = u_i^n - C(u_i^n - u_{i-1}^n), \tag{2.22}$$

has an amplitude $||A||^2 = (1-C)^2 + C^2 + 2C(1-C)\cos\theta$ and is therefore stable if $0 \leq C \leq 1$.

These allowed values of C imply restrictions on the time step, since Equation 2.21 implies that for a scheme to be stable, the numerical domain of dependence must be larger than the physical domain of dependence or, equivalently, the maximum wave speed must be less than the grid speed.

Equation 2.22 is known as the first order upwind method (Courant et al. 1952), and for negative a can be reformulated using the backward spatial difference. Formulating the scheme according to the sign of the sound speed or the direction of information flow ensures that the scheme is upwind. In other words, the physics of the system being modelled is embodied in the discretization scheme so that calculations use mesh points in the direction from which information flows.

Flux vector splitting To use upwind methods, one has to determine which direction the signals are flowing in the grid. This is a more complicated task when solving the isothermal Euler equations than for the simple PDE of Equation 2.18, since instead of a scalar u and a single wave flowing with speed c, we have a vector $\mathbf{U}$ and a number of waves flowing at speeds determined by the eigenvalues of the matrix $\mathbf{A}(\mathbf{U})$ ($u \pm c$ in 1D and $u \pm c$, u in 2D).

There are two main classes of methods presently employed to determine which direction is upwind. One is the Godunov methods, originating with Godunov (1959). Here, a Riemann problem is solved at the interface between every cell at every time step, which automatically determines the direction of information flow. However, it is computationally very expensive.

The alternative (used in the DUAL code) is to use flux splitting techniques, where in every cell,

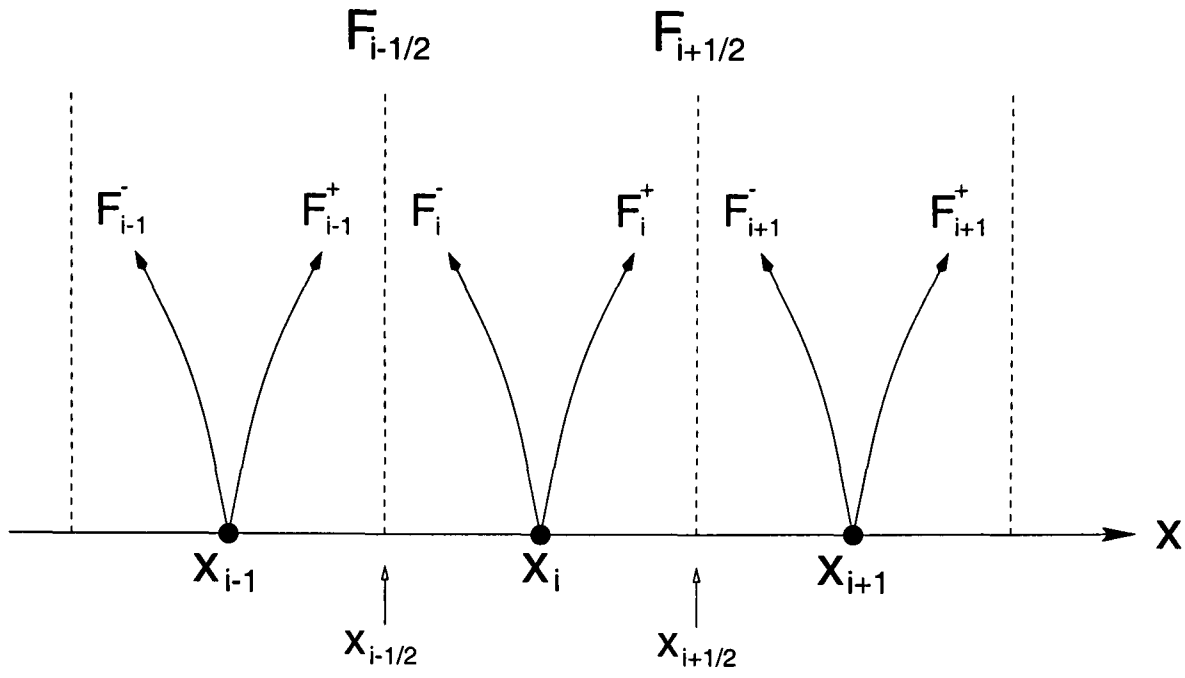


Figure 2.5: Splitting of the flux component within cells x_{i-1}, x_i and x_{i+1}. Dashed lines at $x_{i+1/2}$ and $x_{i-1/2}$ mark the cell boundaries. Fluxes on cell boundaries are composed of the flux travelling from the cell to the left of the boundary to the right and the flux travelling from the cell to the right of the boundary to the left.

the flux $\mathbf{F}(\mathbf{U})$ is split into $\mathbf{F} = \mathbf{F}^+ + \mathbf{F}^-$ according to a predetermined formula. The splitting of van Leer (1982) (see Section A.5) ensures that the Jacobian matrices of the positive and negative fluxes $\partial\mathbf{F}^\pm/\partial\mathbf{U}$ contain only positive or negative eigenvalues respectively and that the fluxes and their Jacobians are continuous, reducing the chance of introducing unphysical discontinuities and leading to smoother numerical results. In supersonic regions, where characteristics all flow in the same direction, $\mathbf{F}^\pm$ is defined to be either $\mathbf{F}$ or $\mathbf{0}$ leading to standard upwind differencing such as Equation 2.22 applied to the Euler equations. And to improve the capture of shocks, which are discontinuous on grid cell scales, van Leer's flux splitting causes one of the eigenvalues of each flux to vanish in subsonic flows. This can be seen from Equation A.13, where if $\mathbf{F}^\pm$ is written as $(F_1^\pm, F_2^\pm, F_3^\pm)^T$ then $F_1^\pm = F_2^\pm/2\,c = F_3^\pm/v$. Thus the flux vector is degenerate and there are two instead of three eigenvalues. The loss of an eigenvalue allows shocks to be captured with just two internal grid cells.

Figure 2.5 illustrates flux splitting. Fluxes at the cell boundary $x_{i+1/2}$ have been split into two components: $\mathbf{F}_i^+$ which is the value of the flux travelling to the right from the cell x_i to the left of the boundary and $\mathbf{F}_{i+1}^-$, the value of the flux travelling to the left from the cell x_{i+1} to the right of the boundary.

Dealing with discontinuities

As mentioned in Section 2.1.2, the solutions to the isothermal Euler equations often become discontinuous and need to be supplemented by the Rankine-Hugoniot conditions. This presents a problem for numerical schemes since data is discretely sampled at grid points u_i^n and so every cell boundary could be considered to be a discontinuity. Good algorithms should recognize where the true discontinuities are and ensure that smooth regions are solved using the Euler equations but that the Rankine-Hugoniot relations are applied across discontinuities.

FS2, the algorithm used in the DUAL code, achieves second-order accuracy in space by approximating the gas flow variables within a cell by a linear rather than a constant function. To determine the gradient of this function, $(\delta u)_i^n$, we could simply take the average of the differences in u_i^n in neighbouring cells,

$$(\delta u)_i^n = \frac{1}{2}(\Delta_+ + \Delta_-),$$
$$\Delta_+ = \frac{u_{i+1}^n - u_i^n}{\Delta x}, \ \Delta_- = \frac{u_i^n - u_{i-1}^n}{\Delta x}. \tag{2.23}$$

This formulation, however, implicitly assumes that the gradient of the solution $u(x)$, varies smoothly between x_{i-1} and x_{i+1}. If there is a true discontinuity in this interval then the computed value of $(\delta u)_i^n$ is meaningless because we have used information from across the shock and should have used the Rankine-Hugoniot conditions. The result is unphysical oscillations in the solution (van Leer 1983).

By imposing the condition that u_i^n is monotonicity preserving i.e. no new extrema are introduced into the solution, $(\delta u)_i^n$ is limited to

$$(\delta u)_i^n \leq \min(2\Delta_+, 2\Delta_-). \tag{2.24}$$

In this case, information is not used from across a discontinuity in order to calculate the solution in a smooth region and so the Rankine-Hugoniot conditions are unnecessary and the spurious oscillations are avoided.

It is preferable to use a smoother limiter, namely

$$(\delta u)_i^n = \frac{(\Delta_+^2 + \epsilon^2)\Delta_- + (\Delta_-^2 + \epsilon^2)\Delta_+}{\Delta_+^2 + \Delta_-^2 + 2\epsilon^2}$$
$$= \left(1 - \frac{(\Delta_+ - \Delta_-)^2}{\Delta_+^2 + \Delta_-^2 + 2\epsilon^2}\right) \cdot \frac{\Delta_+ + \Delta_-}{2}, \tag{2.25}$$

as introduced by van Leer (1983) and used in the FS2 algorithm of Appendix A. The addition of the small bias term ϵ prevents undesirable clipping of smooth extrema and division by zero. The van Leer limiter is biased towards the smaller of Δ_+ and Δ_-, hence if Δ_+ and Δ_- are consecutive finite differences of a smooth solution then $(\delta u)_i^n = \frac{1}{2}(\Delta_+ + \Delta_-)$. However, if a discontinuity lurks within the cell x_i, then $\Delta_\pm \to \infty$ or 0 and $(\delta u)_i^n \to \min(\Delta_+, \Delta_-)$.

Using this limiter means that the Euler equations are applied throughout the smooth regions either side of a discontinuity. Since the equations conserve mass and momentum, the Rankine-Hugoniot conditions are automatically satisfied across the discontinuity.

2.1.4 The FS2 algorithm

FS2 is an upwind finite-difference scheme, used in the DUAL code to solve the discrete isothermal Euler equations with source terms on a 2D Cartesian grid (van Albada and Roberts 1981; van Albada et al. 1982).

The algorithm ensures upwindness through flux-splitting. It utilizes operator splitting to include source terms, dimensional splitting to perform calculations in 2D and is formally second-order in both space and time. Discontinuities in the flow are handled by a limiter rather than by jump conditions and the scheme has no added numerical viscosity.

The FS2 algorithm can be broken down into smaller procedures, each of which can be thought of as acting like an operator on the conserved quantities $\mathbf{U}$. In these terms, the algorithm is

$$\mathbf{U}^{n+2} = \mathcal{S}^{\Delta t/2}\mathcal{Y}^{\Delta t}\mathcal{X}^{\Delta t}\mathcal{S}^{\Delta t}\mathcal{X}^{\Delta t}\mathcal{Y}^{\Delta t}\mathcal{S}^{\Delta t/2}\mathcal{T}\,\mathbf{U}^n \tag{2.26}$$

where Δt is the length of one time step, $\mathbf{U}^n$ is the value of $\mathbf{U}$ at time $t = n\Delta t$ and $\mathbf{U}^{n+2}$ is the value of $\mathbf{U}$ two time steps later. The $\mathcal{T}$ operator calculates the time step, $\mathcal{S}$ deals with source

terms and the $\mathcal{X}$ and $\mathcal{Y}$ operators define hydrodynamic sweeps in the x and y directions. In each case, the output of one operator is used as the input for the next. This transforms the 2D, partial differential Euler equations with sources into two 1D problems without sources and an ordinary differential equation.

The $\mathcal{T}$ operator applies the Courant condition to calculate the time step Δt as

$$\begin{aligned} \Delta t &= \frac{C\Delta x}{\max(w)}, \\ w &= \sqrt{(|u|+c)^2 + (|v|+c)^2}, \end{aligned} \tag{2.27}$$

where C is the Courant coefficient. Hence, the time step varies during the calculation.

Solving for the source terms

The $\mathcal{S}$ operator computes the effect of source terms by ignoring the hydrodynamic fluxes and reducing Equation 2.1 to

$$\frac{d\mathbf{U}}{dt} = \mathbf{S}, \tag{2.28}$$

which is now independent of the Euler equations and equivalent to

$$\mathbf{U}^{n+1} = \mathbf{U}^n + \int_{t^n}^{t^{n+1}} \mathbf{S}(t, \mathbf{U}(t))\, dt. \tag{2.29}$$

During the period in which the potential is changing with time, this integral must be solved using an appropriate ordinary differential equation solver. Using a two stage explicit Runge Kutta method (also known as the improved Euler method),

$$\begin{aligned} k_1 &= \Delta t\, \mathbf{S}(t^n, \mathbf{U}^n), \\ k_2 &= \Delta t\, \mathbf{S}(t^n + \frac{\Delta t}{2}, \mathbf{U}^n + \frac{k_1}{2}), \\ \mathbf{U}^{n+1} &= \mathbf{U}^n + k_2. \end{aligned} \tag{2.30}$$

However, if we consider a static potential, then the solution is even simpler and the operator $\mathcal{S}^{\Delta t}$

is equivalent to

$$\mathbf{U}^{n+1} = \mathbf{U}^n + \Delta t\, \mathbf{S}. \tag{2.31}$$

Obviously, for $\mathcal{S}^{\Delta t/2}$, Δt is replaced by $\Delta t/2$ in the above equations.

Performing $\mathcal{X}$ and $\mathcal{Y}$ sweeps

After eliminating the sources, we are left with the 2D homogeneous version of Equation 2.1,

$$\mathbf{U}_t + \mathbf{F}(\mathbf{U})_x + \mathbf{G}(\mathbf{U})_y = \mathbf{0}, \tag{2.32}$$

with initial conditions $\mathbf{U}^n$. When integrated over one time step we get $\mathbf{U}^{n+1}$. Equation 2.32 can be dimensionally split into two 1D homogeneous equations to be solved sequentially,

$$\mathbf{U}_t + \mathbf{F}(\mathbf{U})_x = \mathbf{0}, \tag{2.33}$$

$$\mathbf{U}_t + \mathbf{G}(\mathbf{U})_y = \mathbf{0}. \tag{2.34}$$

Equation 2.33 has initial conditions $\mathbf{U}^n$ and gives the solution $\mathbf{U}^{n'}$ after sweeping through the grid in the x direction for a time step Δt. Equation 2.34 has initial conditions $\mathbf{U}^{n'}$ and gives the solution $\mathbf{U}^{n+1}$ after sweeping through the grid in the y direction for a time step Δt.

Solution operators $\mathcal{X}$ and $\mathcal{Y}$ can be defined for each of these split equations, where the $\mathcal{X}^{\Delta t}$ operator is written discretely in conservative form as

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \frac{\Delta t}{\Delta x}(\hat{\mathbf{F}}_{i+1/2}^{n+1/2} - \hat{\mathbf{F}}_{i-1/2}^{n+1/2}) \tag{2.35}$$

where $\mathbf{U}_i^n$ is the value of $\mathbf{U}$ at time $t = n\,\Delta t$ and position $x = i\,\Delta x$ and $\hat{\mathbf{F}}_{i+1/2}^{n+1/2}$ is a flux function to be determined. The flux here is defined on the cell boundaries $i+1/2$ and at an intermediate time $n+1/2$. According to van Albada et al. (1982), any first order upwind differencing method can achieve second order accuracy in time by advancing the cell boundary values to the intermediate time level $n+1/2$. In obtaining these values, the interaction between cells can be fully ignored.

The $\mathcal{Y}^{\Delta t}$ operator is obtained by replacing i by j, x by y and $\mathbf{F}$ by $\mathbf{G}$ in Equation 2.35.

The calculation of the flux function $\hat{\mathbf{F}}$, consists of four steps:

- evolve the cell centred values to the intermediate time step to ensure second order accuracy in time,
- reconstruct the intra-cell data by replacing the cell centred values by piecewise linear functions to ensure second order accuracy in space,
- calculate the cell boundary values by interpolation,
- use flux splitting to determine the flux function and ensure that information always flows upwind.

The details of this calculation are contained in Appendix A.

2.2 Gas dynamics in barred galaxies

The motion of gas in a galactic bar can have a significant effect on the host galaxy. Hence by examining the gas dynamics of galaxies with barred potentials, it is possible to gain insights into the central regions of these galaxies.

To calculate the motion of the gas it is necessary to determine the underlying potential in which it moves. In many situations, the fraction of gas by mass is low enough to neglect the contribution to the potential from the self gravity of the gas. Orbit calculations below use the standard toy barred potential of Athanassoula (1992a), consisting of a Kuzmin disk, Hubble bulge and Ferrers bar.

2.2.1 Regular orbits

The structure of a galaxy is primarily built up from stable, periodic orbits of stars and gas. Gas can follow periodic orbits so long as they do not intersect. Intersecting orbits result in the formation of shocks. The periodic orbits that exist in a particular potential can be found by sampling the phase space defined by an orbit's Jacobi integral $E_J = 0.5(u^2 + v^2) + \phi(x, y) - 0.5\Omega^2(x^2 + y^2)$, and the x co-ordinate at which the orbit cuts the bar minor axis, all defined in a frame rotating with the bar. Each orbit is then integrated over time to determine whether it is periodic. The following descriptions of orbits in rotating bars is based on Athanassoula (1992a).

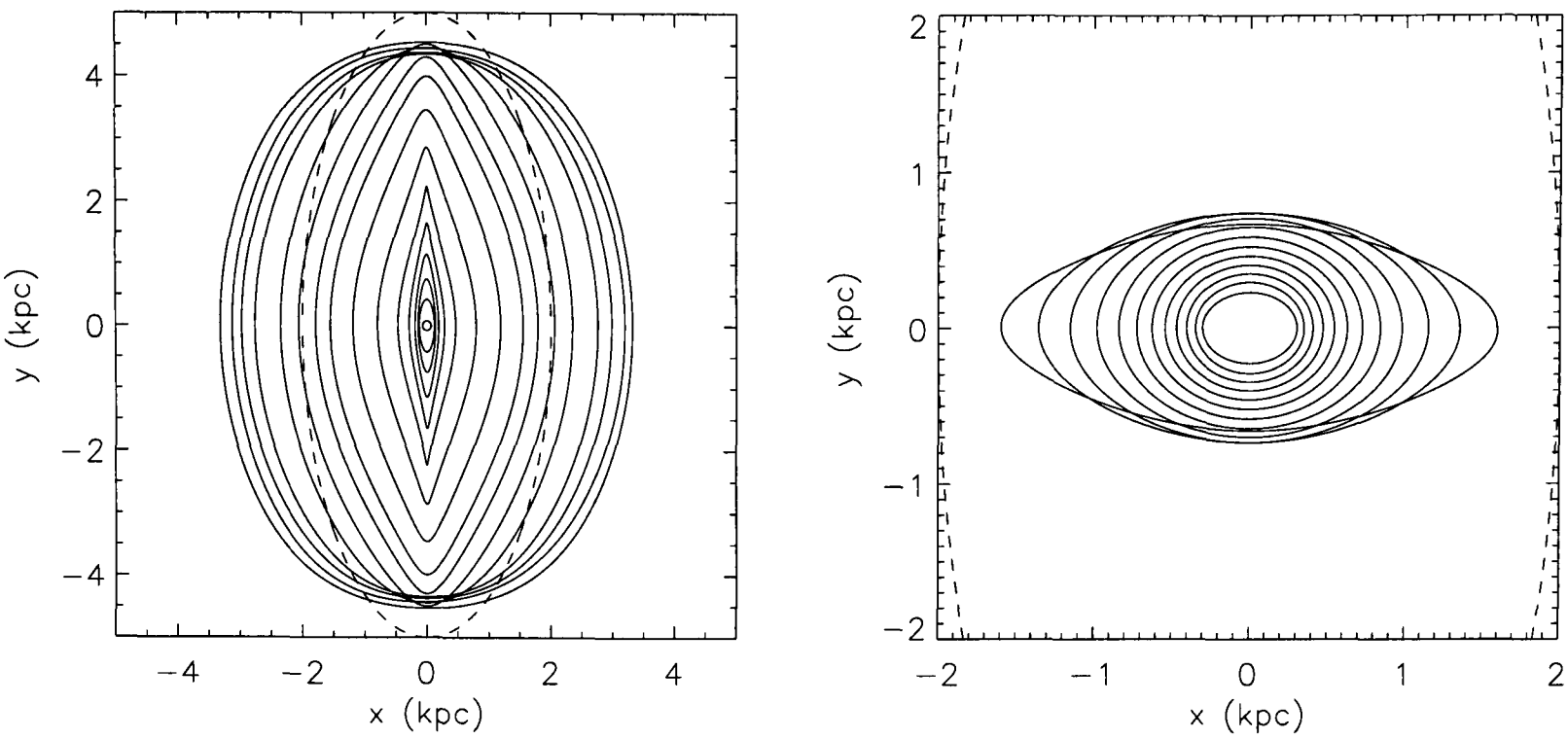


Figure 2.6: Periodic orbits in a barred potential. The left hand panel shows x_1 orbits whereas the right hand one shows x_2 orbits. In both cases, the dotted line highlights the bar.

In the plane $z = 0$ of an axisymmetric potential, gas flows on circular orbits. When a non-axisymmetric component in the form of a bar is introduced, the circular orbits bifurcate into a number of orbit families. In the notation of Contopoulos and Papayannopoulos (1980), the two main families are x_1 and x_2 (depicted in Figure 2.6), but many higher resonance orbits also exist.

The x_1 family consists of prograde orbits elongated along the major axis of the bar. Near the centre, the x_1 orbits are almost circular. As the size of the orbits and their Jacobi integral increases so do the orbits' axial ratio and they become lozenge shaped. This continues until the orbits become cusped. At larger radii, the axial ratio decreases until the orbits become nearly circular again.

The axial ratio and curvature of x_1 orbits increase with total central density and bar axial ratio and decrease with the pattern speed. For sufficiently large values of the bar axial ratio (about 1:5), instabilities begin to develop in x_1 orbits, resulting in loops at the orbits' apocentres which are linked to the lobes seen at the ends of some real bars. It is the high curvature of x_1 orbits at apocentre that leads to the formation of shock fronts in the gas.

The x_2 orbits are only present in certain barred galaxy models. They lie parallel to the bar minor axis and orbit in a prograde sense. They are smaller for low bar axial ratios and pattern

speeds and high central concentrations.

The existence and extent of x_2 orbits is important for the shape of the shock fronts that probably cause the dust lanes seen in many barred galaxies. Models without x_2 orbits exhibit shock fronts close to the bar major axis whereas the shock fronts in models with x_2 orbits are significantly offset from the major axis.

The x_2 orbits provide a useful measure of the existence and extent of the Inner Lindblad Resonance (ILR). The ILR occurs where the star encounters successive crests of the potential at a frequency equal to that of its natural radial oscillations. For axisymmetric models, this occurs when $2\,(\Omega - \Omega_p) = \kappa(R)$, where Ω_p is the bar pattern speed, $\Omega(R)$ is the circular frequency in the unbarred potential at that radius and $\kappa(R)$, the frequency of radial oscillations about an orbit, is

$$\kappa^2(R) = R\frac{d\Omega^2}{dR} + 4\Omega^2 = 2\Omega^2 + \Omega\frac{dv_c}{dR} = \frac{\partial^2\phi}{\partial R^2} + \frac{3}{R}\frac{\partial\phi}{\partial R}. \tag{2.36}$$

There is similarly an Outer Lindblad Resonance (OLR) located where $2\,(\Omega_p - \Omega) = \kappa(R)$.

This definition could be extended to non-axisymmetric models by averaging all quantities over azimuth. However, this leads to a family of non-axisymmetric and axisymmetric models having the same ILR. Athanassoula (1992b) suggests that a better approach is to utilize the x_2 families and say that an ILR exists at the maximum extent of the x_2 family of orbits. Further, if the x_2 family do not persist to the centre of the galaxy then an Inner Inner Lindblad Radius (IILR) exists where they terminate.

2.2.2 Previous simulations of gas dynamics in barred galaxies

By performing gas dynamical simulations using a 1D approximation, Roberts et al. (1979) achieved high enough resolution to discover shocks within the bars that they were modelling. Their shocks are offset forward of the minimum of the potential as seen in real galaxies, and lead to a focusing of the gas at the ends of the bars.

Sanders and Tubbs (1980) carried out two dimensional calculations using the beam scheme over a range of model parameters for a low resolution 40×80 grid. They also obtained offset shocks and deduced that such shocks imply that an Inner Lindblad Radius must be present near

to the bar centre. They further found that to achieve gas morphologies similar to observations, the bar mass distribution must be centrally concentrated and the corotation radius must not be much greater than the bar semi-major axis.

The parameter space for bar models was investigated extensively by Athanassoula (1992b) using an 80 × 160 grid and the FS2 algorithm. She found that the shape and extent of the shock loci are intimately linked to the transition between x_1 and x_2 orbits. To obtain offset shocks the ratio of the corotation radius to the bar major axis, $\mathcal{R} = R_{\rm co}/a$, had to be 1.2 ± 0.2. Away from the shocks, the gas density in the bar region is low. She also found substantial gas inflow of a few km s^{-1} for models with strong shocks.

Piner et al. (1995) performed calculations using the PPM algorithm on a 67 × 40 polar grid which gave them resolution of 8 pc on the inner boundary. They were able to produce nuclear rings at radii of approximately 1 kpc in models with ILR, IILR and fat bars. Ringed models produced lower mass inflow than unringed models.

A further parameter of barred galaxy models, the sound speed of the gas, was investigated by Englmaier and Gerhard (1997) using an smooth particle hydrodynamics (SPH) code. They find that shocks which were previously assumed to be offset can become centred if the sound speed rises above a critical value, $\sim$ 20 km s^{-1} or 10 % of the circular velocity.

Regan et al. (1999) showed that the dense part of the ISM responds to hydrodynamic forces in the same way as the diffuse component. Their observations of CO gas kinematics in seven barred galaxies verified a number of predictions from hydrodynamic models: the gas flows along the dust lanes; at the end of a dust lane, the gas sprays towards the far side of the bar; both fast and slow gas velocities are seen near the contact point of the dust lanes with a nuclear ring.

The simulations of Athanassoula and Bureau (1999) are used to develop diagnostics of edge-on barred galaxies from their position-velocity (PV) diagrams at various viewing angles. Certain structures in the PV diagrams are caused by the presence of x_1 and x_2 orbits and provide evidence for the existence and properties of a bar. Particularly characteristic structures are apparent if a nuclear spiral is present.

Motivated by the detection of nuclear spirals in a number of disk galaxies, Englmaier and

Shlosman (2000) explain the spirals' existence and shape using density wave theory and hydrodynamic simulations. The theory allows for spiral structure in non-self-gravitating gas inside an ILR. The simulations generate nuclear spirals extending from the bar shocks inwards to radii of $50 - 100$ pc, with a pitch angle that varies inversely with central concentration and directly with sound speed.

Maciejewski et al. (2002) confirmed that either nuclear rings or nuclear spirals can exist at the centre of barred galaxies depending on the gas sound speed. They also find that secondary bars can exist inside the ILR of a kpc sized bar but that, contrary to expectation, the smaller bar prevents gas inflow into the very centre of the galaxy.

2.3 N-body theory

2.3.1 Stellar dynamics as an N-body problem

The dynamics of stars in a spiral galaxy can be modelled by considering only the their mutual gravitational attraction i.e. by solving an N-body problem. Since the force of gravity follows an inverse square law, the motion of an individual star will be dictated as much by the stars on the far side of the galaxy as by those close by. In fact, the attraction of the nearest stars is rather weak and stellar collisions are extremely rare.

If we assume that the Galaxy is a disk of radius 10 kpc and thickness 1 kpc and that all $N = 10^{11}$ stars in it are like the Sun, then the mean free path for stellar collisions is $\lambda_\star \sim 1/n_\star\sigma_\star$, where the number of density of stars $n_\star = 0.3\,\mathrm{pc}^{-3}$, and their collisional cross section $\sigma_\star = 4\pi R_\odot^2$. Hence, the mean time between collisions $\tau_{\mathrm{coll}} = \lambda_\star/v \sim 10^{19}$ yrs, where $v = 40\,\mathrm{km\ s}^{-1}$ is the random velocity of stars.

N-body theory assumes that the mean potential of the galaxy determines the force on individual stars. However, even though the time between collisions is vast compared to the Hubble time, individual stars will still be deflected by each other. The time over which the combination of individual deflections leads to significant deviations from the expected motion in the mean galactic

potential is known as the relaxation time,

$$\tau_{\rm relax} = \frac{N}{8 \ln N}\, \tau_{\rm cross}, \tag{2.37}$$

where the crossing time $\tau_{\rm cross}$ is the time taken for a star to cross the Galaxy (Binney and Tremaine 1987). For our Galaxy, it would take $\sim 5 \times 10^8$ crossing times to perturb the smooth stellar dynamics. Since the Galaxy is only a few hundred crossing times old, it can be effectively modelled with N-body theory.

When N-body simulations of galaxies are performed, it is computationally too costly to use 10^{11} particles to represent the 10^{11} stars and instead the motion of each particle represents many stars of approximately similar position and velocity. In my simulations, $N = 2 \times 10^5$ and hence $\tau_{\rm relax} \sim 2000$ crossing times, much longer that the 8 crossing times of the simulations. However, artificially low values of N generate Poisson noise in N-body simulations which can lead to the development of unphysical structures.

2.3.2 Softened gravity

Since we are trying to model a collisionless system, the two body encounters and binaries that are so important in an *collisional* setup are wholly artificial here. Any attempt to model them would require a prohibitively small time step and needlessly take up large amounts of computing time. Therefore, a simple softened potential is used to weaken short range forces. At a distance r from a star of mass m, the softened potential is given by

$$\phi(R) = -\frac{Gm}{(R^2 + s^2)^{1/2}}. \tag{2.38}$$

This softened potential is equivalent to the potential in a plane offset by a distance s from a point mass, thereby providing an effective thickness to the 2D simulations. Hence, particles moving in a plane under softened forces behave as though distributed in a layer of half-thickness s or as if they were Plummer spheres of Plummer length s (Sellwood 1987).

Softened forces do not have much effect on the relaxation times of N-body simulations. Al-

though short range scattering is reduced, the bulk of the relaxation due to long range effects remains (Dehnen 2001). However, they do reduce the resolution of a simulation and affect the stability of a simulated galaxy as described in Section 2.4.

2.3.3 Particle-mesh codes

For an N particle system and the potential of Equation 2.38, the force on the ith particle is

$$\mathbf{F}_\alpha = \sum_{\beta}^{N} \frac{G m^2 \mathbf{R}_{\alpha\beta}}{(|\mathbf{R}_{\alpha\beta}|^2 + s^2)^{3/2}}, \tag{2.39}$$

where $\mathbf{R}_{\alpha\beta}$ is the displacement between the αth and β^{th} particles. The simplest method of solving Equation 2.39 is to sum directly over all particles. However, there are $\frac{1}{2}N(N-1)$ terms in each summation, so this is a very costly scheme computationally.

The method used in the DUAL code is a common one, known as particle-mesh or PM, whereby the particle masses are interpolated to nodes of a mesh or grid and then the forces are calculated at those points rather than at the position of the particles. Thus the number of force calculations to perform is limited by the number of nodes on the grid rather than by the number of particles. The forces are then interpolated back to the particle positions and the particles moved accordingly. Since it is the movement of the particles which takes the most time in these schemes, the computational cost just scales as N plus a constant overhead for the force calculation. The disadvantage with the PM method is the reduction in resolution and the correspondingly lower accuracy.

Using a grid produces softening-like effects in addition to those due to the softened potential. Due to the interpolation of the particle masses onto the grid, the density used in the force calculation is not the input density but its convolution with the interpolation kernel. Similarly, the forces used to move the particles are the forces calculated on the grid convolved with the interpolation kernel (Knebe et al. 2001). This reduces the forces at small distances from large structures e.g. near to the galaxy centre.

Although the softening-like effects of the grid changes the potential felt by the particles, it can still be approximated by Equation 2.38 if s is replaced by the somewhat larger value s_{grid}. Hence

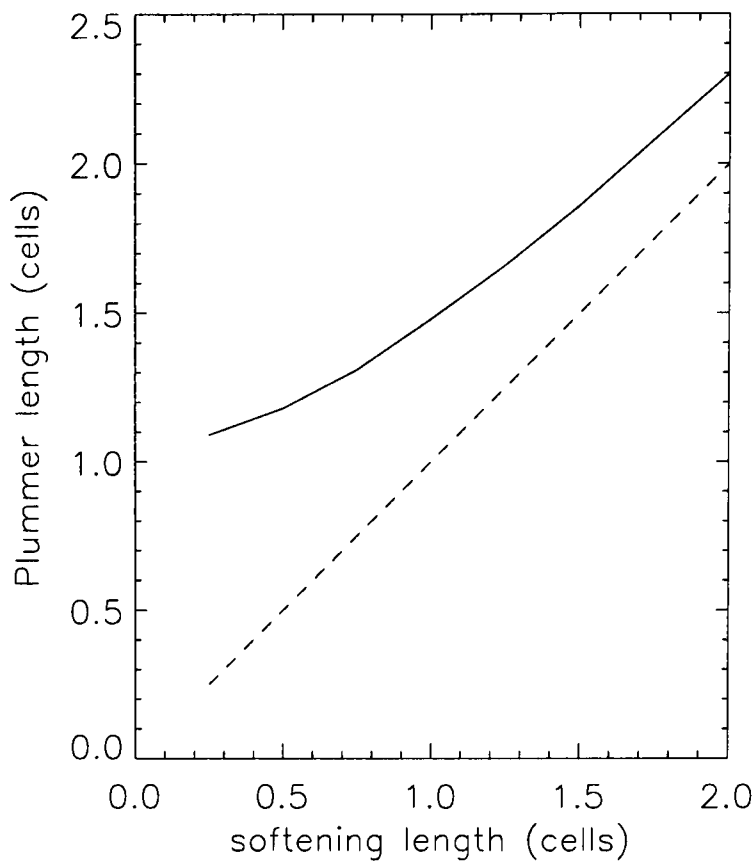


Figure 2.7: Relationship between the softening length s and the resulting Plummer length $s_{\rm grid}$. The dotted line marks the line of equal softening and Plummer lengths.

the effective thickness of the disk is also increased to $s_{\rm grid}$. The relationship between the softening length s and the Plummer length $s_{\rm grid}$ is shown in Figure 2.7. Limiting the softening length to $0.5 < s < 2.0$ cell lengths (Sellwood, private communication) restricts the Plummer length to $1.2 < s_{\rm grid} < 2.3$ cell lengths.

The next three paragraphs provide details of the schemes used in the DUAL code to interpolate the masses and the forces, calculate the forces on the grid and move the particles.

Interpolating the masses and forces The masses are interpolated onto the grid using the bilinear particle-in-cell scheme (PIC). A particle's mass is shared amongst the four nearest grid points in proportion to its fractional x and y displacements across the cell towards each grid point. The force on a particle is taken to be the weighted average of the forces on each of the four grid points with the weights equal to the fraction of the mass assigned to each point.

Calculating the grid forces Rewriting Equation 2.39 for the force $\mathbf{F}_k$ at each of the $k = 1, 2 \ldots K$ grid nodes gives

$$\mathbf{F}_k = m_k \sum_{l=1}^{K} G_{kl}\, m_l,$$
$$G_{kl} = \frac{G\,\mathbf{R}_{kl}}{(d^2 + |\mathbf{R}_{kl}|^2)^{3/2}}, \tag{2.40}$$

where m_k is the mass contained in the kth cell and $\mathbf{R}_{kl}$ is the displacement vector between the kth and l^{th} cells. Since G_{kl} depends only on the positions of the cells it only needs to be calculated once at the start of a N-body calculation. In fact, $\hat{G}_{kl}$, the Fourier transform of G_{kl} is calculated and stored instead.

The Fourier convolution theorem states that the Fourier transform of the convolution of two functions is equal to the product of the Fourier transforms of the same functions. In Equation 2.40, $\mathbf{F}_k$ is the discrete convolution of G_{kl} and m_l. Hence, using 2D fast Fourier transforms (FFT), $\mathbf{F}_k$ can be calculated efficiently by taking the inverse Fourier transform of $\mathbf{F}_k = G_{kl}\,\hat{m}_l$.

Moving the particles Once the forces on each particle have been calculated, their positions and velocities can be updated by solving Newton's second law. This is done using the popular leapfrog scheme, where for each particle

$$\mathbf{v}^{n+1/2} = \mathbf{v}^{n-1/2} + \frac{\mathbf{F}^n\,\Delta t}{m},$$
$$\mathbf{x}^{n+1} = \mathbf{x}^n + \mathbf{v}^{n+1/2}\Delta t, \tag{2.41}$$

where $\mathbf{F}^n$ is the force at the nth time step. This is a second order accurate scheme needing only one evaluation of the forces per time step. To accomplish this, as illustrated in Figure 2.8, the particle positions and forces are defined at integral time step whereas the velocities are offset by half a time step.

2.3.4 Previous N-body simulations of spiral galaxies

There are many examples of previous N-body simulations of spiral galaxies. I mention three.

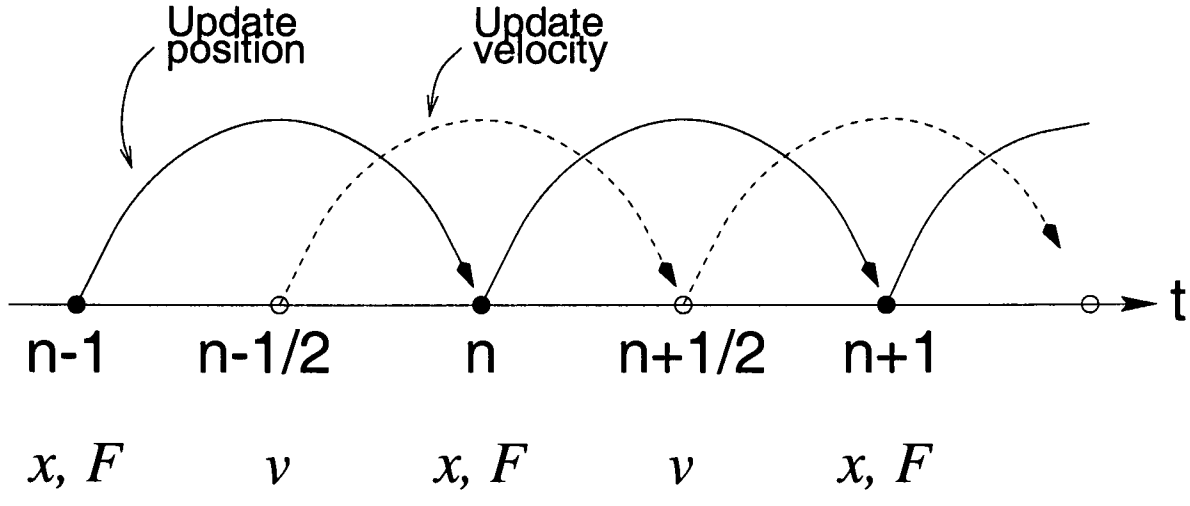


Figure 2.8: The second order leap frog time integration scheme. Positions x and forces F are stored at integral time steps and velocities v at half time steps.

An important early work was Hohl (1971). He used 100,000 particles to show that a bar could easily form in an unstable stellar disk. He was also able to create spiral structure by adding simple stellar cooling routines to his models.

Using a 90 $\times$ 70 polar grid, Sellwood and Carlberg (1984) created recurrent spiral structures which continued for $\sim$ 10 rotation periods, by continually adding stellar particles moving on circular velocities in order to mimic star formation due to infalling gas.

Recently, Debattista and Sellwood (2000) modelled the stellar disk and the dark halo using 300,000 N-body particles on a 129 $\times$ 129 $\times$ 129 Cartesian grid. They observed the bar slowing down due to the dynamical friction of the halo and hence, since almost all real bars have been observed to be fast, inferred that only minimal halos exist in galaxy centres.

2.4 Galactic stability

To determine whether a galaxy model is locally stable to gravitational collapse and thermal or centrifugal expansion, it is helpful to derive the dispersion relation for the frequency and wavelength of perturbing waves in the galaxy. The dispersion relation for a homogeneous, zero thickness gas disk in the limit of tightly wound spiral structures (the WKB approximation) is (Binney and Tremaine 1987),

$$(m\Omega - \omega)^2 = \kappa^2 - 2\pi G\,\Sigma\,|k| + k^2c^2, \tag{2.42}$$

where ω is the angular frequency of a perturbing wave, m is the azimuthal wavenumber or the number of identical spiral arms in the disk and k is the radial wavenumber.

For axisymmetric disturbances $m = 0$. If $\omega^2 < 0$ then ω is not real and the disk is unstable whereas if $\omega^2 > 0$, ω is real and the disk is stable. The line of neutral stability exists where $\omega = 0$ and thus

$$\kappa^2 - 2\pi G\,\Sigma\,|k| + k^2 c^2 = 0. \tag{2.43}$$

The disk is stable if there are no solutions to Equation 2.43 for all $k > 0$, implying that

$$Q_g \equiv \frac{c\,\kappa}{\pi\,G\,\Sigma} > 1, \tag{2.44}$$

where Q is the dimensionless measure of stability first introduced by Toomre (1964). The physical picture is that stability is ensured if thermal and centrifugal forces can overcome the inward force of self gravity.

The relations for a similar stellar disk were also derived by Toomre (1964) and follow those of the gas disk. Along the line of neutral stability, $\omega = 0$ and

$$\frac{|k|\,\sigma_R^2}{2\pi G\,\Sigma} = 1 - \exp\left(-\frac{k^2\,\sigma_R^2}{\kappa^2}\right)\,I_0\left(\frac{k^2\,\sigma_R^2}{\kappa^2}\right), \tag{2.45}$$

where σ_R is the radial velocity dispersion of the stars and I_0 is the modified Bessel function of the first kind of order zero. The disk is stable if there are no solutions to Equation 2.45, leading to

$$Q_\star \equiv \frac{\sigma_R\,\kappa}{3.36\,G\,\Sigma} > 1. \tag{2.46}$$

Stability relations can also be calculated for N-body simulations with the softened gravity of Equation 2.38. Miller (1971) and Miller (1974) show that the effect of gravitational softening is similar to that of a velocity dispersion, and that a dynamically cold ($\sigma_R = 0$) but softened system can be stabilized if

$$Q_{\rm soft} \equiv \frac{\kappa^2\,s}{2.31\,G\,\Sigma} > 1. \tag{2.47}$$

Chapter 3

Description of the DUAL code

3.1 Conceptual overview

The DUAL code is designed to simulate the dynamics of gas and/or stars in the plane of a spiral galaxy. The gas is modelled by solving the 2D Euler equations of hydrodynamics with gravitational source terms but neglecting self gravity. Analytic source terms can be supplemented by an N-body calculation which solves the 2D Poisson equation and Newton's Second Law to give the motion and gravitational potential of a distribution of stellar particles.

This Chapter describes the usage of the hydrodynamic and N-body solvers separately and together. The N-body solver itself is an amended version of a code written by Jerry Sellwood (Sellwood 1981). Its outline will be described, as will the subroutines which provide the interface to the hydrodynamic calculations.

The basic structure of DUAL is described in Section 3.2. The hydrodynamics are detailed in Section 3.3 which includes the workings of the C++ classes `Node`, `Grid`, `Setup`, `Dim`, `Sweep`, `Source`, `FS2` and `TwoDimPot`. Most of DUAL is written in C++ whereas the N-body solver is written in Fortran. Hence Section 3.4, which deals with the workings of the N-body solver, also contains information about the interface between the C++ and Fortran parts of the code. Use of the combined hydrodynamic and N-body solver is dealt with in Section 3.5. The driver functions used in particular implementations of DUAL are described in Section 3.6 and issues relating to performance, parallization and tests of DUAL are contained in Section 3.7.

This Chapter is best read in conjunction with the source code which can be downloaded from `www-thphys.physics.ox.ac.uk/users/SimonBerman`.

Table 3.1: The directory structure for the DUAL source code.

dual3-0/	driver files	
	input files	
	makefile, rebuild	
	classes/	C++ classes, functions.cc, getAnaPot.cc
	nbody/	galaxy.a
		c2d/ c2d*.f, admin.h
		gen/ foo.f, fortran.h, f77sub.h, admin.h
		nag/ FFT files
	tools/	libutil.a, libnr.a, header.h, util.h
blitz/		

3.2 Structural details

3.2.1 The source code files for DUAL

Table 3.1 shows the directory structure used in the DUAL source code. The files which are common to every implementation of DUAL are in the `classes/`, `tools/` and `blitz/` directories. Those files in `c2d/`, `gen/` and `nag/` are needed if the N-body solver is being used. The implementation specific files all sit in the uppermost directory `dual3-0/` and consists of the particular driver file and any input files.

The source code for each C++ class is stored in two files in the `classes/` directory. The header file, normally called `foo.h` for class `foo`, contains the declarations of the member functions and member data that make up the class. The source file is normally called `foo.cc` and contains the definitions of the member functions.

3.2.2 Using C++

By using C++ to write most of DUAL, I have tried to ensure that the code is written in a modular format so that it is simple to amend if any of the algorithms are changed. Although I have not fully exploited the object oriented nature of C++ there are a number of features of the language which have proved to be rather useful.

The use of *classes*[1] makes it possible to group common data and functions together. This removes the need for either excessive numbers of arguments being passed from function to function or having many global variables. The result more closely resembles the mathematics of the algorithm being coded, as can be seen in member functions of class `Sweep`.

Objects[2] have most obviously been used in the `Node` class which groups the density and x and y velocities of the gas in a grid cell into a single data structure. These quantities or combinations of them such as $\mathbf{Q}$, $\mathbf{U}$ or $\mathbf{v}$ can then be accessed using a variety of intuitively named member functions.

Lastly, the technique of *inheritance*[3] allows many specific potentials to be defined which all inherit the same basic structure of `TwoDimPot`.

One disadvantage of using C++ is that there is often a reduction in performance over Fortran or C, particularly when using inheritance. In order to overcome this I have made use of the Blitz++ library (Veldhuizen 2000) for scientific computing. This uses template techniques to achieve high performance on a par with Fortran 77/90. It also allows arrays to be used in a safe and intuitive manner. The disadvantage is that the Blitz++ library must be downloaded and compiled before running DUAL. Blitz++ arrays are declared as

```
Array<float,2> anArray(I,J);  // 2D array of floats
Array<Node,1> nodeArray(I);   // 1D array of Nodes
```

where `I` and `J` are the number of elements of the array in each dimension.

3.2.3 Installation and the makefile

Installation of DUAL is not particularly difficult, although it is time consuming to download the Blitz++ array class. The code has been tested and run on two operating systems: IRIX 6.5.12 and SunOS 5.7 both of which are UNIX based systems. The makefile distributed with the source code includes options for both of these systems. To install,

---

[1] A *class* is a user defined type containing data and functions that manipulate that data. `Node`, `Grid` and `Sweep` are all classes.

[2] An *object* is a particular instance of a class.

[3] *Inheritance* allows classes of subtypes to be defined which include all the member data and functions of the base class to which additional ones can be added.

- download and install the latest snapshot of the Blitz++ library, which can be found at `www.oonumerics.org/blitz/`,
- download the DUAL code from `www-thphys.physics.ox.ac.uk/users/SimonBerman`,
- ensure that the C++ compiler, home directory and extra libraries are correctly defined in the makefile,
- run `rebuild` to compile the `galaxy.a` N-body library,
- run `make` to compile the DUAL classes, the libraries `libnr.a` and `libutil.a` and the example programmes `morph`, `androm` and `nbodyrun`.

If the source code is adapted to run problems other than those described in this thesis, the makefile has to amended accordingly.

To run DUAL on SunOS, simply type `dual [run]`, where `[run]` is a three digit number identifying the simulation. All options are defined in the `.ini` input file. Under IRIX, a small shell script is needed like the example `dual001` found in the `dual3-0/` directory. Before use, the shell directory, run number and working directory must be amended. Further, `OMP_NUM_THREADS` must be set to the number of processors being used. The shell script can be run using the `batchq` command.

3.2.4 Miscellaneous functions

There are a small number of mathematical and utility functions in the `tools/` directory which are called on at various points during the running of DUAL:

`double bessi0(double)`
`double bessi1(double)`
`double bessi(int n,double)`
	return the modified Bessel function of the first kind of order zero, one or `n`,

`double bessk0(double)`
`double bessk1(double)`
`double bessk(int n,double)`
	return the modified Bessel function of the second kind of order zero, one or `n`,

`int numberOfLines(string f)`	return the number of lines in file `f`,
`float lineNumber(string w,string f)`	return the number of the first line in file `f` to contain the word `w`,
`float GetData(string w,string p,string f)` `int GetIntData(string w,string p,string f)`	return the first 10 characters of the first line in the file `p` to contain the word `w` and append the line to file `f`.

The declarations for these functions are contained in the header files `nr.h` and `util.h` and the corresponding library files are `libnr.a` and `libutil.a`.

One further file which should be mentioned is `header.h` which is included in every C++ file used in DUAL. It contains the code needed to use the C++ standard library and the Blitz++ library, some simple aliases for vector and array types and a definition of `pi`.

3.3 Hydrodynamics

3.3.1 The dynamic variables - class `Node`

The basic data structure used by DUAL is the `Node`. Each `Node` represents the state of the gas in a particular grid cell and consists of

```
class Node {
private:                        // in addition to member functions
   float        density;
   vector2      velocity;     // x and y velocities
   vector3      Q;            // primitive quantities
   vector3      U;            // conserved quantities
   static float soundSpeed;
};
```

where `vector2` and `vector3` are vectors of 2 and 3 elements.

`U` is the vector of conserved quantities $\mathbf{U} = (\Sigma, \Sigma u, \Sigma v)^T$ of Equation 2.1. Including it as member data saves numerous calculations of $\mathbf{U}(1)$ and $\mathbf{U}(2)$ during the hydrodynamic sweeps. It also enables the vector $\mathbf{U}$ to be coded in a much more intuitive way in class `Sweep`. Since the vector of primitive quantities $\mathbf{Q} = (\Sigma, u, v)^T$ is also called many times during a sweep, it is also sensible to include $\mathbf{Q}$ as a data member. The variable `soundSpeed` is `static` implying that the same value applies to every instance of a `Node` object.

There are no member data for position because in an array of `Nodes` it is trivial to determine the physical position of a `Node` from its position in the array. The class also contains member functions for accessing and setting the member data and `GetinvQ` and `GetinvU` for determining the values of $\mathbf{Q}$ and $\mathbf{U}$ in a `Node` rotated 180° about the centre of the galaxy.

3.3.2 Managing the grid, the nodes and the time - class `Grid`

Class `Grid` contains data relating to the computational grid and functions which manage it, the array of `Nodes` within it and the time step. Its member data include,

```
class Grid {
private:                        // in addition to member functions
   int        iSize, jSize;     // number of Nodes in x,y directions
   float      deltax,           // grid cell size
              deltat,           // time step
              theTime,          // time after initialization
              courant;          // Courant stability parameter
};
```

The computational grid is structured as shown in Figure 2.4. If the simulation is assumed to have 2-fold rotational symmetry, only half of the galaxy needs to be modelled and $J = 2\,I$. Otherwise, it is usual (though not necessary) to set $J = I$. The size of each cell Δx is determined once the user has specified the size of the grid, which normally corresponds to the size of part or all of a galaxy. Hence, $\Delta x = 2\,R_G/J$ where R_G is the shortest distance from the centre to the edge of the grid.

Important functions in class `Grid` include,

`CalcTimeStep`	uses Equation 2.27 to calculate the time step Δt,
`AdvanceTime`	adds either Δt or a specified amount to `theTime`,
`SetNodeArray`	an *overloaded function*[4] with a number of different combinations of arguments. It can be used to fill an array of `Nodes` from a velocity array and either a density array or a constant density. It is also possible to fill the ghost `Nodes` used to enforce the boundary conditions of Section 3.3.4 by including additional ghost `Node` arrays in the argument lists,
`GetScalarArray`	extracts an array of density, x-velocity or y-velocity values from the inputted `Node` array.

A number of functions are enclosed by `#ifdef NBODY` and `#endif` statements and are detailed in Section 3.5. These lines are only compiled if the identifier `NBODY` is defined in the makefile. Conditional compilation appears in a number of classes and ensures that the code needed to run the N-body solver is only compiled if the N-body solver is being used.

[4] *Overloaded functions* have the same name and generally perform similar tasks but differ in the number and/or types of their arguments

3.3.3 The FS2 algorithm - class FS2

Class FS2 controls the execution of an amended version of the FS2 algorithm of Equation 2.26:

```
class FS2 {
public:                                          // plus constructor
   void      UpdateGrid(Array<Node,2>&);
   void      InitialUpdate(Array<Node,2>&);
   void      FinalUpdate(Array<Node,2>&);
private:
   Grid&     gridOne;                            // manages the grid and time
   Sweep&    theSweep;                           // controls the hydro sweeps
   Source&   theSource;                          // controls the source terms
};
```

where its member data are simply *references*[5] to other data structures.

The amended version of the FS2 algorithm is designed to ensure that the potential is only calculated once per time step, a particular advantage when using a costly N-body solver. The basic idea, due to a private communication between myself, James Binney and Witold Maciejewski, is to merge the leading and trailing $\mathcal{S}^{\Delta t}$ operators of Equation 2.26 into the $\mathcal{S}$ operator of Equations 3.3 so that source terms are always calculated over a complete time step. Using a second order Runge Kutta scheme, the source terms are calculated at integer time steps n and serve to integrate the dynamical variables from $n - 1/2$ to $n + 1/2$.

However, this only produces the first half of the sequence outlined in Figure 3.1 and leads to problems. After one application of the $\mathcal{YX}$ operators the scheme is first order accurate and only becomes second order accurate when the $\mathcal{XY}$ operators are applied to restore symmetry. Therefore, the application of the source operator $\mathcal{S}$ after $\mathcal{YX}$ will also be only first order accurate. But, by extending the algorithm to cover four time steps in the manner of Figure 3.1, the error in the source terms can be cancelled. This is possible because in the second half of the algorithm, the direction of the sweeps are reversed. Hence, the source term errors now arise when $\mathcal{S}$ is applied after the $\mathcal{XY}$ operators and, by symmetry, cancel those due to the earlier application of the $\mathcal{YX}$ operators.

Unfortunately, the source operator now starts and ends on half-integer time steps. The solution

[5] A *reference* is an alternative name for an object. A reference `r` to an object `b` of type `T` is defined as `T& r = b;`. Any operation carried out on `r` is also carried out on `b`.

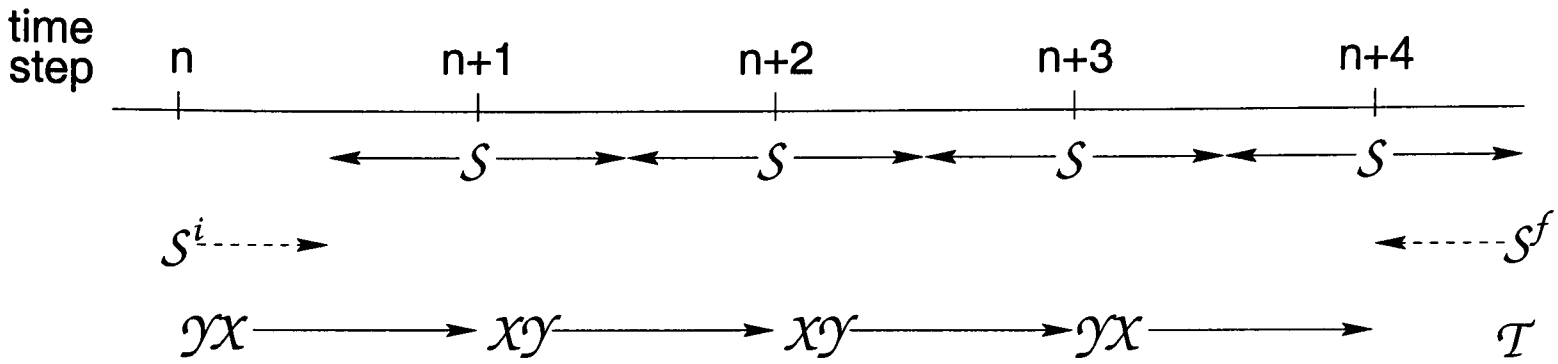


Figure 3.1: Schematic of the amended FS2 algorithm. The operators $\mathcal{S}, \mathcal{X}, \mathcal{Y}$ and $\mathcal{T}$ correspond to the second order Runge Kutta source term calculation, the x and y hydrodynamic sweeps and the calculation of the time step, respectively. Each operator acts at the time it appears in the figure and performs an integration over the duration of the arrow associated with it. The scheme is supplemented by the initializing source operator $\mathcal{S}^i$ and the finalizing source operator $\mathcal{S}^f$ as described in the text.

is to begin with an application of the Euler method for half a time step. This procedure is enacted in the member function `InitialUpdate` or the source operator $\mathcal{S}^i$ and uses the source terms $\mathbf{S}^0$, at $n = 0$, to integrate the dynamical variables $\mathbf{U}$ to $n = 1/2$,

$$\mathbf{U}^{1/2} = \mathbf{U}^0 + 0.5\,\Delta t\,\mathbf{S}^0. \tag{3.1}$$

The simulation is brought to a close by running the member function `FinalUpdate` or the source operator $\mathcal{S}^f$ which runs a second half-time step Euler method, but back in time from $n = N + 1/2$ to $n = N$,

$$\mathbf{U}^N = \mathbf{U}^{N+1/2} - 0.5\,\Delta t\,\mathbf{S}^{N+1/2}. \tag{3.2}$$

The central repeating part of the amended FS2 algorithm lasts for four time steps and is coded in the member function `UpdateGrid`. Thus the entire algorithm is

$$\begin{aligned} \mathbf{U}^{4n} &= \mathcal{S}^f\,(\mathcal{F})^n\,\mathcal{S}^i\,\mathcal{T}\,\mathbf{U}^0, \\ \mathcal{F} &= \mathcal{T}\,\mathcal{S}\,\mathcal{YX}\,\mathcal{S}\,\mathcal{XY}\,\mathcal{S}\,\mathcal{XY}\,\mathcal{S}\,\mathcal{YX}. \end{aligned} \tag{3.3}$$

3.3.4 A hydrodynamic sweep - class Sweep

Class `Sweep` contains functions for performing a one dimensional sweep across the grid according to the prescription of Appendix A. These functions are called via the member function `SweepGrid(XY,int)` where `XY` is an *enum*[6] defined as

```
enum XY { x = 1, y = 2 };
```

which allows the x or y direction of the sweep to be specified. Since it only updates one row or column of the grid, `SweepGrid` is called repeatedly from `FS2::UpdateGrid` in the following manner[7]

```
// 1D sweep of Nodes in the x direction over all values of j
for (int j = 0; j < jSize; j++) nodeArray(all,j) = theSweep.SweepGrid(x,j);
```

where `all` is a range object from the Blitz++ array class which ensures that all of the elements in the appropriate dimension are accessed.

The other member functions of Class `Sweep` are

`CalculateBias`	calculates the bias parameter ϵ as described in Appendix A,
`Q(i)`	returns $\mathbf{Q}_i$ subject to boundary conditions,
`DeltaQ(i)`	returns the slope vector $(\delta\mathbf{Q})_i$ of Equation A.4,
`Average(a,b)`	returns the average of a and b as defined by the van Leer limiter of Equation A.6,
`U(i)`	returns $\mathbf{U}_i$ subject to boundary conditions,
`FluxPlusMinus(±,i,±)`	returns the flux functions $\mathbf{F}^+(\mathbf{U}^{n+1/2}_{(i\pm1/2)-})$ or $\mathbf{F}^-(\mathbf{U}^{n+1/2}_{(i\pm1/2)+})$ of Equation A.1,
`QBoundary(Q,±,i,±)`	returns the cell boundary values $\mathbf{Q}^{n+1/2}_{(i\pm1/2)-}$ or $\mathbf{Q}^{n+1/2}_{(i\pm1/2)+}$ of Equation A.9,
`QHalfTime(Q,i)`	returns the half time step vector $\mathbf{Q}^{n+1/2}_i$ of Equation A.2,
`Q2U(Q)`	converts from $\mathbf{Q}$ to $\mathbf{U}$,
`U2Flux(U)`	converts from $\mathbf{U}$ to $\mathbf{F(U)}$,
`U2SplitFlux(U,M,±)`	converts from $\mathbf{U}$ to the split flux $\mathbf{F}^{\pm}$ of Equation A.11 which is determined by the Mach number M,

and Figure 3.2 indicates how each of them are called.

---

[6] An *enum* is a user defined set (or enumeration) of integers

[7] The scope operator `::` indicates which class a member function belongs to. So `FS2::UpdateGrid` is the member function `UpdateGrid` from the `FS2` class.

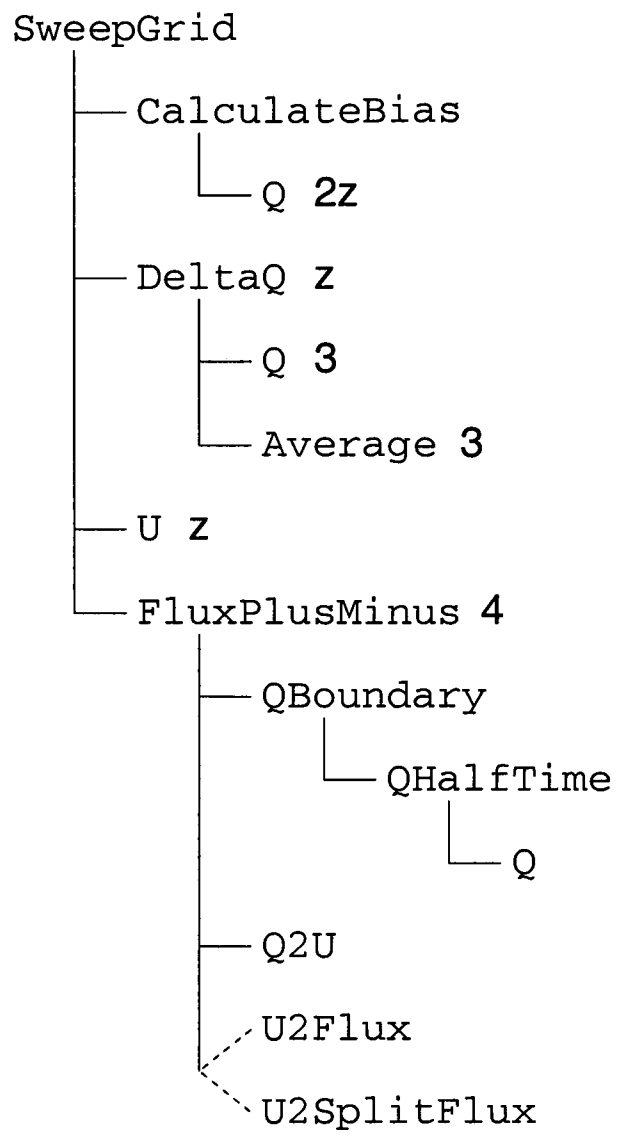


Figure 3.2: Calling tree for `SweepGrid` functions. The values next to some of the functions indicate how many times they are called from the function above them. z is the number of `Nodes` in the row or column being swept. Dotted lines indicate conditional function calls.

The member data of class `Sweep` include,

```
class Sweep {
private:                                  // in addition to member functions
   Array<Node,2>&  nodeArray,northNodes,southNodes,eastNodes,westNodes;
   Grid&           gridOne;               // manages the grid and time
   float           lambda;                // Δt/Δx
   const bool      rotSym;                // rotational symmetry of grid
   const int       plus,zero,minus;       // +1, 0, -1
   float           c;                     // bias parameter for slope limiting
   int             kpos;                  // nodeArray row or column to sweep
   XY              sweep;                 // direction of sweep
};
```

where the identifier `const` indicates that a member datum remains constant throughout the lifetime of the `Sweep` object and `bool` is a boolean type which can be only `true` or `false`.

Boundary conditions

The boundary conditions imposed on the grid are realized in the code through the use of two rows of ghost cells which surround the grid and are not updated according to the Euler equations but held at constant values. In a grid which is not rotationally symmetric (`rotSym` = 0) those ghost cells are stored in the `northNodes`, `southNodes`, `westNodes` and `eastNodes` arrays. Whenever the function `Q` calls on cells over the grid boundary i.e. when `i` $\geq$ `iSize`, `i` $<$ `0`, `j` $\geq$ `jSize` or `j` $<$ `0`, the ghost arrays are accessed.

However, if the grid is rotationally symmetric, the `eastNodes` remain unfilled. Instead, when `i` $\geq$ `iSize`, `Q` returns `nodeArray(jSize-1-ipos,jSize-1-jpos).GetinvQ()`, which is the vector **Q** from the ith node rotated 180° around the galaxy, $\mathbf{Q}^{\text{INV}} = (\Sigma, -u, -v)^T$.

3.3.5 Defining the gravitational potential - class `TwoDimPot`

DUAL allows any combination of 2D analytic potentials to be used irrespective of whether or not the N-body solver is included. Individual potentials are coded as separate C++ classes and inherit a basic structure from the base class `TwoDimPot`, which is declared as

```
enum haloType {exponentialDisk, isochroneHalo, plummerSphere,
               hubbleBulge, kuzminDisk, ferrersBar};

class TwoDimPot { // abstract base class for 2D potentials.  G = 1
public:
   virtual        ~TwoDimPot() {}
   virtual float  operator()(const float x,const float y,vector2*=0) const
   virtual bool   axisym() const { return false; }
};
```

The *virtual*[8] function `operator()` redefines the meaning of the brackets operator `()` so that it returns the value of the potential at a position (x, y). If a *pointer*[9] to a `vector2` is included in the list of arguments, the `vector2` will be set to the x and y potential gradients at that position. The function `axisym` should return true if the potential in question has cylindrical symmetric and false if not. Both functions must be redefined in any implementation of a class `TwoDimPot`.

If an axisymmetric potential is used, it is sensible to write functions for the potential and

[8] A *virtual* function is one which must be redefined in any inherited classes.
[9] A *pointer* `T*` stores the address or location in memory of an object of type `T`.

potential gradient in terms of the radius and then convert them to their x and y equivalents. At present, `TwoDimPot.h` contains declarations for the classes `ExponentialDisk`, `IsochroneHalo`, `PlummerSphere`, `HubbleBulge`, `KuzminDisk` and `FerrersBar`, and it a simple matter to use these classes as guidelines to writing user defined potentials. If user defined potentials will be used to drive N-body particles then the `getAnaPot.cc` file must also be updated (see Section 3.4.5).

The use of `TwoDimPot` objects in DUAL is described in Sections 3.3.6 on source terms, 3.3.7 on setting up the initial gas distribution and 3.6 on the driver programme.

3.3.6 Including source terms - class `Source`

Class `Source` contains the functions which perform the integrations of dynamic variables due to source terms. It also includes code to update the grid time and the N-body solver. The source terms are predominantly gravitational but can include Coriolis forces arising from a rotating frame of reference and gas recycling terms (see below). Member data include,

```
class Source {
private:                                  // in addition to member functions
   const float          patternSpeed,     // pattern speed of source terms
                        recycle;          // coefficient for gas recycling
   Array<float,2>&      initGasDen;       // initial gas density distribution
   Array<vector2,2>&    initGasVel;       // initial gas velocity distribution
   const TwoDimPot**    initPot;          // initial potential components array
   const int            qSize;            // number of initial components
   const TwoDimPot**    pot;              // potential components array
   const int            pSize;            // number of potential components
   bool                 axisym;           // axisymmetric potential flag
   Array<vector2,2>     initPG;           // initial non-axisymmetric forces
   Array<vector2,2>     finalPG;          // final non-axisymmetric forces
};
```

where `TwoDimPot**` is a pointer to an array of pointers to `TwoDimPot` objects. Gravitational forces are accessed through `pot`, where pointers enable the number of components included in the final potential to be specified in the driver programme and stored as `pSize`.

If the analytic potential is not axisymmetric, the non-axisymmetry is introduced during the first half a rotation period of the pattern. It is assumed that a non-axisymmetric potential is costly to calculate. Hence, the initial axisymmetric potential of `qSize` components, accessed via

the `initPot` pointers, is stored in the `initPG` array and the final non-axisymmetric potential, accessed via the `pot` pointers, is stored in the `finalPG` array. During the half rotation period the potential is linearly interpolated between `initPG` and `finalPG` and following this time, `finalPG` is used. If the potential is axisymmetric then it is always accessed via the `pot` pointers and calculated as needed.

The functions `ForwardEuler`, `BackwardEuler`, `RungeKutta` and `Integrate` carry out the amended FS2 algorithm as described in Section 3.3.3, with `Integrate` containing the mathematics common to all three functions.

Other functions include,

`AdvanceNBody`	calls functions in the N-body solver which integrate the particles over one time step. It is described more fully in Section 3.4.
`S`	returns the vector of source terms at the grid cell (i, j) as defined in Equation 2.2, adding the gas recycling term of Equation 3.4 if necessary.
`GetGravPotGrad`	returns the potential gradient vector at a grid cell (i, j). The potential gradient is calculated by summing the forces due to each component of the potential and adding those due to the N-body particles if required.

Gas recycling

During every simulation, the gas gradually loses angular momentum to numerical viscosity and falls inwards to the galactic centre. Based on Athanassoula (1992b), DUAL includes a routine which compensates for this by crudely simulating the interaction of stars with the ISM: star formation occurs and gas is removed in areas of high density; stellar mass loss due to stellar winds and supernovae, and hence gas replenishment, is assumed to take place at a steady rate across the galaxy. These phenomena are modelled by setting

$$\mathbf{R} = \alpha \begin{pmatrix} \Sigma_0^2 - \Sigma^2 \\ \Sigma_0^2 u_0 - \Sigma^2 u \\ \Sigma_0^2 v_0 - \Sigma^2 v \end{pmatrix}, \tag{3.4}$$

where α is the gas recycling coefficient set in the input files, Σ_0 is the initial gas surface density and u_0 and v_0 are the initial x and y velocities. $\mathbf{R}$ is then added to the source term vector $\mathbf{S}$ in Equation 2.2 and member function `S`.

3.3.7 Initializing the gas density and velocity - class `Setup`

Class `Setup` is used to initialize the gas surface density and velocity fields in the main grid and in the ghost `Nodes`. Its member data include,

```
class Setup {
private:                                  // in addition to member functions
   const int          iSize, jSize;       // grid size without ghosts
   const float        patternSpeed;       // pattern speed
   const vector2      gasDen;             // gas scale length and density
   const TwoDimPot**  pot;                // potential components array
   const int          pSize;              // no. of potential components
};
```

and its functions are,

`CalcInitialDensity`	calculates the initial gas surface density field, either as an exponential distribution of scale length `gasDen(0)` and scale surface density `gasDen(1)` or, if `gasDen(0) = 0`, as a uniform field of surface density `gasDen(1)`,
`CalcInitialVelocity`	calculates the initial gas velocity (u_0, v_0) from the rotation speed v_c due to the axisymmetric potential ϕ and the pattern speed Ω_p at the position of each grid cell (x, y),
`GetGravPotGrad`	calculates the forces in the same manner as `Source::GetGravPotGrad`.

The rotation speed and gas velocities are calculated using,

$$
\begin{aligned}
v_c &= (x\,\phi_x + y\,\phi_y)^{1/2}, \\
u_0 &= -\frac{v_c\,y}{R} + \Omega_p\,y, \\
v_0 &= \frac{v_c\,x}{R} - \Omega_p\,x,
\end{aligned}
\tag{3.5}
$$

where $R^2 = x^2 + y^2$.

3.3.8 Ensuring that variables are dimensionless - class `Dim`

It is helpful in computer simulations to cast physical variables in a dimensionless form with values close to unity. This is done by setting appropriate physical constants equal to unity and defining scaling variables in one or more dimensions.

For example, DUAL can be set up so that G is set to unity and the `Dim` class is used to define a scale length and scale mass. In the hydrodynamical version of DUAL, a less sensible set up is used in which G is set to unity, a scale length and scale velocity are defined and then an extra scale surface density is introduced. This means that volume densities and surface densities are dealt with using different units.

The scale units and derived quantities are represented as member data within the `Dim` class,

```
class Dim {
private:
   float  scaleLength, scaleDensity, scaleVelocity, scalePotential,
          scaleTime, scaleMass, scaleVolDens, scaleLum;
public:                                  // plus other member functions
   Dim(float,float);                     // G = 1, length, mass
   Dim(float,float,float);               // G = 1, length, surf dens, velocity
   void   DimForShocktube();             // change variables to shocktube test
   void   PrintToFile(ofstream&);        // prints scale variables to file
   float  LessFoo(float foo)             const { return foo / scaleFoo; }
   float  RestoreFoo(float foo)          const { return foo * scaleFoo; }
   float  DegToRad(float angle)          const { return angle * pi / 180; }
   float  RadToDeg(float angle)          const { return angle * 180 / pi; }
};
```

where `foo` refers to some or all of length, surface density, velocity, potential, time, mass, volume density and luminosity. The `Dim` class contains member functions for casting physical variables into a dimensionless format and for restoring them to their original dimensionality. There are also functions for converting from radians to degrees and back. Similar functions exist for converting 1D and 2D arrays.

3.3.9 The `main` programme

The main driver programme for a hydrodynamic simulation must be tailored to fit the users requirements. However, it can be heavily based on `andromeda.cc` and will normally follow the

outline described here and illustrated in Figure 3.3. Each instance of DUAL is labelled with a `name` which acts as the prefix for all input and output files. This is the same as or an abbreviation of the file name containing the driver programme e.g. `androm` and `andromeda.cc`. Each simulation run is given a three digit label `[run]` which is appended to `name` for use on all input and output files e.g. `androm001`. The input and output files are distinguished by their suffix e.g. `.ini`.

The `main` programme sets up all the relevant variables and objects, ensures that all variables are dimensionless, determines the initial conditions for the gas, sets the update scheme in motion and prints out the final results. Figure 3.3 describes the main subroutines called during `main`. Most of these routines have been described as member functions of one of the above classes. However, there are a few more subroutines contained in the file `functions.cc` which are common to all implementations of DUAL. In particular, `Update` contains the `for` loop which runs the FS2 algorithm by calling `FS2::UpdateGrid` as well as the initial and final routines `FS2::InitialUpdate` and `FS2::FinalUpdate`. `Update` also determines when to dump results to the output files.

The input file

The only input file used in purely hydrodynamic implementations of DUAL is `name[run].ini`. The variables defined within it are divided into those that must be set in any implementation and those which depend on the particular potential and initial conditions. The first set are,

```
80     // number of nodes in x direction:                    iSize
10000  // galaxy radius [pc]:                                galRad
3      // end time +ve=[bar rotations] -ve=[time steps]:     timeEnd
2      // time between data dumps [timeEnd]:                 dump
0.5    // Courant stability parameter, 0 < theta < 1:        courant
10     // sound speed [km/s] (10):                           soundSpeed
3e-10  // recycle coeff [pc^2/solar mass/yr] (3e-10):        recycle
```

where the square brackets indicate the units of each variable (the units of `dump` are the same as those of `timeEnd`). These units are implementation dependent. `galRad` is used to determine the size of each grid cell. The identifiers on each line following the : are the labels given to each variable within the driver programme.

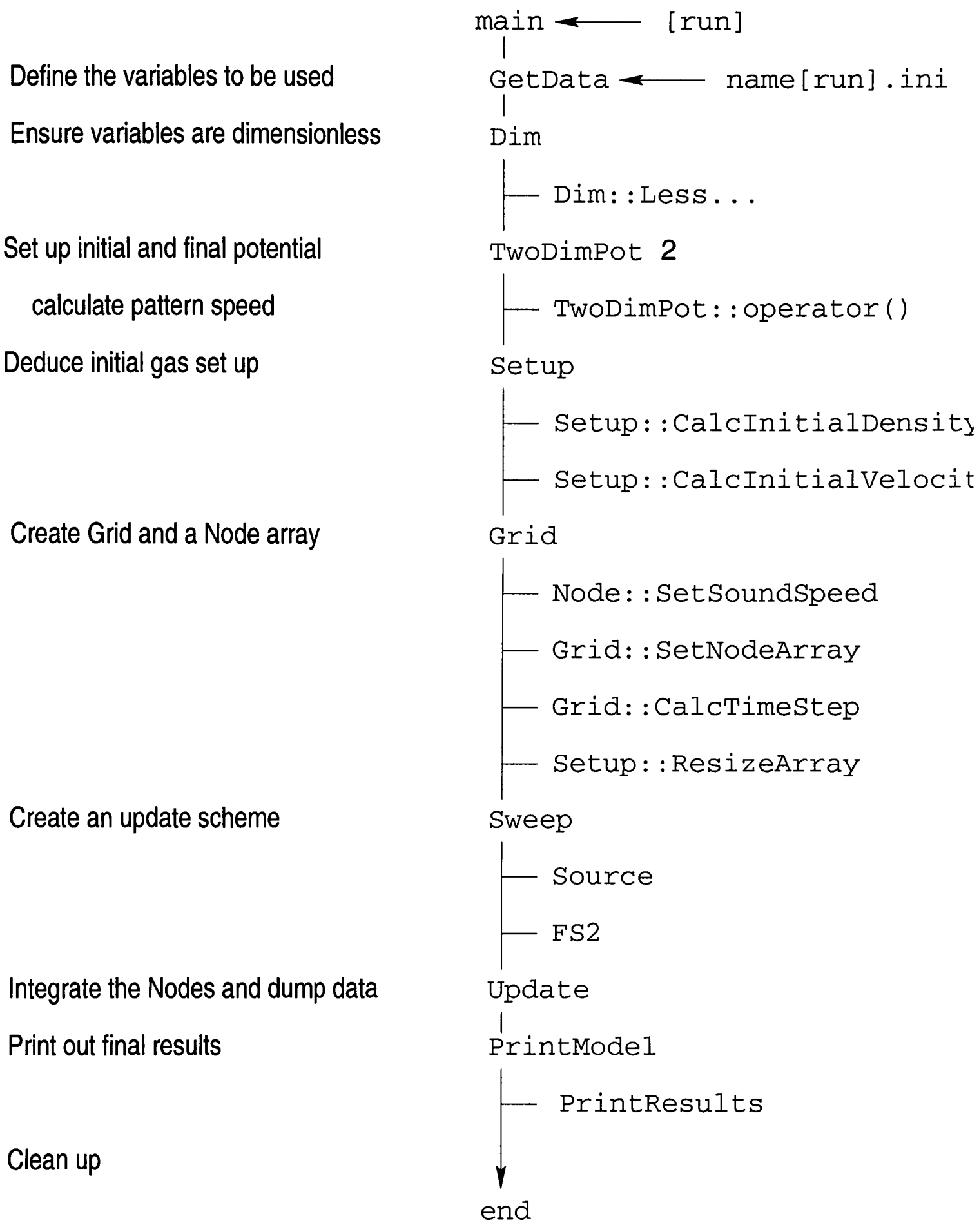


Figure 3.3: Calling tree for `main` hydrodynamics programme. The values next to some of the functions indicate how many times they are called. `[run]` is the three digit label given to each simulation. `name` is the prefix given to all input and output files. `Dim::Less...` refers to a number of different functions which can be used to ensure that variables are dimensionless.

Output files

Each run generates eight ASCII output files called `name[run].xxx`, where `xxx` is one of,

`.dat`	input parameters and calculated variables	appended during the run
`.tim`	time after every call to the FS2 algorithm	appended at every time step
`.gas`	gas surface density field	appended at every dump
`.xvl`	gas x velocity field	appended at every dump
`.yvl`	gas y velocity field	appended at every dump
`.div`	divergence of the gas velocity	appended at every dump
`.xpt`	x potential gradient	written at end of the run
`.ypt`	y potential gradient	written at end of the run

3.4 The N-body solver

3.4.1 Overview

The N-body solver used in DUAL is an amended version of a Fortran code written by Jerry Sellwood. It is a standard 2D Cartesian particle-mesh code as described in Section 2.3.3. The Poisson equation is solved using 2D Cartesian Fast Fourier Transforms, the accelerations being calculated directly rather than evaluated as numerical differences of the potential. The time integration follows the leap frog algorithm and so particle positions are advanced for half a step at the start of the simulation and backed up half a step before the particle coordinates are output. Interpolation on and off the grid is done via the particle-in-cell method. Lastly, particles which leave the grid during the simulation are discarded. As well as the addition of Fortran wrappers (described in Section 3.4.5), the original N-body solver has been amended in a number of ways, detailed in Section 3.4.4.

All N-body subroutines are found in the `c2d/`, `gen/` or `nag/` directories, where the file `foo.f` contains the subroutine `foo`. All Fortran subroutine names are written in lower case.

3.4.2 Dimensions and time step

As well as setting $G = 1$, there are two scale variables defining the internal dimensions of the N-body solver: `lscale` is the number of grid cells per unit of distance and `ts` is the time step measured in the units of time defined by the user. Both are determined from values set in

the `name[run].ini` file. Unlike hydrodynamics, the time step in an N-body simulation is not restricted by a Courant condition and stays fixed throughout the duration of a simulation. Its size is a compromise between speed, which argues for a large time step, and physics, which argues that the time step should be many times smaller than the smallest dynamical time scale in the simulation.

3.4.3 Selected routines and data structures

Most of the source code for the N-body solver has been fairly well commented by Jerry Sellwood and will not be described here. However, I will mention the subroutines that are called from the C++ files and some of the data structures that are used.

`galaxy` was originally the Fortran driver routine. It sets up the N-body grid and the Green's function for the Poisson solver using data contained in a short input file `galaxy.dat`, reads in the initial positions and velocities of the particles from the `gal[run].ini` input file and interpolates the particle masses onto the grid.

`dimens` is a subroutine in `galaxy.f` which contains parameters defining the maximum permitted number of particles and grid size. At present they are set as `mbod = 500000` and `mx = my = 256` respectively and must be amended if larger numbers are used.

`findf` calculates the forces on the grid due to the particles and any analytic potential.

`step` advances the positions and velocities of the particles by one time step, reassigns their mass to the grid and dumps particle data if requested.

`unload` centres particles' positions and velocities and dumps them to the `.fin` output file.

The x and y forces, the potential and the interpolated particle mass at each grid cell are stored in four sections of a 1D array called `grids`. Using Fortran indices[10], the first element of the section of the array in which each variable is stored is `grids(k*iSize*jSize + 1)`. The value of `k` is `0, 1, 2` or `3` for x forces, y forces, potential or mass respectively. To determine where in `grids` a variable corresponding to a particular grid cell is located, use the fact that the element `(i,j)` in a 2D C++ array in which `i` and `j` are numbered from `0` and rows are filled before columns, corresponds to the element `(j*iSize + i + 1)` in a 1D Fortran array.

During the time integration, the particles are operated on in groups of `mbuff = 1000`. The data for each group in turn is read from a linked list, `ptcls`, and stored in a 2D array called `oldc`. In this array, using Fortran indices, the particles' dynamical variables are stored at `oldc(k,l)`,

[10]The first element of an array `A` is `A(1)` in Fortran but `A(0)` in C++

where `k = 1, 2, 3` or `4` for x position, y position, x velocity or y velocity, respectively and `l = 1, 2 ... mbuff`.

Each run generates four output files called `galaxy.xxx` where `xxx` is one of the following,

`.lis`	ASCII	brief summary of progress	appended throughout run
`.fin`	ASCII	final coordinates of all the particles	written at end of run
`.qvc`	ASCII	radial profile of Q and v_c	appended at every dump
`.res`	binary	particle coordinates and potentials	appended at every dump

plus two additional binary files, `galaxy.grd` and `galaxy.grt`. These last two files contain the Green function used for the field determination (`.grd`) and its more useful Fourier transform (`.grt`). If these files are deleted after a run, the software will re-create them should the need arise, but will not do so if they exist. They should be deleted if the grid size or softening length is changed because then new versions are required.

3.4.4 Amendments to the N-body solver

The most important changes are to the `getacc` subroutine, which adds any external forces to those of the particles. This subroutine now accesses the `TwoDimPot` C++ class to get the background potential. The positions and velocities of the particles are initialized to match user defined surface density, radial velocity dispersion and circular speed profiles as closely as possible. However, the initialization routine used for the project of Chapter 5 produces too little surface density in the central regions. Since the velocities of the particles are matched to a pre-defined circular speed curve but there is less mass at the centre than expected, the particles expand outwards, creating spurious breathing modes. Thus the `getacc` routine was corrected to ensure that the axisymmetric part of the force felt by the particles always gives the pre-defined circular speed curve. This procedure is described in more detail in Section 5.2.3.

Both the particle radial force and the total radial force are tabulated as a function of radius in an amended `step` subroutine. Then, in `measure`, these and the radial velocity dispersion profiles (tabulated in the `icheck` subroutine) are used to calculate and dump to the `.qvc` file the quantities $Q(R)$ and $v_c(R)$. All of the additional global parameters that have been introduced by these amendments are included in the `admin.h` header file.

3.4.5 The C++ Fortran interface

Accessing Fortran from C++

Since the N-body code was written in Fortran 77 but DUAL was written in C++ a number of wrapper files are used as an interface between the two languages. These wrappers allow Fortran functions and global variables to be accessed from C++ code. The basic idea, due to Carsten Arnhlom (Arnholm 1997), is to define macros which make it possible to declare C++ prototypes for Fortran subroutines. The three files which do this are,

`fortran.h`	defines the macros and declarations that underpin the interface,
`f77sub.h`	declares the functions that are used in DUAL,
`f77char.h`	defines the `Character` class, enabling a `char` to be passed as an argument.

where the first and last files were written by Arnholm.

Two examples from the file `f77sub.h` illustrate how this interface works:

```
#define SUBROUTINE_F77 extern "C" void
#define SUBROUTINE inline void
typedef int     INTEGER;
typedef float   REAL;

SUBROUTINE_F77  galaxy_();
SUBROUTINE      galaxy()                          { galaxy_(); }

SUBROUTINE_F77  getxpg_(float&, int&);
SUBROUTINE      getxpg(REAL& v1, INTEGER& v2)  { getxpg_(v1, v2); }
```

where all variables are passed by reference (the standard in Fortran 77) and a `typedef` simply creates a synonym for a particular type. The declaration of the function `galaxy` enables access in C++ code to the Fortran function of that name. `getxpg` is a wrapper function which sets `v1` to the value of the x potential gradient at position `v2` in the `grids` array of the N-body solver.

The other wrapper files used in DUAL are,

`addtstep()`	adds 1 to the N-body time step counter,
`getmass(REAL& v1, INTEGER& v2)`	sets `v1` to the mass at `grids(v2)`,
`gettstep(INTEGER& v1)`	sets `v1` to the N-body time step counter,
`getypg(REAL& v1, INTEGER& v2)`	sets `v1` to the y force at `grids(v2)`,
`sethalo(REAL&, REAL&, REAL&, INTEGER&)`	sets the halo in the N-body solver,
`setpwd(CHARACTER v1, INTEGER& v2)`	sets the present directory to `v1`,
`settstep(INTEGER&)`	sets the N-body time step counter to `v1`,

all of which are declared in `f77sub.h`.

Accessing C++ from Fortran

All of the above involved accessing Fortran from C++. However, there is one point where the Fortran code needs to access a C++ function. This occurs during `getacc` where access is needed to the `TwoDimPot` C++ class to determine the background forces acting on the particles.

Class objects cannot be accessed from Fortran code. Hence a wrapper file, `anapot.f`, calls the C++ function `getanapot_` defined in the file `getAnaPot.cc`. This creates the necessary `TwoDimPot` objects and returns the forces due to them at the appropriate positions. A flag, `halotype`, is defined in `TwoDimPot.h` and set in the driver programme to indicate which potential should be accessed. The parameters of the potential are also set in the driver programme and passed through `anapot` to `getanapot_` as required. At present, `anapot` can only call axially symmetric potentials.

3.4.6 The `main` programme

It is perfectly possible to run a simulation involving just the N-body particles and a background potential. Even if the final goal is to combine N-body and hydrodynamic techniques, it is useful to be able to run the particles by themselves to check that they are performing correctly.

The file `nbodyrun.cc` contains a `main` programme which sets up all of the relevant variables and the halo, sets the update scheme in motion, integrates the N-body particles and prints out the final results. Figure 3.4 highlights the subroutines that are called during `main`.

`plummer2soft`	performs the conversion shown in Figure 2.7 between the desired effective height of the simulated galaxy (the Plummer length) and the softening length which produces it when combined with grid effects.
`AmendDatFile`	amends an old version of the `galaxy.dat` input file to ensure that the values in it are correct.

`galaxy`, `findf`, `step` and `unload` are all Fortran functions called via declarations in the header file `f77sub.h` and described in Section 3.4.3.

The input files

The two input files needed to run the N-body solver are `name[run].ini` and `gal[run].ini`. The variables defined in `name[run].ini` are divided into those that must be set and those which depend on the particular potential and initial conditions. Those that must be set are

```
200000   // number of N-body particles:                        numParticles
250      // number of nodes in x direction:                    iSize
7.5      // galaxy radius [scale lengths]:                     galRad
160      // end time +ve=[scale times] -ve=[time steps]:       timeEnd
2        // time between data dumps [timeEnd]:                 dump
0.01588  // time step length [scale time]:                     timeStep
0.0748   // scale height of stellar disk [scale lengths]:      height
```

where the square brackets indicate the units of each variable (the units of `dump` are the same as those of `timeEnd`). These units are implementation dependent. The identifiers on each line following the : are the labels given to each variable within the driver programme. `height` is set equal to the Plummer length and used to derive the softening length.

The `gal[run].ini` file consists of a header line and then a line consisting of the (x, y, u, v) coordinates for each particle. The header line looks like

$t_0 \quad M_T \quad N \quad 4$

where t_0 is the initial time of the run, M_T is the total mass of the particles, N is the number of particles and 4 represents the number of coordinates per particle.

3.5 Combining hydrodynamic and N-body solvers

3.5.1 Hydrodynamic functions which access N-body routines

The reason for combining the hydrodynamic and N-body solvers is to enable the N-body particles to contribute to the background potential felt by the gas. This is accomplished primarily within the function `Source::GetGravPotGrad` by adding the forces due to the particles to those of the analytic background potential. Since the `Source` class controls the hydrodynamic time stepping it

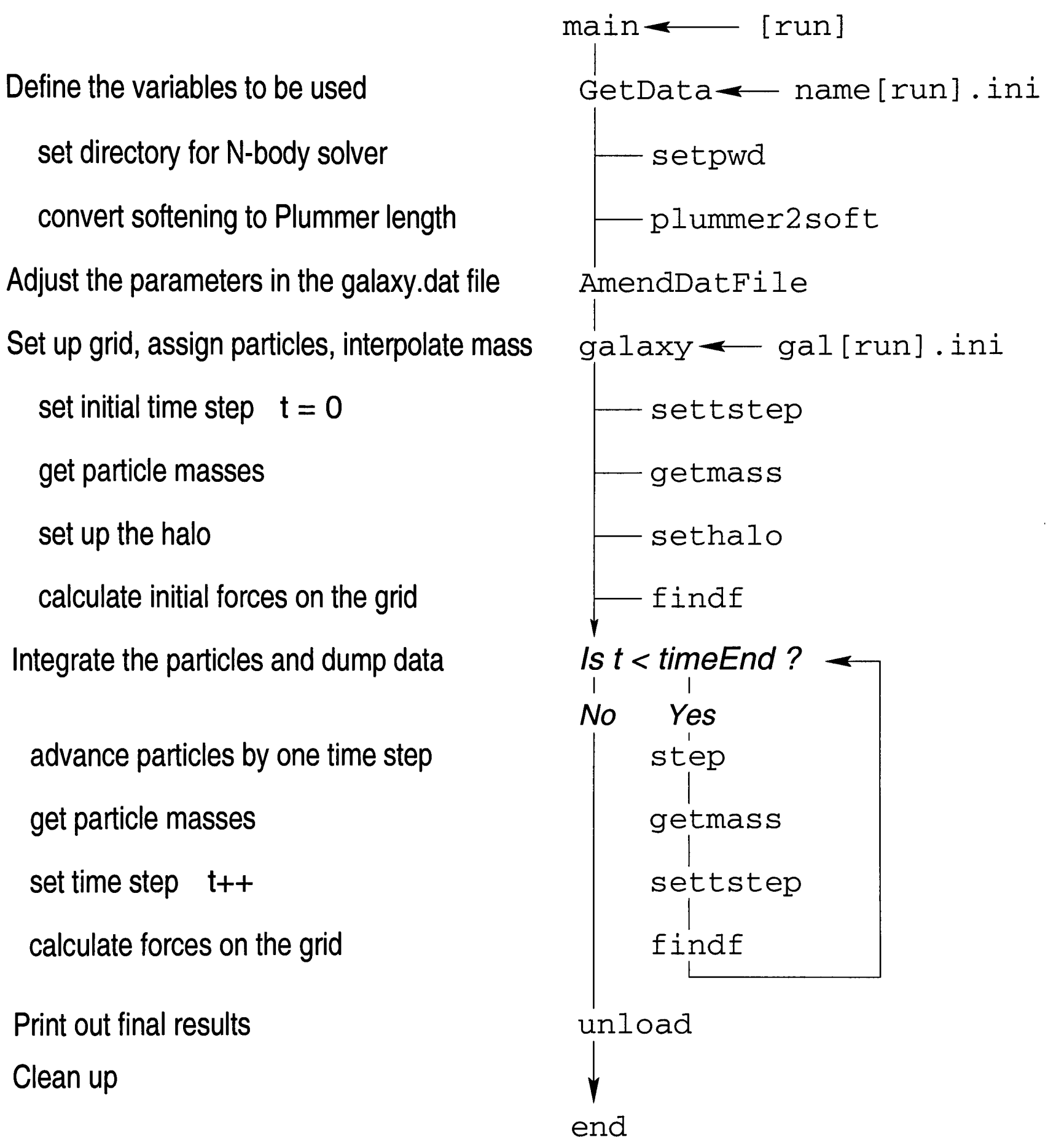


Figure 3.4: Calling tree for `main` N-body programme. `[run]` is the three digit label given to each simulation. `name` is the prefix given to all input and output files.

is logical that it also moves the N-body particles forward in time. Hence, `Source::AdvanceNBody` calls the subroutines `step`, `Grid::StoreMasses`, `addtstep` and `findf`.

During simulations, the hydrodynamic and N-body integrations must be kept in step with each other. Hence the time step in one solver must be a multiple of that in the other. Figure 3.1 shows that the N-body time step cannot be larger than the hydrodynamic time step since the potential is needed at every hydro time step. The most sensible practice is to set the two time steps equal to one another, and this is what is done.

The hydrodynamic time step Δt is controlled by the Courant condition and updated during a simulation. Yet the Courant condition only sets an upper bound on Δt and we are allowed to set Δt to be lower than this, although it is inefficient to do so. Since the time step as calculated by the Courant condition rarely deviates by more than 30% from its initial value, the initial time step is made artificially low by multiplying by the N-body time coefficient $0 < \varphi < 1$ and checked periodically. This leads to the following algorithm:

Calculate the initial time step Δt using the Courant condition	`Grid::CalcTimeStep`
Multiply the initial time step by φ	`Grid::SetDeltat`
After every FS2 sequence, propose a new time step $\Delta t_{\rm new}$ calculated from the Courant condition	`Grid::CalcTimeStep`
compare Δt with $\Delta t_{\rm new}$	`Grid::CheckTimeStep`
if $(\Delta t > \Delta t_{\rm new}$ or $5\,\Delta t < \Delta t_{\rm new})$ stop the N-body solver, reset Δt to $\varphi\,\Delta t_{\rm new}$ and restart the N-body solver	`Grid::ResetNBody`

All of the code fragments in the C++ files that are only used if N-body particles are present, are enclosed within `#ifdef NBODY` and `#endif` statements.

3.5.2 The `main` programme

The `main` programme for the combined hydrodynamic and N-body solvers is, as expected, an amalgamation of the two previously described `main` programmes and is illustrated in Figure 3.5. N-body particles are added into the potential in `Source::GetGravPotGrad`, buried deep in the update section of `main`. The fixing and keeping track of the time step leads to the addition of `Grid::SetTimeStep` in Figure 3.5. There is no mention of the pattern speed since it is set to zero.



Figure 3.5: Calling tree for `main` combined hydrodynamics and N-body programme. `[run]` is the three digit label given to each simulation. `name` is the prefix given to all input and output files. The values next to some of the functions indicate how many times they are called. `Dim::Less...` refers to a number of different functions which can be used to ensure that variables are dimensionless.

Inputs and outputs

The input files that are needed to run the combined programme are `name[run].ini`, `gal[run].ini` and a version of `galaxy.dat`. The variables defined in `name[run].ini` either must be set or depend on the particular potential and initial conditions. Those that must be set are

```
200000  // number of N-body particles:                          numParticles
250     // number of nodes in x direction:                      iSize
7.5     // galaxy radius [scale lengths]:                       galRad
160     // end time +ve=[scale times] -ve=[time steps]:         timeEnd
2       // time between data dumps [timeEnd]:                   dump
0.6     // N-body time coefficient:                             nbodydt
0.5     // Courant stability parameter, 0 < theta < 1:          courant
0.025   // sound speed [scale length/scale time] (10 km/s):     soundSpeed
0       // recycle coeff [pc^2/solar mass/yr] (3e-10):          recycle
0.0748  // scale height of stellar disk [scale lengths]:        height
```

where the square brackets indicate the units of each variable (the units of `dump` are the same as those of `timeEnd`). These units are implementation dependent. Note the absence of `timeStep` since this is calculated internally. The Plummer length necessary to derive the softening length is given here by the scale height of the disk, `height`. As described above, `gal[run].ini` consists of a header line and then a line consisting of the (x, y, u, v) coordinates for each particle.

There are twelve `ASCII` output files. The first nine are called `name[run].xxx` and the last three are called `gal[run].xxx` where `xxx` is one of the following,

`.dat`	input parameters and calculated variables	appended during the run
`.tim`	time after every call to the FS2 algorithm	appended at every time step
`.sta`	stellar surface density	appended at every dump
`.gas`	gas surface density field	appended at every dump
`.xvl`	gas x velocity field	appended at every dump
`.yvl`	gas y velocity field	appended at every dump
`.div`	divergence of the gas velocity	appended at every dump
`.xpt`	x potential gradient	written at end of the run
`.ypt`	y potential gradient	written at end of the run
`.lis`	brief summary of progress	appended during the run
`.fin`	final coordinates of all the particles	written at end of the run
`.qvc`	radial profile of Q and $v_c(R)$	appended at every dump

3.6 Particular implementations of DUAL

There are a number of parts of DUAL that must be tailored to each specific implementation, whether using the N-body or the hydrodynamic solver or both. In particular, amending the `makefile`, the driver file (here called `name.cc`), the input file `name[RUN].ini` and, if a new potential is being used, the `TwoDimPot` class in `TwoDimPot.h` and `TwoDimPot.cc`. If using the N-body solver, `getanapot.cc` and the particle input file `gal[RUN].ini` may also need to be changed.

In the `makefile`, presuming that the correct paths have been defined for the compiler libraries, the variable `NBODY` should be defined if using the N-body solver and remain undefined otherwise.

In `name.cc`, a number of parameters need to be set:

`pwd`	the present working directory,
`name`	the label given to this implementation of DUAL,
`jSize`	the number of y nodes e.g. `jSize = iSize` or `2 * iSize`,
`rotSym`	the rotational symmetry of the grid to be either `true` or `false`,
`...`	the parameters which define the potential being used,
`patternSpeed`	the pattern speed of the potential.

If using the hydrodynamic solver, the following must also be set:

`...`	the parameters which define the initial gas surface density,
`Dim()`	the correct constructor for class `Dim`,
`Dim::LessFoo`	the use of `Dim` to make all dimensional parameters dimensionless,
`pSize, qSize`	the number of components in the initial and background potentials,
`setupPot, pot`	the two arrays of `TwoDimPot` pointers which define the potentials.

3.6.1 Modelling the Andromeda galaxy

For the project of Chapter 4, in which gas flows in the bulge of the Andromeda galaxy were modelled, only the hydrodynamic part of DUAL was used. Much effort was spent on designing the `M31Pot` class, derived from `TwoDimPot` and coded in the `m31pot.h` and `m31pot.cc` files. `M31Pot` contains member functions which calculate the forces due to the potential described in Section 4.2.

In terms of the parameters mentioned above, `name = androm, jSize = 2 * iSize, rotSym = true, pSize = 1` and `qSize = 1`. The pattern speed was calculated using,

```
vector2  potGrad;                                    // y axis at corotation
float    coRotRad = 1.2 * m31->y_size();             // m31 is a M31Pot object
m31->operator()(float(0), coRotRad, &potGrad);
float    patternSpeed = sqrt(potGrad[1] / coRotRad);
```

3.6.2 Modelling stellar and gaseous morphology

In Chapter 5, I describe a project to model the stellar and gaseous morphologies in barred spiral galaxies. This was accomplished using both the hydrodynamic and N-body solvers in DUAL. The background halo potential was modelled using the `IsochroneHalo` class. The construction of the particles' positions and velocities for the input file `gal[RUN].ini` is described in Section 5.2.3.

In terms of the parameters described above, `name = dual`, `patternSpeed = 0`, `rotSym = false`, `jSize = iSize`, `pSize = 2` and `qSize = 1`.

3.7 Tests and performance

3.7.1 Axisymmetry test

A simple and effective test for hydrodynamic codes used in spiral galaxy simulations is to model the hydrodynamics of an axisymmetric galaxy to determine how much gas infall there is due to numerical viscosity. To do this, I set up the grand design galaxy model described in Chapter 5, but without the N-body forces. Instead, I kept the potential at its initial analytic and axisymmetric distribution. The simulation lasted for 160 time units, or 1.6 Gyr if the disk scale length and disk mass are set to 5 kpc and 2×10^{11} $M_\odot$.

By the end of the simulation, the gas surface density had changed by more than 1% in only 148 out of the 62500 grid cells (or 0.2%). The surface density of the central four cells rose by just 50% implying a very low numerical viscosity.

3.7.2 Shocktube tests

The analytic results of the one dimensional isothermal shocktube problem of Section 2.1.2 are used here to gauge the ability of DUAL (and the FS2 algorithm within it) to capture shocks.

The shocktube problem is solved in two scenarios: directly onto a one dimensional, 50 cell grid, 20 m in length and obliquely across a two dimensional, 80×80 cell grid, 20×20 m in area.

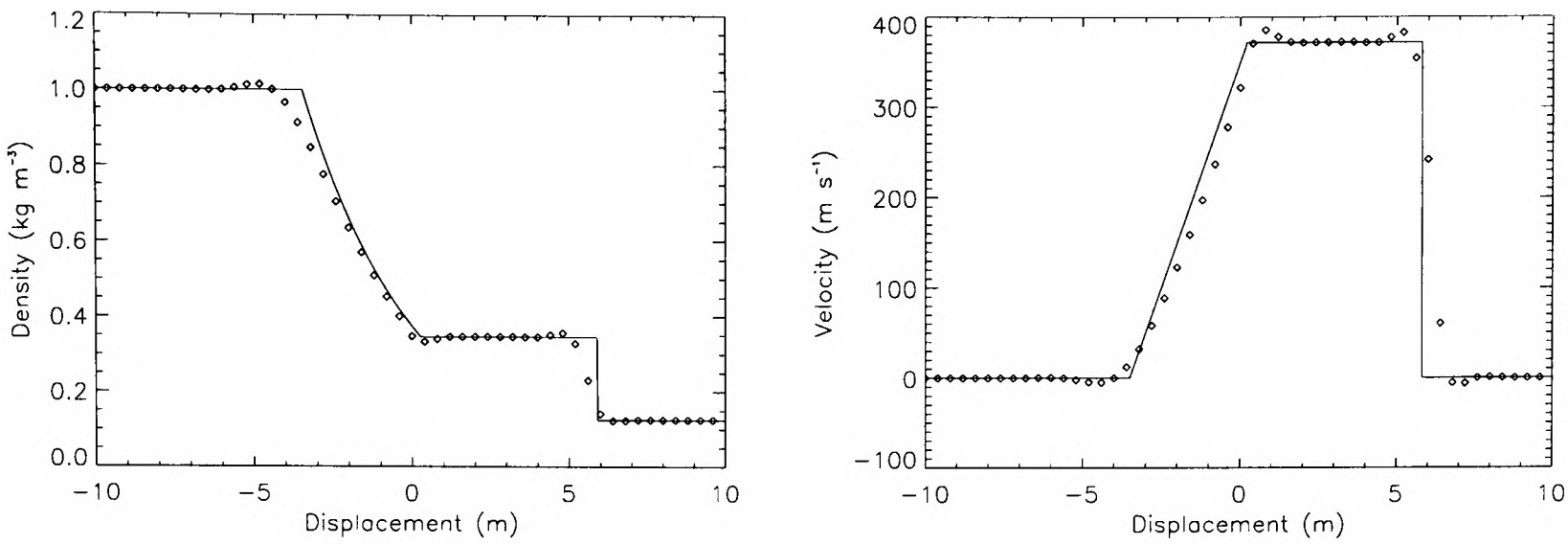


Figure 3.6: Density and velocity for one dimensional isothermal shocktube test after 0.01 s, with sound speed $c = 350\ \mathrm{m\,s^{-1}}$. The solid line shows the analytic result of Section 2.1.2 and each diamond indicates a grid cell in the numerical result. The right and left boundaries maintain their initial conditions.

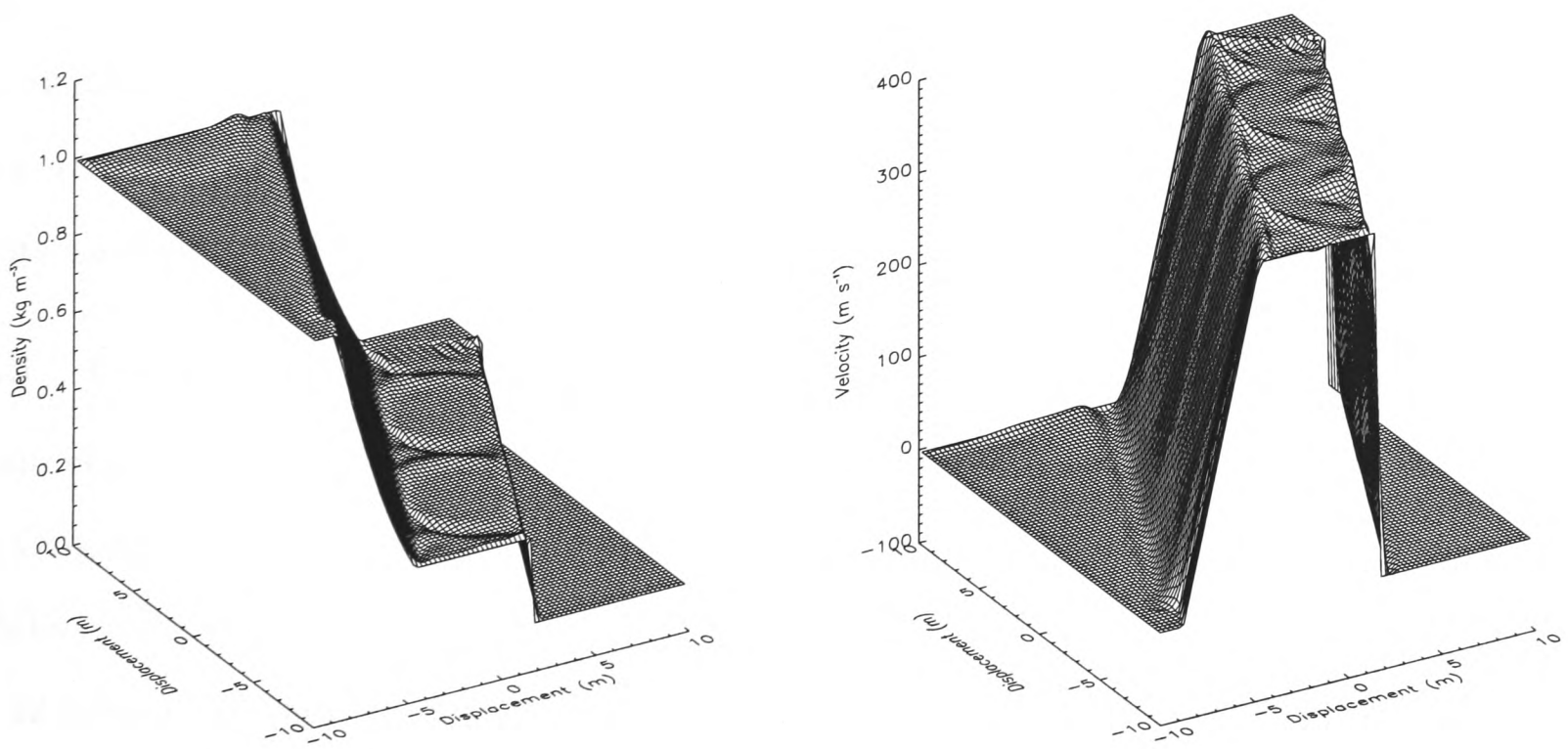


Figure 3.7: Density and velocity parallel to the motion of the shock for a one dimensional isothermal shocktube obliquely on a two dimensional grid after 0.01 s, with sound speed $c = 350\ \mathrm{m\,s^{-1}}$. The boundary conditions for the simulation were simply the analytic solution which can be seen at the front and back boundaries. The right and left boundaries maintain their initial conditions.

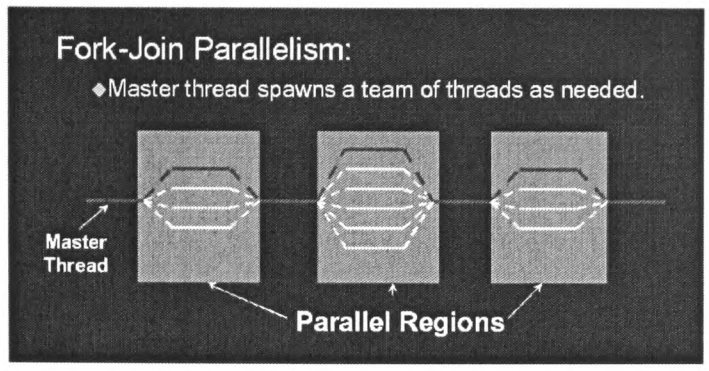


Figure 3.8: Fork-join parallelization model.

In both cases the sound speed $c = 350\ \mathrm{m\,s^{-1}}$ and the solution is viewed after 0.01 s. Results are shown for the density and the velocity parallel to the motion of the shock: the one dimensional grid in Figure 3.6 and the two dimensional grid in Figure 3.7.

Figure 3.6 shows that away from the shock, DUAL reproduces the correct density and velocity and that, in general, the analytic solution is reproduced by the code. The shock itself is captured in three grid cells rather than the two that van Leer (1982) says is possible. This may be due to the low resolution of the simulation, since in the higher resolution Figure 3.7, the shock is captured in just two cells. A bigger worry is the spurious wobbling of the velocity profile in Figure 3.6 near to the shock and the rarefaction wave and the wave formation at the top of the velocity profile in Figure 3.7. But in both cases the amplitude of the anomaly is only about 5% of the undisturbed velocity and similar to the amplitude of velocities perpendicular to the shock motion.

3.7.3 Oscar and parallelization

The simulations described in this thesis have all been run on Oscar, a 96 node Silicon Graphics Cray Origin2000 parallel computer. It consists of 96 MIPS 195 MHz R10000 processors each with 256 Mbytes RAM for a total shared memory of 24 Gbytes and peak computing capacity greater than 32 Gflops. My simulations typically use 10 processors or less. Oscar supports the Fortran 77 and C++ programming languages via its SGI MIPSpro compilers which implement the OpenMP standard for shared-memory parallel programming directives[11]. Directives are embedded into the DUAL source code via preprocessor `#pragma` commands. Four points within DUAL take advantage

[11]In shared memory machines every processor has access to all of the memory. This contrasts with distributed memory machines, where each processor has its own independent memory and if one processor requires data contained in another processor's memory, messages must be passed between them.

of Oscar's parallel capabilities. All four involve sharing a `for` loop out amongst some of Oscar's processors. An example from class `Sweep` will serve to illustrate how the directives work:

```
#include <omp.h>
#pragma omp parallel shared(maxSize,result) private(z) {
   #pragma omp for schedule(static)
   for (z = 0; z < maxSize; z++) {
      result(z) = U(z) - (lambda * (FluxPlusMinus(plus,z,plus) +
                                    FluxPlusMinus(minus,z,plus) -
                                    FluxPlusMinus(plus,z,minus) -
                                    FluxPlusMinus(minus,z,minus)));
   }
}
```

where `omp.h` is the header file containing definitions of the parallel directives. The `omp parallel` directive splits the master thread into a master and slaves as shown in Figure 3.8. The `shared` variables are common to all processors but `private` variables can be different for each processor. The `omp for` directive informs the compiler that a parallel `for` loop is to be undertaken and `schedule(static)` splits the `for` loop into groups of roughly equal numbers of consecutive loop indexes - one group is allocated to each processor. Hence, in this example, each processor works on a small section of the 1D hydrodynamic sweep of a particular row or column.

Since there are overheads involved in splitting and rejoining many threads, it would be preferable to parallelize the outer loop of the integration in `FS2::UpdateGrid`. This would entail passing a `Sweep` object into shared memory, but this is not allowed since the `Sweep` class does not have a default constructor (one with no arguments). However, timing data below shows that parallelization of the inner loop produces considerable efficiency on its own.

The `Sweep` class also contains parallel implementations of the calculation of the bias term ϵ and the slope vectors δQ. The integration of source terms in `Source::Integrate` is also parallelized, where the iteration space is here divided up in two dimensions.

The effect of parallelization on timing

On a single Oscar processor, a typical run of one of the simulations detailed in Chapter 5 takes 3 hrs 45 min. Depending on whether a pure hydrodynamic or a combined hydro plus N-body

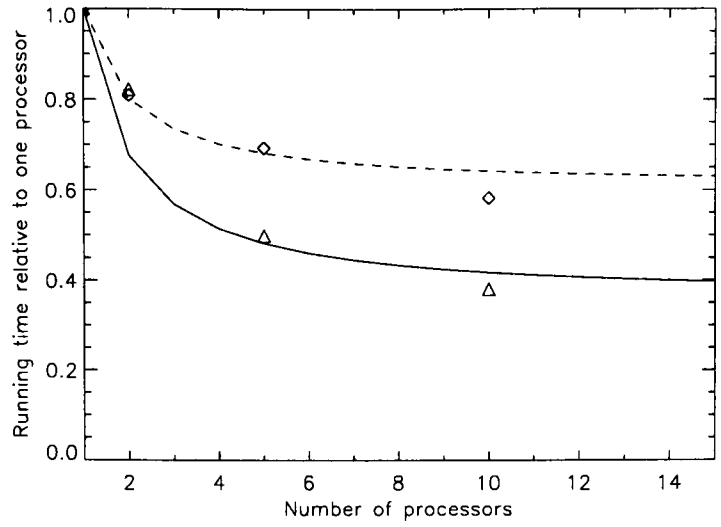


Figure 3.9: Performance of DUAL with number of processors. The solid line shows the expected timings for a purely hydrodynamic simulation and the dashed line for a combined run, both based on Equation 3.6. The triangles and diamonds are the actual run times.

simulation is run, the time spent in major function groups breaks down as follows:

		Just hydro	Combined
N-body	functions controlling the N-body solver	–	37.7%
`Sweep`	the 1D hydrodynamic sweep	56.0%	33.4%
Maths.	mathematical library functions	29.5%	16.0%
`Source`	the calculations of source terms	8.6%	6.2%
`Grid, Node`	`Grid` and `Node` functions	3.7%	1.2%
Misc.	many little used functions	2.2%	5.5%

Since the functions in the `Sweep` and `Source` groups are parallelizable, the time spent within them should scale roughly as the inverse of the number of processors used, n_p. Hence the time to run a typical hydrodynamic simulation t_H or a combined hydrodynamic and N-body simulation t_C should depend on the number of processors as

$$t_H = (0.646/n_p + 0.354)\, t_1, \tag{3.6}$$

$$t_C = (0.396/n_p + 0.604)\, t_1,$$

where t_1 is the time taken using just one processor. Figure 3.9 shows the expected timings and the actual timings as a function of processor number in both cases. When using many processors, the actual timings are faster than the predictions since some of the mathematical library functions are called during the parallelized time.

Chapter 4

Hydrodynamic simulations of the bulge of M31

4.1 Observational motivation

The Andromeda Galaxy (also known as M31 or NGC 224) is the nearest large spiral galaxy to our own and the dominant member of the Local Group. Like our own Galaxy, Andromeda consists of a compact nucleus, a warped exponential disk exhibiting spiral structure, an $R^{1/4}$ de Vaucoulers bulge consisting of older, metal rich stars, a spherical distribution of globular clusters and a surrounding halo of dark matter. Basic observational data is summarized in Table 4.1, a brief review of M31 can be found in Walterbos (2001) and more extensive ones in van den Bergh (2000) and Hodge (1992).

Due to its similarity and proximity to our own Galaxy, and the fact that L^* galaxies like Andromeda and the Milky Way make up the bulk of the luminosity budget of the universe, it is important to understand Andromeda in as much detail as possible. Since it is a mere 690 kpc away, it is possible to obtain high resolution observations, but this can also prove to be a disadvantage as large numbers of observations are necessary to cover the entire galaxy. Its large inclination angle (77°) means that although it is difficult to determine positional data on its apparent minor axis, more information is available about its kinematics.

The fact that Andromeda exhibits a dark halo can be deduced from its rotation curve. It rises to 225 km s^{-1} by a radius of 400 pc, reaches 280 km s^{-1} at a radius of 10 kpc and settles back to 225 km s^{-1} out to the limits of observations at about 30 kpc, well beyond the limits of the optical disk. It is not known how massive the halo is in the central regions of M31, but this Chapter aims to make a significant contribution towards finding out.

As mentioned in Chapter 1, there are a number of observations which indicate that the bulge of M31 is not spherical. The isophotal twisting reported by Lindblad (1956) can be clearly seen in Figure 4.1, reproduced from Hodge and Kennicutt (1982). Lindblad finds that the position angle

Table 4.1: Observed parameters of M31

Hubble type♮	Sb	
Position (1950)*	α_0	$00^h40^m00^s.28$
	δ_0	$40°59'42''.9$
Inclination angle†	i	77°
Position angle†	θ	38°
Distance†	D	690 kpc ($1' = 200$ pc on line of nodes)
Systemic velocity†	$v_{\rm sys}$	$-315\,{\rm km\ s^{-1}}$
Disk B band scale length*	R_d	5.8 ± 0.3 kpc
Disk B band central surface brightness*	I_d	21.6 mag arcsec^{-2}
Total B band absorption‡	A_B	1.3 mag
Bulge effective radius*	R_e	2.0 kpc
Bulge B band surface brightness*	I_e	22.2 mag arcsec^{-2}
Axis ratio of bulge isophotes*	β	1.54
Bulge to disk isophote angle**	ψ	10°
Max. semi-major axis of bulge isophotes§	$R_{\rm max}$	$11'.4$

♮ Sandage and Bedke (1994)
* Loinard et al. (1999)
† Hodge (1992)
* Walterbos and Kennicutt (1988)
‡ Burstein and Heiles (1982) and van Genderen (1973)
** Lindblad (1956)
§ Stark (1977)

of the central bulge region is twisted with respect to the majority of the disk isophotes by an angle $\psi = 10°$. He shows that this is consistent with a central triaxial bulge having axis lengths in the ratio $a : b : c = 2.7 : 1.5 : 1.0$ (my final model has axis lengths in the ratio $a : b : c = 2.8 : 1.6 : 1.0$).

The position velocity diagram in Figure 4.2 is a combination of the CO survey of Loinard et al. (1995) and the H I survey of Brinks and Shane (1984). These show three separate components: the s-shape of a differentially rotating disk, a shallow, almost straight line attributed to the warp in the outer part of the disk and a high velocity component in the inner 5 kpc which shows signatures of non-circular motion.

Finally, Ciardullo et al. (1988) observe a face-on spiral structure in the ionized gas of M31's bulge, contrasting with the highly inclined spiral of the disk. Stark and Binney (1994) suggest that this spiral is caused by the transition from x_1 to x_2 orbits in a barred potential.

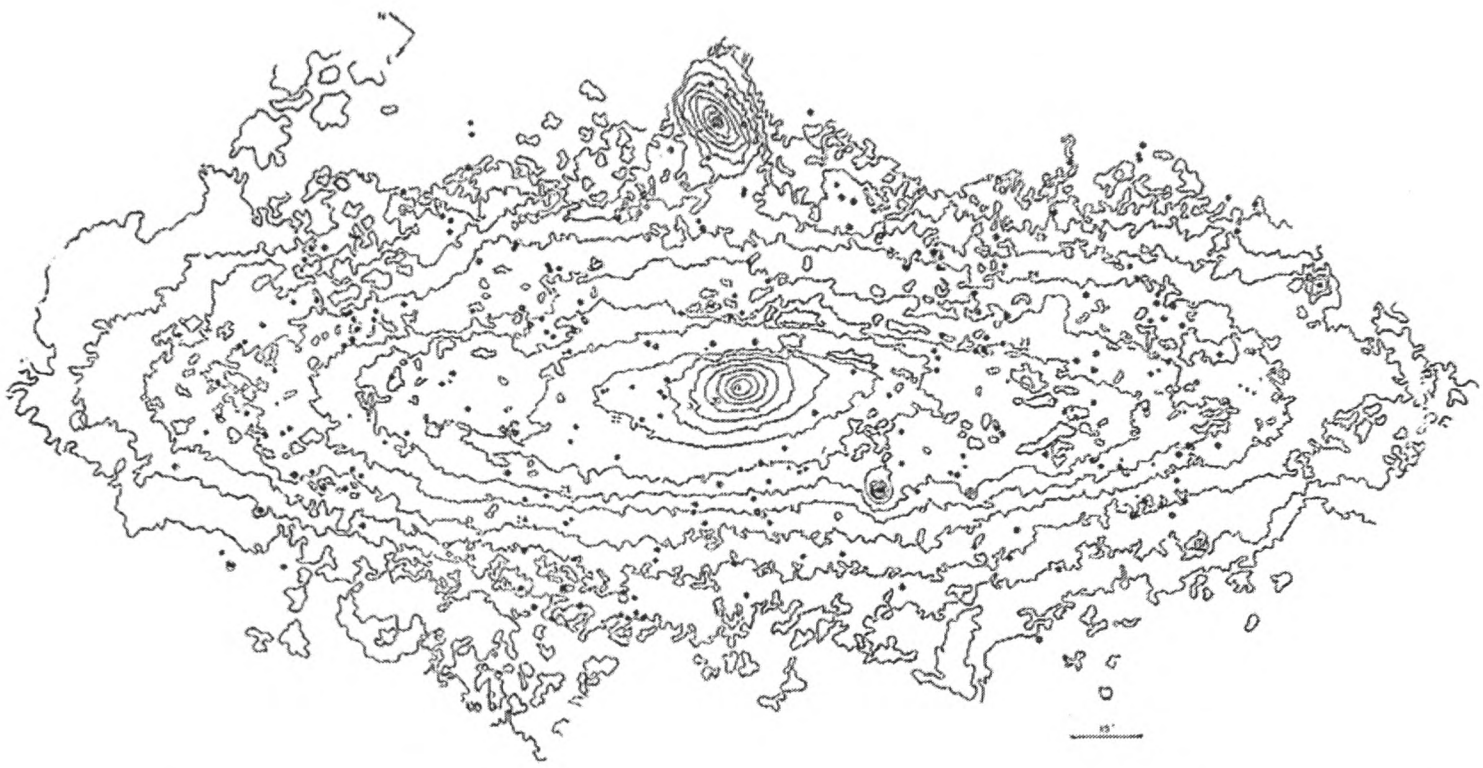


Figure 4.1: Isophotal map for M31 reproduced from Hodge and Kennicutt (1982) and showing the twisting of the central isophotes. Isophotes are given for $B = 19.0$ to 25.5 in half magnitude steps. The bar in the bottom right hand corner corresponds to 10′.

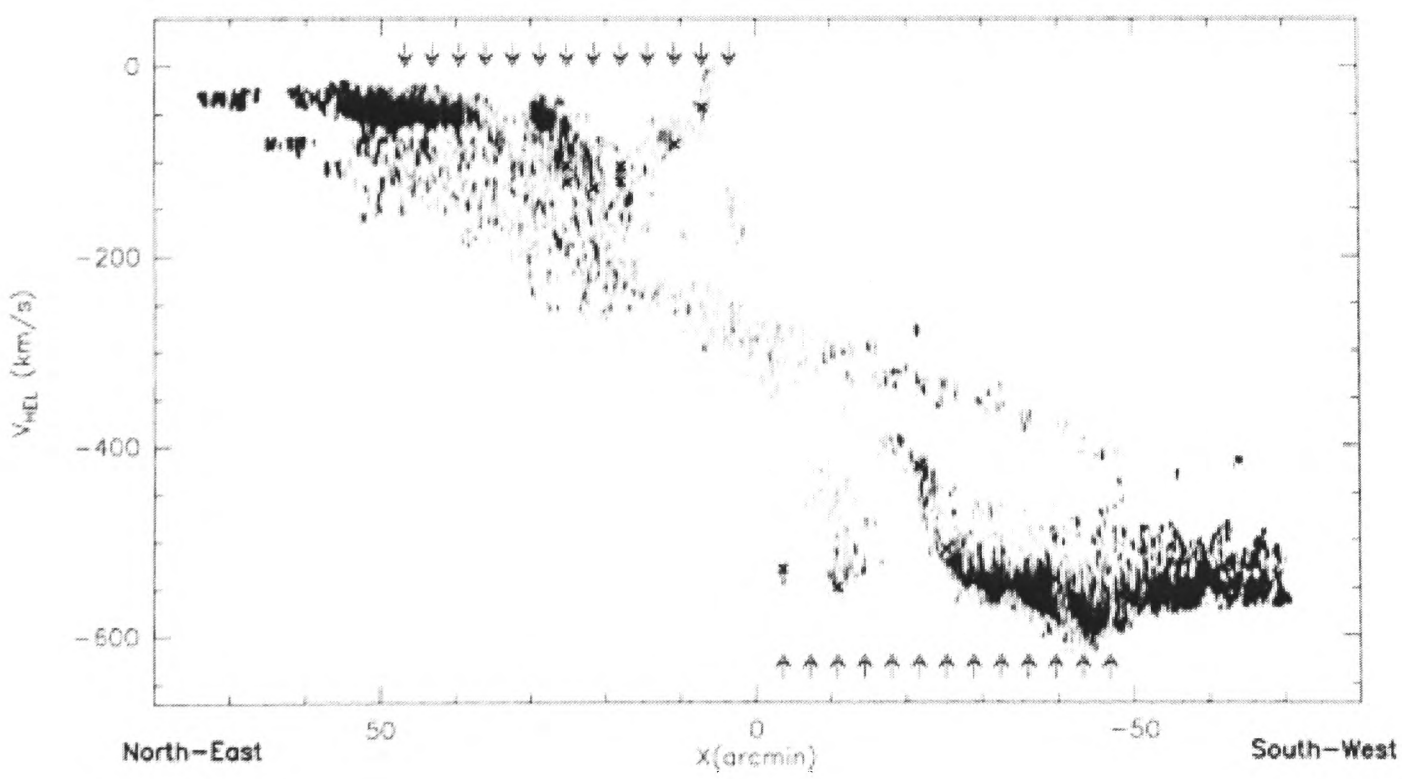


Figure 4.2: Position-velocity diagram of H I(greyscale) and CO (crosses) along the line of nodes of the disk of M31. Reproduced from Loinard et al. (1995).

This chapter describes an investigation of the central part of M31. In particular, a model of the bulge is detailed that improves upon Stark (1977) and Stark and Binney (1994) and uses the hydrodynamics part of the DUAL code described in Chapter 3. The most significant advances on previous models are the use of 2D, isothermal hydrodynamics in modelling the interstellar medium, and the use of the surface brightness profile of Walterbos and Kennicutt (1988) to derive the stellar potential of the bulge. The models of the bulge that are described below are characterized by three parameters: the orientation angle ϕ, the B band mass-to-light ratio Υ_B, and the rotation ratio $\mathcal{R}$. All of these will be described in more detail later.

In Section 4.2, I describe the mass distribution used to model M31 and the initial conditions of the hydrodynamic simulations. Section 4.3 discusses the results of these simulations and Section 4.4 makes comparisons with observations and describes the best fitting model. Section 4.5 comments on the constraints that the simulations place on the existence of a central dark halo and the implications for cold dark matter theories. Section 4.6 concludes.

4.2 Modelling M31

The model of M31 has just three components: a triaxial bulge, an axisymmetric component combining the axisymmetric disk and spherical dark halo, and the gaseous interstellar medium. The bulge and axisymmetric parts form the background potential that acts on the interstellar medium. The ISM is represented by a non self gravitating, isothermal fluid. This is few enough parameters to allow a fairly comprehensive investigation of the parameter space.

Given the possibility of ambiguity, lengths which are observable on the sky will be quoted in arcmins whereas lengths in any other frame will be quoted in pc.

4.2.1 Triaxial bulge

The aim here is to derive the 2D potential generated by a triaxial mass distribution which also gives the surface brightness profile observed in the bulge of M31. All of the values of the observationally determined parameters used in the model are detailed in Table 4.1.

The B band surface brightness profile of the bulge of M31 follows the $R^{1/4}$ law of de Vau-

couleurs (1958):

$$I(s) = I_e \exp(-7.67[(s/R_e)^{\frac{1}{4}} - 1]), \tag{4.1}$$

where the R_e is the radius of the isophote containing half of the total luminosity, I_e is the surface brightness at R_e and s is the semi-major axis length of any isophote. However, the value of I_e in Table 4.1 must be corrected for absorption in M31 and our own Galaxy. Using the values of $A_B = 0.32$ mag for our Galaxy in the direction of M31 (Burstein and Heiles 1982) and $A_B = 0.98$ mag for the bulge region of M31 (van Genderen 1973), the apparent magnitude of the B band bulge surface brightness becomes $I_e = 22.2 - 1.3 = 20.9$ mag arcsec^{-2}.

It is now necessary to transform I_e^{M} in mag arcsec^{-2} into I_e^{L} in $L_\odot$ pc^{-2}. Using $M_\odot = m_\odot + 5 - 5\log D_\odot$ and taking the absolute B band solar magnitude to be $M_{\odot B} = 5.48$ (Allen 1973), the B band apparent brightness of our Sun is derived to be $m_{\odot\mathrm{B}} = 27.052$. From this,

$$\log I_e^{\mathrm{L}} = -0.4\,(I_e^{\mathrm{M}} - m_{\odot\mathrm{B}}), \tag{4.2}$$

gives $I_e = 289\ L_\odot$ pc^{-2}.

Contopoulos (1956) and Stark (1977) both show that any galaxy in which the luminosity is constant on similar ellipsoids has isophotes which are similar ellipses, all identically oriented. He further shows that the problem of finding the volume brightness distribution given the isophotes has a unique solution assuming that the axes of the bulge are parallel with the disk.

To recover this solution, an Abel-type integral equation has to be solved. The solution is itself an integral, which can be numerically integrated using Romberg's method and then multiplied by Υ_B to give a 3D mass density profile.

There are a number of observationally determined parameters in Table 4.1 that are needed to determine the particular triaxial mass distribution of the bulge of M31. The B band surface brightness profile $I(s)$ has already been described. The geometry of the bulge isophotes is described by their axis ratio β, and the semi-major axis of the outermost isophote R_{max}. Two angles transform the observers coordinate frame to that of the bulge: the inclination angle i and the angle between the major axis of the bulge isophotes and the major axis of the disk isophotes ψ.

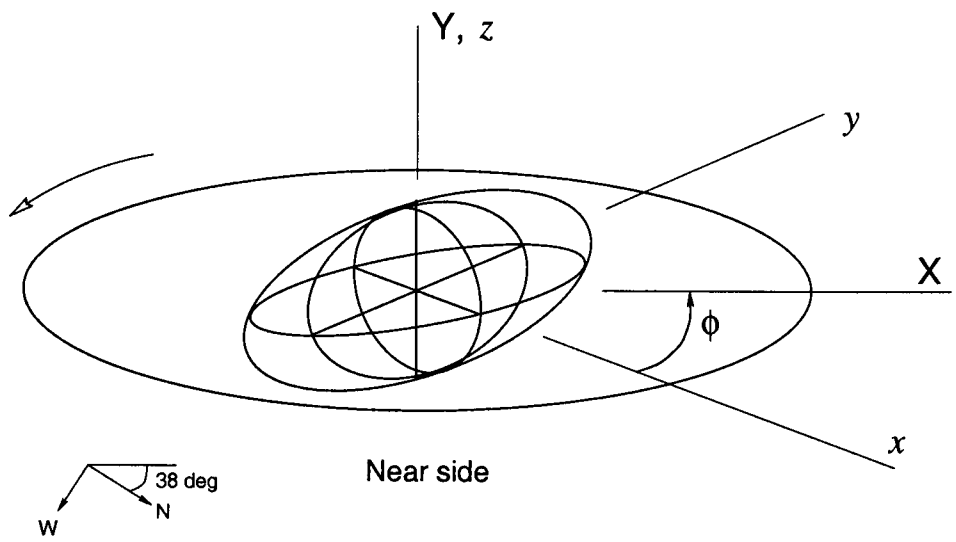


Figure 4.3: Orientation of M31's bulge within the inclined disk. X and Y indicate the coordinates in the observers frame whereas x, y, z are in the frame of the bulge. The angle between the minor axis of the bulge and the line of nodes of the disk is highlighted as ϕ.

One last angle ϕ defines the orientation of the bulge within the disk. ϕ is the angle in the frame of the bulge between the minor axis of the bulge and the disk's line of nodes and it forms a family of mass distributions which are consistent with the observed isophotes. However, it is not known from observations and remains to be determined by comparison with gas kinematics. Figure 4.3 sketches one possible solution.

Table 4.2 shows the effect that changes in ϕ have on other physical variables in the model. Larger values of ϕ are associated with fatter, shorter and more rapidly tumbling bulges.

Table 4.2: Effect of changing ϕ on the size, shape and speed of the bulge

ϕ	a^*	b^*	c^*	a/b	$\Omega^\dagger_{1.0}$	$\Omega^\dagger_{1.2}$	$\Omega^\dagger_{1.4}$
6°	5.00	2.13	0.96	2.35	51.2	44.3	37.4
12°	3.84	2.08	1.21	1.86	62.4	51.0	45.8
18°	3.36	2.02	1.29	1.66	70.2	56.2	49.3
24°	3.10	1.95	1.33	1.58	76.3	60.4	52.2
30°	2.92	1.88	1.36	1.55	81.3	63.8	54.7
36°	2.78	1.79	1.38	1.55	85.6	66.8	56.8
42°	2.68	1.69	1.40	1.58	89.6	69.5	58.7
48°	2.58	1.55	1.42	1.66	93.4	72.1	60.5

* semi major axis lengths (kpc)
† bulge pattern speed (km s^{-1} kpc^{-1}) for $\mathcal{R} = 1.0, 1.2$ or 1.4

The inner isophotes, which are twisted with respect to those of the outer disk, are used to

define the spatial extent of the bulge. Hence, the bulge model is truncated at an elliptical radius $R_{\max} = 11'.4$, the length of the semi-major axis of the largest twisted isophote. For the model of maximum likelihood, this corresponds to a bulge semi-major axis of 3.5 kpc.

Since the hydrodynamic calculations are carried out in 2D, only the forces in the plane $z = 0$ are relevant. The force due to the bulge at a point (x, y) is found by splitting the bulge into 300 concentric triaxial shells and summing the contributions of all shells interior to that point. The force due to exterior shells vanishes. From Hunter et al. (1988), the component of the force acting at (x, y) in the x direction due to an ellipsoidal shell of semi-axes a_i, b_i, c_i, where $a_i > b_i > c_i$ and c_i lies perpendicular to the plane of the disk is

$$dF_x = -\frac{2\,\pi\,G\,\rho(a_i)\,b_i\,c_i\,da_i}{\sqrt{(a_i^2+\kappa)(b_i^2+\kappa)(c_i^2+\kappa)}}\frac{\partial\kappa}{\partial x}\,, \tag{4.3}$$

where $\rho(a_i)$ is the tabulated density of the shell with semi-major axis length a_i and κ is the largest positive root of the equation

$$\frac{x^2}{(b_i^2+\kappa)} + \frac{y^2}{(a_i^2+\kappa)} = 1. \tag{4.4}$$

A similar relationship exists for the force in the y direction.

The region inside the central shell is modelled as a homogeneous ellipsoid of density ρ_c. This central region is split into three smaller shells and the volume weighted densities of these smaller shells are summed to give ρ_c.

4.2.2 Axisymmetric component

A simple interpretation of observations of gas velocities in M31 (Brinks and Burton 1984), suggests that the circular speed $v_c(r)$ along the line of nodes of the disk is equal to a constant velocity v_0, at radii greater than R_0. At radii less than R_0, the circular speed falls linearly to zero. The axisymmetric component has a mass distribution contrived to ensure that the model's circular speed behaves thus. To determine v_0 and R_0, a two component straight line is fitted to the CO observations of Loinard et al. (1995). The data are folded about $X = 0$ and the four most central points are excluded from the fit as they fall inside the bulge and will therefore be experiencing

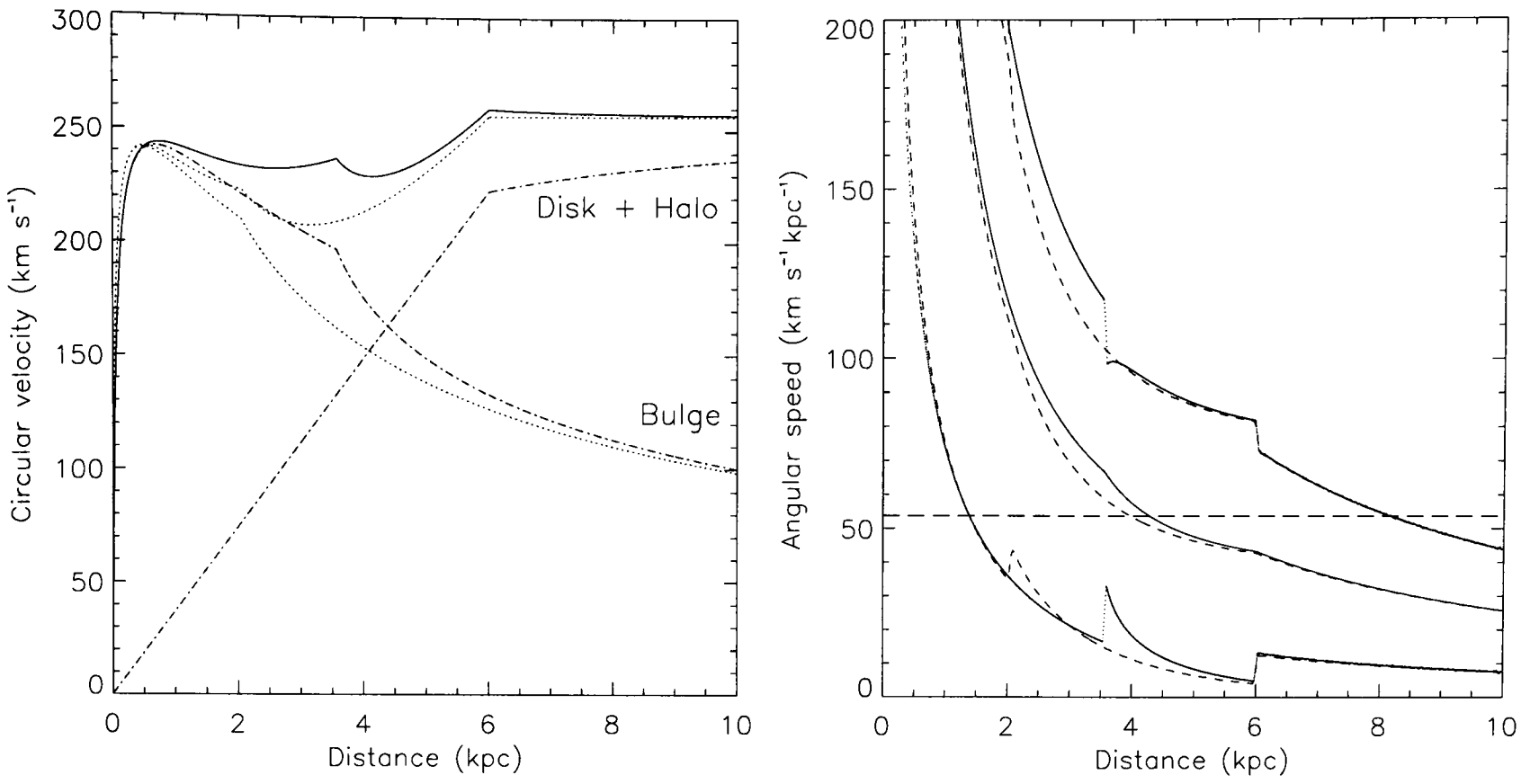


Figure 4.4: The model of maximum likelihood along the major (solid and dot-dashed lines) and minor (dotted lines) axes of the bulge. The left hand panel shows the circular speed. The right hand panel shows Ω and $\Omega \pm \kappa/2$, determined from numerical derivatives of the potential. The long dashed line marks the bulge pattern speed $\Omega_p = 54$ km s^{-1} kpc^{-1}.

severe non-axisymmetric motion. The best fit occurs for $v_0 = 248$ km s^{-1} and $R_0 = 30.0'$. This value of v_0 is deprojected by a factor $\sin i$ for use in the calculation of the forces due to the axisymmetric component.

Since the bulge forces have already been calculated, the forces due to the axisymmetric component can be found by considering that at radii greater than R_0, $v_A^2 + v_B^2 = (v_0/\sin i)^2$, where v_A and v_B are the contributions to the circular speed along the line of nodes from the axisymmetric component and the bulge, respectively. At radii less than R_0, the forces due to the axisymmetric component are assumed to fall linearly to zero.[1]

The left hand panel of Figure 4.4 shows the circular speed curve for the maximum likelihood model (as determined later) on both the major and minor axes and decomposed into contributions from the bulge and the axisymmetric components. The right hand panel shows the angular velocity profile for the same model.

[1]It should be noted that since I have not shown that the forces corresponding to the axisymmetric component can be derived from an everywhere positive mass distribution, it cannot be assumed that such a mass distribution exists.

Section 4.5 considers whether and how the axisymmetric component can be broken up into separate contributions from the disk and dark halo.

4.2.3 Hydrodynamic parameters

As argued in Chapter 2, the cold dense ensemble of clouds within the ISM of M31 can be assumed to behave as an isothermal gas responding to the fixed potential described above. In this case, the sound speed $c_g \sim 10$ km s^{-1} is taken to be the cloud dispersion velocity, and the surface density Σ to be the average density of the surrounding clouds.

The response of the gas is determined using the DUAL code described in Chapter 3. The 2D Cartesian grid represents just half of a two fold rotationally symmetric galaxy. The 80 $\times$ 160 grid cells cover 10 $\times$ 20 kpc so that each grid cell is a square, 125 pc on a side. This is approximately the same resolution as that of the CO observations described later. The bulge is positioned so that its major axis is aligned with the long axis of the grid. The self gravity of the gas is neglected in all simulations. The grid is assumed to be in a frame rotating with the bulge at a pattern speed Ω_p and gas recycling is used. Hence Equations 2.2 and 3.4 apply.

Defining the pattern speed fixes a corotation radius $R_{\rm co}$, where the gravitational and centrifugal forces are equal and, in the rotating frame, gas can remain motionless. This radius occurs at the distance along the major axis at which $\frac{\partial \Phi_{\rm eff}}{\partial y} = 0$, where the effective potential $\Phi_{\rm eff} = \Phi - \frac{1}{2}\Omega_p^2 R^2$. Following Athanassoula (1992b), I set $R_{\rm co} = \mathcal{R}\,a$, where a is the length of the bulge semi-major axis. However, unlike the galaxy models of Athanassoula (1992b), M31 is a real galaxy, modelled with a somewhat arbitrary bulge edge and so it cannot be assumed that the value of $\mathcal{R}$ will be 1.2 ± 0.2 as quoted in Athanassoula (1992b). Hence the rotation ratio $\mathcal{R}$ remains to be determined by comparison with gas kinematics.

Simulations start with the gas at a uniform surface density $\Sigma_0 = 1 M_\odot$ pc^{-2} and on circular orbits at a speed given by $\beta v_A(r)$ with β chosen such that $\beta^2 v_A^2(R_M) = GM(R_M)/R_M$, where $R_M = 10$ kpc and $M(R_M)$ is the mass of the model at R_M. As stated in Chapter 3, the boundary conditions fix the dynamical variables at their initial values on two rows of ghost cells outside the boundary of the grid. The value of all hydrodynamic and computational parameters used in the

Table 4.3: Hydrodynamic and computational parameters

Number of grid cells	$I \times J$	80×160
Gas recycling parameter†	α	0.3 pc^2 $M_\odot^{-1}$ Gyr^{-1}
Initial gas density	Σ_0	1 $M_\odot$ pc^{-2}
Constant mass radius	R_M	10 kpc
Sound speed‡	c_g	10 km s^{-1}
Galaxy radius	R_G	10 kpc
Courant number	C	0.5

† Athanassoula (1992b)
‡ Cowie (1980)

simulations are listed in Table 4.3.

In order to ensure a smooth transition from an axisymmetric to a non axisymmetric potential, the bulge is linearly introduced over half a rotation period and the model is then evolved to a quasi steady state over three rotation periods, corresponding to about 0.3 Gyr. A comparison of runs at three and four rotation periods shows very little change in the velocity fields over much of the galaxy. However, in the region in which gas orbits around the Lagrangian points, significant differences are found. Oscillations with amplitudes $\sim$ 10 km s^{-1} occur in the total gas velocity with a period of 70 Myr. These oscillations persist until at least ten rotation periods.

4.3 Simulation results

4.3.1 Selection of models

In all, 87 simulations were run within the three dimensional parameter space (ϕ, Υ_B, $\mathcal{R}$) where $6° < \phi < 36°$, $4.5 < \Upsilon_B < 8.5$ and $1.0 < \mathcal{R} < 1.4$. Appendix B contains parameters of all of the runs. The model which most closely fits the observations is detailed in Table 4.4 and the method used to determine this is described in Section 4.4.

Table 4.4: The bulge model of maximum likelihood

Phi angle	ϕ	15°
Bulge mass-to-light ratio	Υ_B	6.5
Rotation ratio	$\mathcal{R}$	1.2
Semi-major axis	a	3.5 kpc
Bulge mass	M_{bulge}	$2.3 \times 10^{10}\, M_\odot$
Pattern speed	Ω_p	53.7 km s^{-1} kpc^{-1}

4.3.2 Gas flow in selected models

Section 2.2 described the two main families of periodic orbits that exist in the plane $z = 0$ of triaxial or barred potentials. x_1 orbits exist at all radii and are elongated along the bulge major axis and x_2 orbits exist inside the Inner Lindblad Resonance (ILR) and are oriented perpendicular to the major axis. These gas flows can be seen in the simulations of M31. Figure 4.5 shows the velocity field for the model of maximum likelihood: gas near the outline of the bulge is moving in orbits elongated along the major axis whilst, near the centre of the bulge, the gas flows along orbits elongated along the minor axis.

When gas falls inwards and passes the ILR, it shifts from x_1 to the lower energy x_2 orbits. In doing so, it creates the 'spray' region seen as the diagonal gas flows just above and to the right, and below and to the left, of centre in Figure 4.5. Since this area contains transient gas which is not on any periodic orbits, it is an area of very low density and corresponds to the low density regions in Figure 4.5.

As the gas travels through the 'spray' region, it slows down and climbs out of the bulge's potential well. At the end of the 'spray' region, the gas reaches its minimum kinetic energy and abruptly changes direction, following the gas it has encountered back down into the potential well. This is seen in the left hand panel of Figure 4.5 at the peak of the negative of the velocity divergence and is a signature of shocked gas. In the right hand panel, this corresponds to the high density lanes of gas, formed where the shock causes a pile up of gas. Athanassoula (1992b) shows

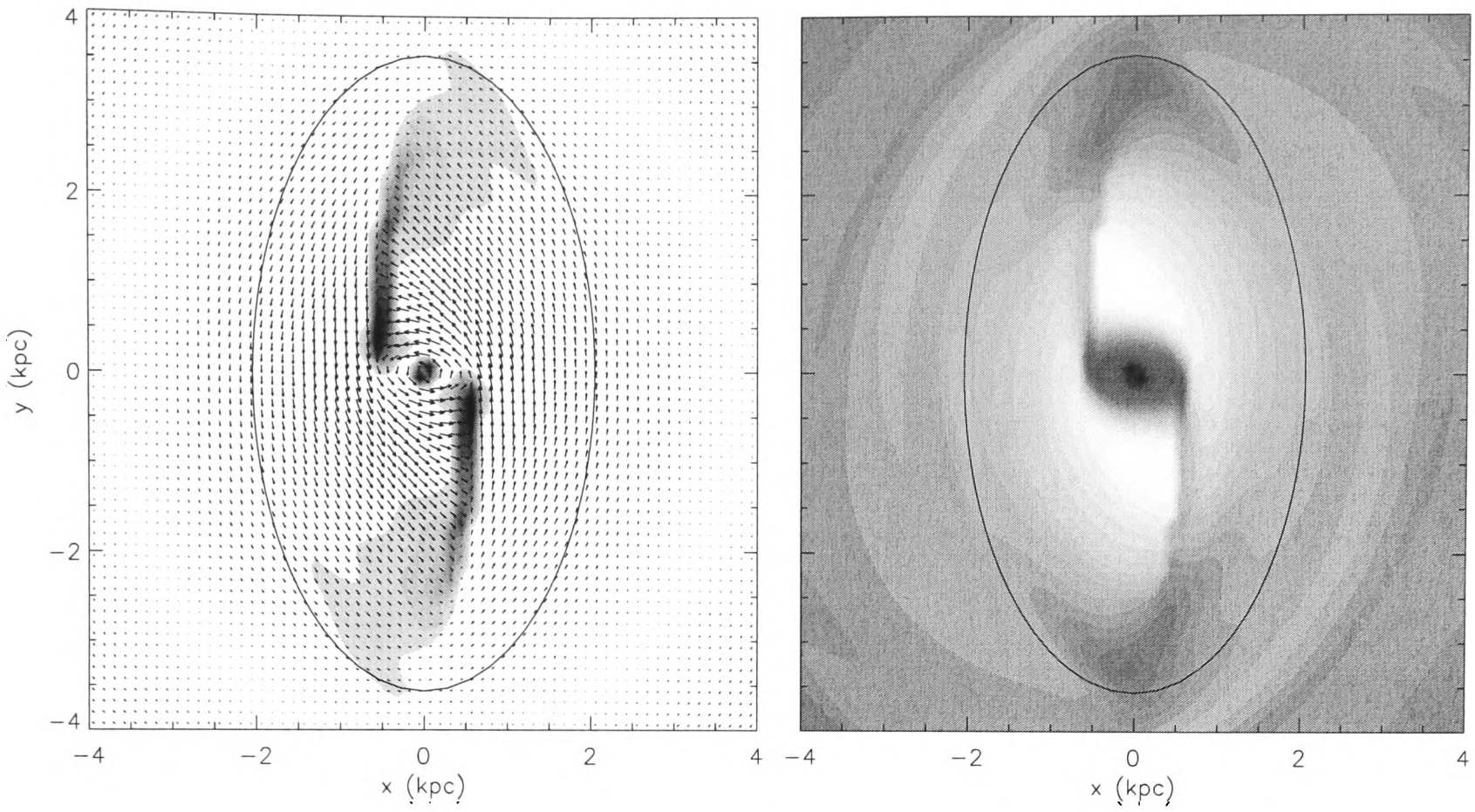


Figure 4.5: The central 4 kpc of the model of maximum likelihood in the frame rotating with the bulge. The left hand panel shows the velocity field in which minus the divergence of the velocity is shown as a greyscale, highlighting the shocks. The right hand panel shows the logarithmic density as a greyscale suggesting the existence of dust lanes. The solid ellipse marks the bulge outline.

that these shocks are intimately linked to the dust lanes which appear in many barred galaxies.

Figure 4.6 shows the velocity field for a selection of four different models in the frame rotating with the bulge. Each model has $\phi = 22°$ implying that the semi-major axis of the bulge $a = 3.2$ kpc for every model. The solid ellipse in each panel marks the extent of the bulge and the greyscale is minus the divergence of the velocity, highlighting the shocks.

The top two panels are slower bulges ($\mathcal{R} = 1.4$ and $\Omega_p = 50$ and 52 km s^{-1} kpc^{-1}) whereas the lower two are faster ($\mathcal{R} = 1.0$ and $\Omega_p = 70$ and 77 km s^{-1} kpc^{-1}). The left hand panels describe lighter bulges ($\Upsilon_B = 5.5$) and the two on the right are heavier ($\Upsilon_B = 7.5$).

Comparing the top two panels with the bottom two shows the effect of a faster bulge on the shock morphology. A faster bulge has a smaller corotation radius and ILR where gas transfers between x_1 and x_2 orbits. Since the transfer of gas between families of periodic orbits is the primary cause of the shocks seen in Figure 4.6, shocks in fast bulges occur closer to the centre, irrespective of whether the bulge is heavy or light. Faster bulges also have pronounced secondary

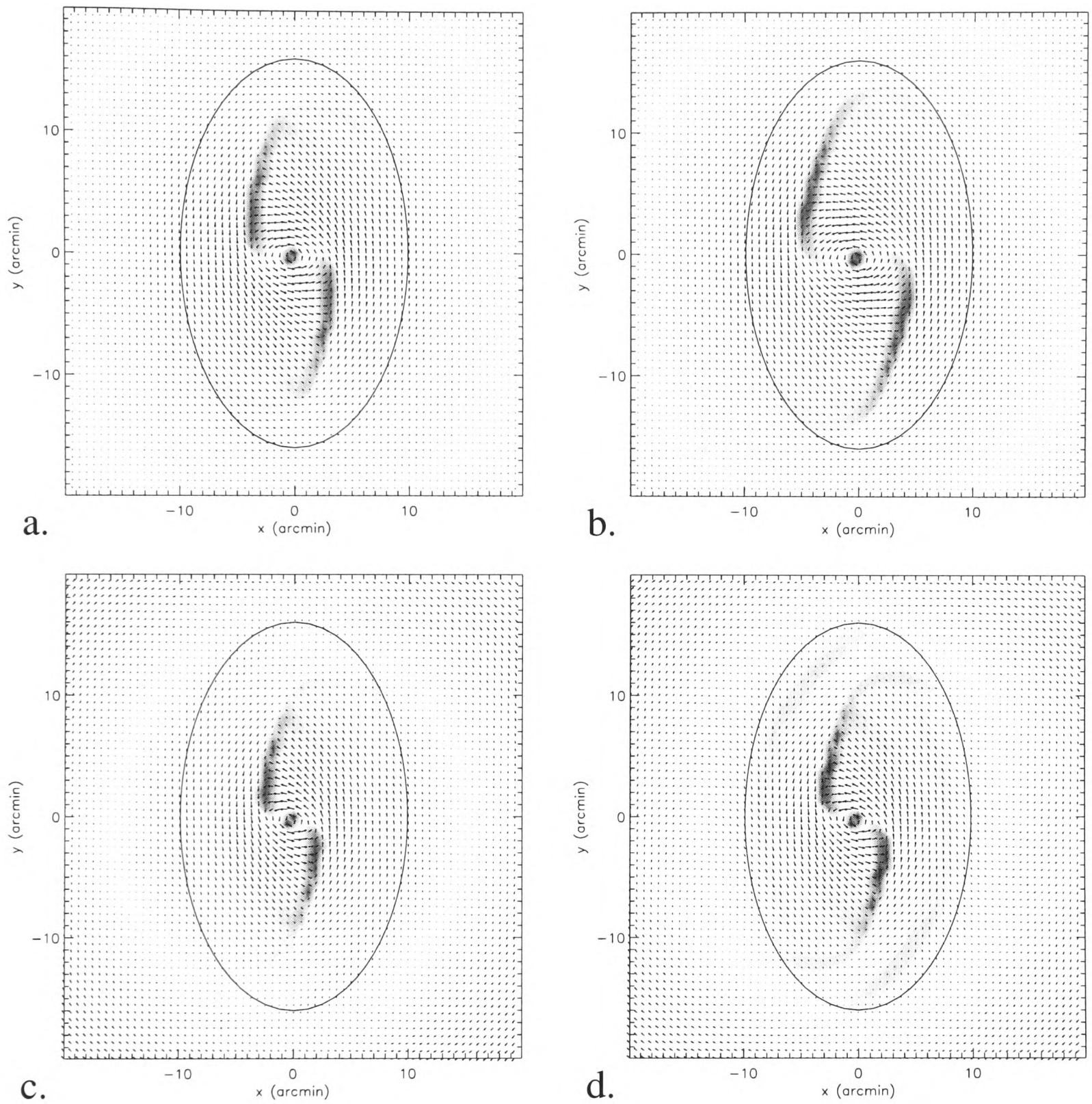


Figure 4.6: Velocity fields for four different runs in the frame rotating with the bulge. The models in panels (a) and (b) are slower bulges with $\mathcal{R} = 1.4$ whereas those in panels (c) and (d) are faster with $\mathcal{R} = 1.0$. Panels (a) and (c) are lighter bulges with $\Upsilon_B = 5.5$ and those in panels (b) and (d) are heavier with $\Upsilon_B = 7.5$. The pattern speeds for the four models are $\Omega_p = 50$, 52, 70 and 77 km s^{-1} kpc^{-1} for panels (a), (b), (c) and (d) respectively. In all four models $\phi = 22°$, and the semi-major axis of the bulge $a = 3.2$ kpc. The solid ellipse in each panel marks the bulge and minus the divergence of the velocity is shown as a greyscale, highlighting the shocks. Every computational cell is shown.

shocks emerging near the end of the major axis. Comparing right and left panels shows that lighter bulges have weaker shocks, close to the major axis and slightly straighter than heavier bulges.

4.4 Comparison with observations

4.4.1 Observations along the line of nodes of the disk

Molecular gas is likely to be confined to the plane of the disk and follow the periodic orbits described above. Since CO is a good tracer of molecular gas in Andromeda (Neininger et al. 1998), CO data is used to judge the success of the hydrodynamic simulations.

Although there have been recent CO maps made of most of the Andromeda galaxy (Loinard et al. 1999; Dame et al. 1993), there is little CO data available in the central region. The reason for this deficiency could be either that the CO is very cold and difficult to detect in the inner disk (Loinard and Allen 1998) or simply not there (Melchior et al. 2000).

However, high resolution data (Loinard et al. 1995) does exist along the line of nodes of the disk. This data was taken using the 30-m IRAM millimeter radio telescope in 1993. 26 positions along the apparent major axis of M31 were surveyed and CO was detected in 16 of them with a resolution of $35'' \times 2.6\,\mathrm{km\ s^{-1}}$. All 16 data points are used in the comparison with the models.

Each model was rotated by ϕ and inclined by i to match the viewing position on the sky. A cut was then taken along the line of nodes of the disk, involving an interpolation from off-axis positions, and a comparison made with the CO data. Figure 4.7 shows this comparison for the best fit model.

The model fits the data well except for the point at X = -21.6″, v_{los} = -422.4 km s^{-1} which cannot be fit. This is not unexpected since the model only takes 2D gas dynamics into account. There are many dynamical and astrophysical phenomena that could impart extra radial or vertical motions to a molecular cloud and leave it with the position and velocity of the anomalous point. These phenomena include supernovae ejecting shells of gas, vertical instabilities in the disk or the existence of a CO disk with a significant scale height.

To make a quantitative statement about how good a fit a particular model is and which model gives the best fit to the data, a likelihood function is derived, where the errors in the position

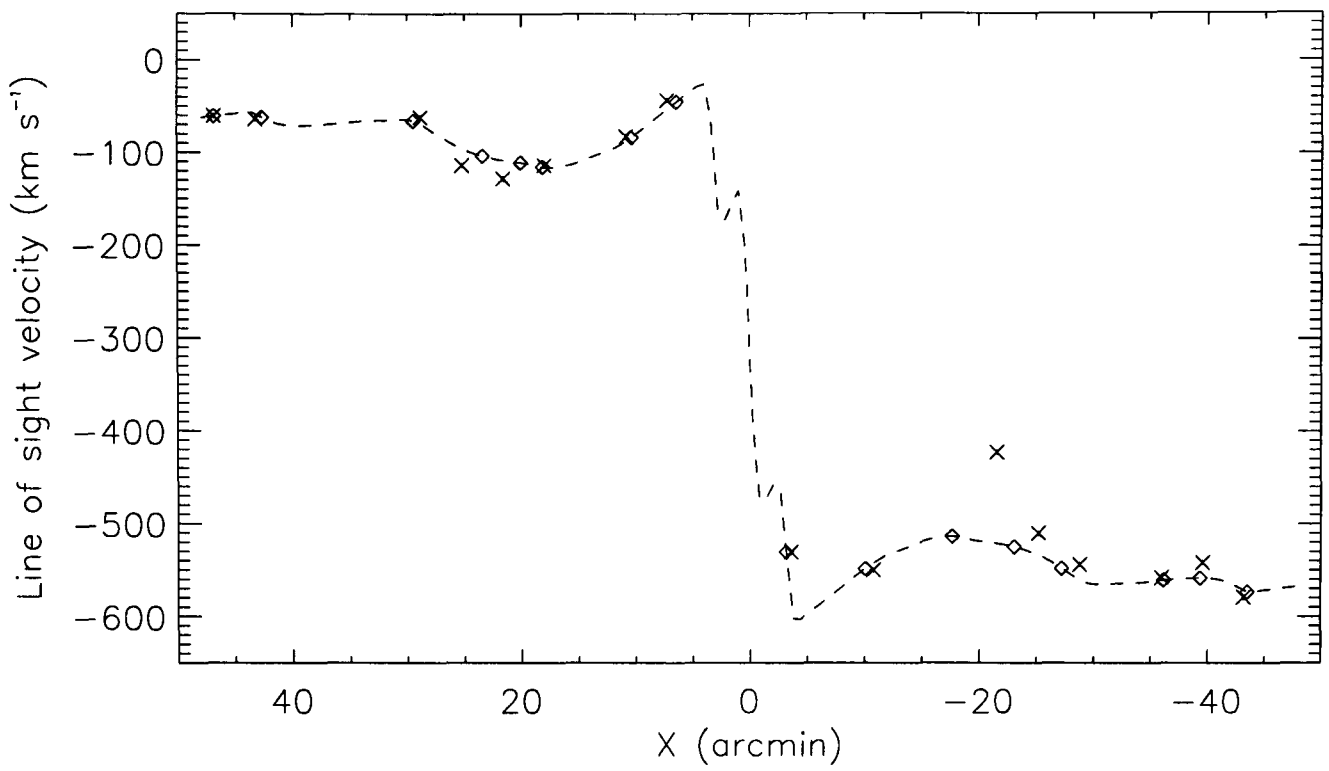


Figure 4.7: The line of sight gas velocity on the line of nodes of the disk. The dashed line is data from the model of maximum likelihood after transforming to the observing frame. Crosses are CO detections from Loinard et al. (1995). Diamonds are the closest points on the dashed line to each data point in terms of data error bars. The coordinate system is that of Baade and Arp (1964).

and velocity of the CO data are assumed to be Gaussian. Given that there are errors in both X and $v_{\rm los}$, the difference between the data and the model is found by comparing each datum with the point in the model closest to it. This takes place within error space, where all distances are divided by the error on X ($35''$) and all velocities by the error on $v_{\rm los}$ (2.6 km s^{-1}).

This procedure gives a value of the likelihood for each model and hence the model of maximum likelihood can be determined. For this model, $\phi = 15°$, $\Upsilon_B = 6.5$ and $\mathcal{R} = 1.2$. Other parameters are detailed in Table 4.4. The value of χ^2/N for this model would be 4.7 or 2.7, with or without the far flung data point respectively. Given the various assumptions that have gone into the simulations, it is unlikely that values much closer to $\chi^2/N = 1$ could be attained.

The 50%, 75%, 90% and 99% confidence surfaces are now formed in the 3D parameter space by the points at which the likelihood has fallen by 1.2, 2.1, 3.1 and 5.7 from its maximum value. The values for these falls in likelihood are half of those given by the χ^2 distribution, with the number of degrees of freedom set to three. Cuts through this 3D parameter space are shown in Figures 4.8 and 4.9 for $\mathcal{R} = 1.0$, 1.2 and 1.4. They show that all of the models with $\mathcal{R} = 1.4$

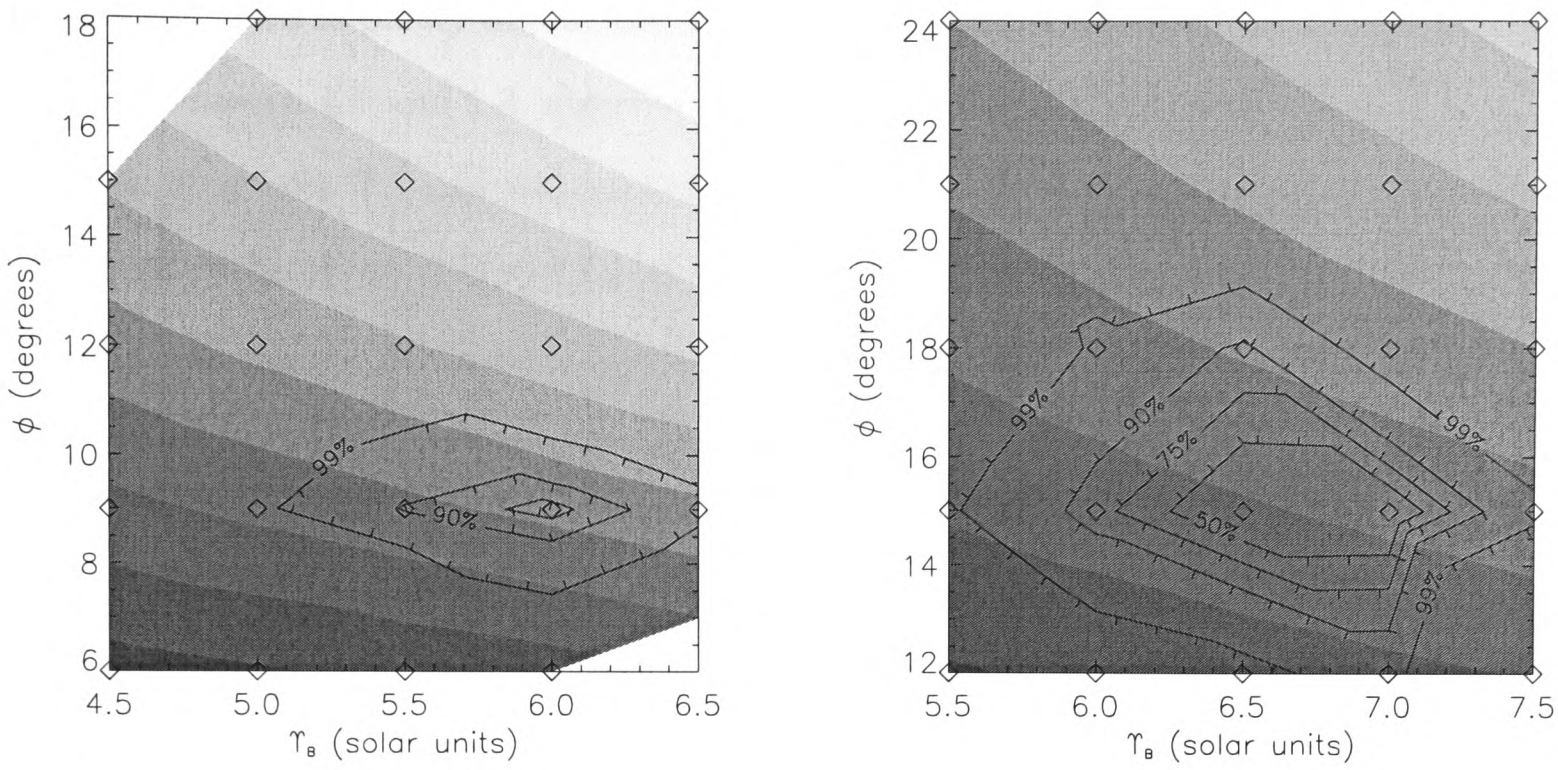


Figure 4.8: Confidence regions for $\mathcal{R} = 1.0$ and 1.2. The anomalous point has been removed from the data set. Diamonds mark the position of each model. Contours show the 50%, 75%, 90% and 99% confidence levels, whilst the greyscale represents the bulge pattern speed Ω_p of each model (the scale is shown Figure 4.9).

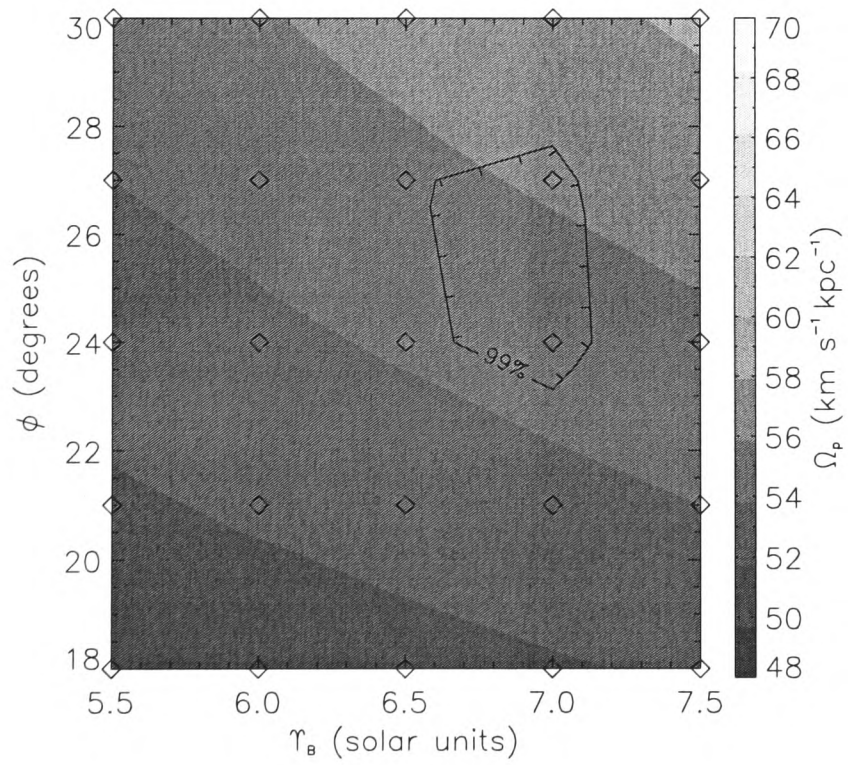


Figure 4.9: Confidence regions for $\mathcal{R} = 1.4$. The anomalous point has been removed from the data set. Diamonds mark the position of each model. The contour shows the 99% confidence level, whilst the greyscale represents the bulge pattern speed Ω_p of each model.

are excluded with 90% confidence. Only two of the models with $\mathcal{R} = 1.0$ are included in the 90% confidence region. This is partly due to the undersampling in $\mathcal{R}$, but it is clear that the favoured value of $\mathcal{R}$ lies between 1.0 and 1.3, showing that M31's bulge is a fast rotator.

The values of ϕ and Υ_B can be constrained separately by considering the χ^2 distribution with one degree of freedom. This gives, at the 2σ or 95% confidence level, $9° < \phi < 18°$ and $5.9 < \Upsilon_B < 7.2$. Similarly, the pattern speed is $53 < \Omega_p < 57\,\mathrm{km\ s^{-1}\ kpc^{-1}}$ with 95% confidence.

4.4.2 Observations away from the line of nodes of the disk

The values of χ^2/N derived above are, unfortunately, not very robust. It is possible to vary χ^2/N by about two by taking the result of a particular model at four rather than three bar rotations. This is enough to turn a model which fits the data very well into one which does not. Therefore it becomes much more difficult to determine whether a particular model is a good representation of M31 or not and how well the above results can be trusted. A much stronger case could be made for the best fit model if it were consistent with observations of CO velocities away from the line of nodes of the disk. Moreover, it is important to utilize these velocities since dynamical models of the bulge of M31 have not been constrained by three dimensional (X, Y, V) observations before.

The data used to make this comparison are CO observations of Loinard et al. (1999).[2] This survey was made using the FCRAO radio telescope and covers the entire southern half of M31. The size of the data cube is $(146 \times 75 \times 36)$ in (X, Y, V) and the resolution is 3.25 km s^{-1} in velocity and $50''$ in X and Y. In all of the comparisons below, the model has been interpolated to fit the positions of the observations.

Since the observing beam is $50''$ wide, observations taken at any particular value of Y actually pick up all the emission between $Y - 25''$ and $Y + 25''$. Hence, the position-velocity diagrams derived from the data cube all have multiple velocities at each position X.

Figure 4.10 shows (X, V) observational data integrated between $Y - 25''$ and $Y + 25''$ for $Y = \pm y \times 50''$ and $y = 1, 2, 3$. On top of this are the best fit model's predictions at $Y \pm 25''$ which should, if the model is correct, enclose the observations on every plot. Although the general

[2]The data cube of the CO survey of Loinard et al. (1999) can be found in FITS format at `http://www.astrosmo.unam.mx/~l.loinard/fcrao.html`

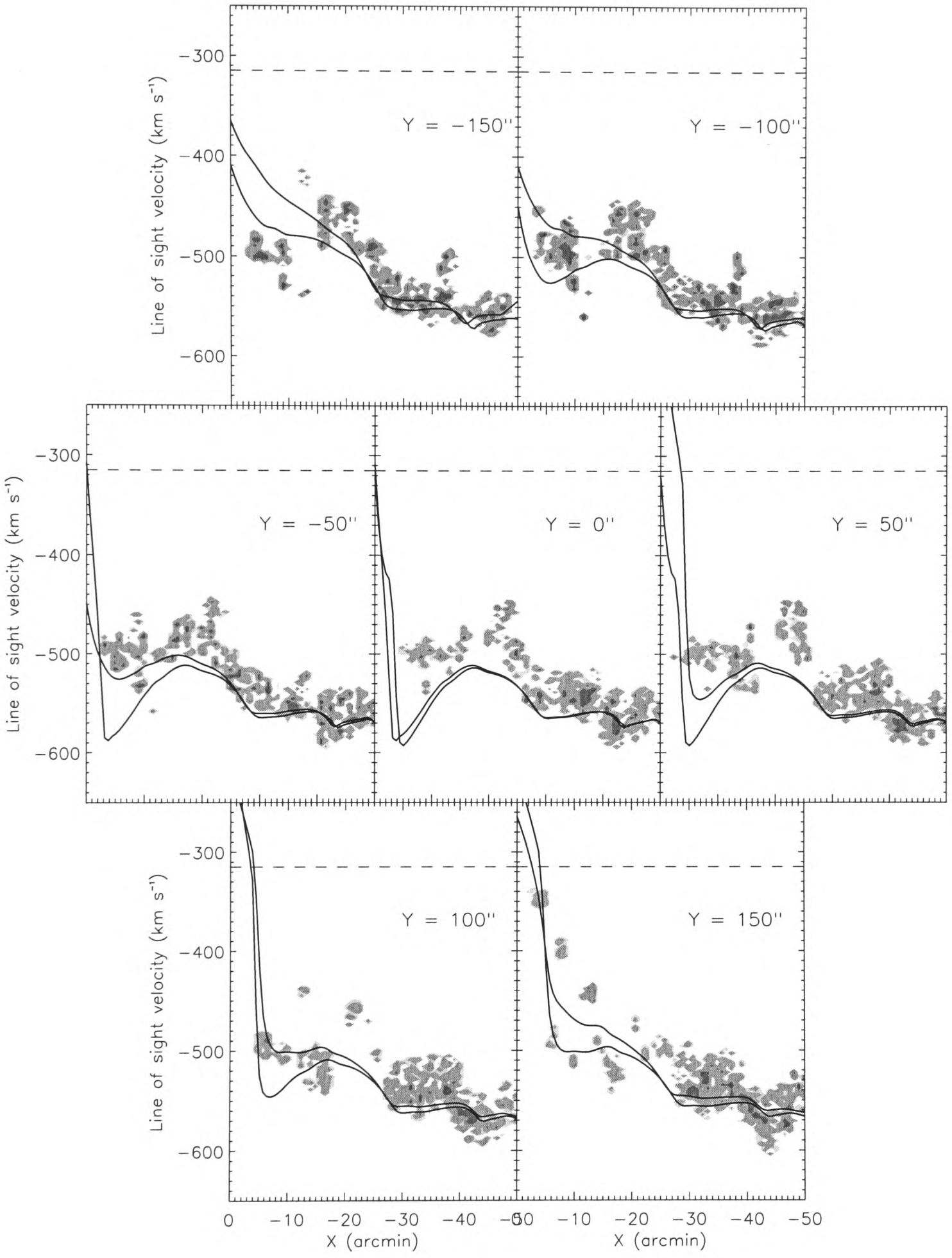


Figure 4.10: CO velocities parallel to the line of nodes of the disk from Loinard et al. (1999). Observations were integrated between $Y - 25''$ and $Y + 25''$ where Y is indicated on each subplot. Solid lines indicate the best fit model predictions for $Y \pm 25''$. The dashed line indicates the systemic velocity.

trends of the predictions and the observations are similar, the model data enclose a narrower range of velocities than the observations and are further away from the systemic velocity than the observations. Similar discrepancies are found between the observations and models with different parameters.

The same pattern of velocities are described by Schaap et al. (2000) for the spiral galaxy NGC 2403. Figure 1 of Schaap et al. (2000) includes the position-velocity map for the galaxy in which a similar 'beard' structure is evident. The arguments presented in that paper conclude that the beard is either caused by a thick gas layer with a half width at half maximum (HWHM) intensity of 500 pc or a lower density layer with HWHM of 1.75 kpc rotating more slowly than the standard thin disk. This conclusion is motivated by galactic fountain models (Spitzer 1990) in which hot gas from supernova explosions and galactic winds rises from the disk into the halo. As it rises, the gravitational attraction towards the centre of the galaxy lessens, the gas moves outwards and, by conservation of angular momentum, its azimuthal velocity decreases. Examples of thick molecular disks have been seen directly in CO in the Milky Way Galaxy (Dame and Thaddeus 1994) and the edge–on spiral galaxy NGC 891 (Garcia-Burillo et al. 1992).

To recover the smaller magnitude and larger range of velocities in the observations, the bulge model is augmented by the kinematics of a thick disk, rotated and inclined to the observer's frame. The vertical density distribution follows Schaap et al. (2000) in taking a Gaussian form

$$\rho(z) = \rho_0 \exp(-z^2/2\,z_0^2),$$
$$\rho_0 = \frac{\Sigma_0}{(2\pi)^{1/2}\,z_0}, \tag{4.5}$$

where ρ_0 and Σ_0 are the density and surface density respectively at $z = 0$, and z_0 is the vertical scale height of the gas. The half width at half intensity is $z_{1/2} = 1.18\,z_0$. The intensity at (X, V) is summed over every point between $Y - 25''$ and $Y + 25''$.

Figure 4.11 compares the observational data with the results from a model with $z_0 = 400$ pc, the lowest value of the scale height that adequately encloses the vast majority of the observations for $X > -50'$. For this model, $z_{1/2} = 472$ pc. This is very large when compared to the thick

disk observed by Dame and Thaddeus (1994) which had a HWHM of between 71 and 133 pc depending on position, but smaller than the thick disk of NGC 891 (Garcia-Burillo et al. 1992), where HWHM is larger than 1 kpc.

I now follow Schaap and drop the implicit assumption that the thick disk is corotating with the $z = 0$ plane by reducing the velocities of the gas where $z \neq 0$. Velocities parallel to the z axis stay fixed at zero. A slower thick disk is implied by the galactic fountain models and was also invoked by Swaters et al. (1997) to explain the H I observations of NGC 891. I adopt the linear velocity distribution used in both papers

$$\mathbf{v}(z) = \mathbf{v}_0(1 - \gamma\, z), \tag{4.6}$$

where $\mathbf{v}_0$ is the velocity in the $z = 0$ plane, γ is the velocity lost per kpc and z is the height.

By slowing the gas down above and below the plane of the disk, the range of velocities at a particular scale height is extended. It can be seen in Figure 4.12 that a similar range of velocities to both the observations and the corotating model is produced if the scale height is just 200 pc and $\gamma = 0.2$ i.e. the gas loses 20% of its speed per kpc away from the plane of the disk. This implies that $z_{1/2} = 236$ pc which is closer to but still larger than the values quoted in Dame and Thaddeus (1994) and much smaller than the value for NGC 891.

Including a thick disk enables the bulge model to match the (X, Y, V) CO data of Loinard et al. (1999) reasonably well. The most noticeable discrepancy is found around $X = -22'$, $V = -450$ km s^{-1} for most values of Y, a position that the original model found difficult to account for even along the line of nodes of the disk. It is plausible that some vertical or turbulent gas motions can account for these remaining discrepancies.

Galactic fountain models may not be justifiable for the centre of M31 since it is not thought to harbour a great deal of star formation. It is possible, however, that the central CO disk is warped or simply thicker than was previously thought. Pringle et al. (2001) contend that much of the ISM in spiral galaxies is molecular, but too cold to be detected in CO. Hence, only the parts of molecular clouds illuminated by newly formed stars would be visible in CO. The CO–bright thin

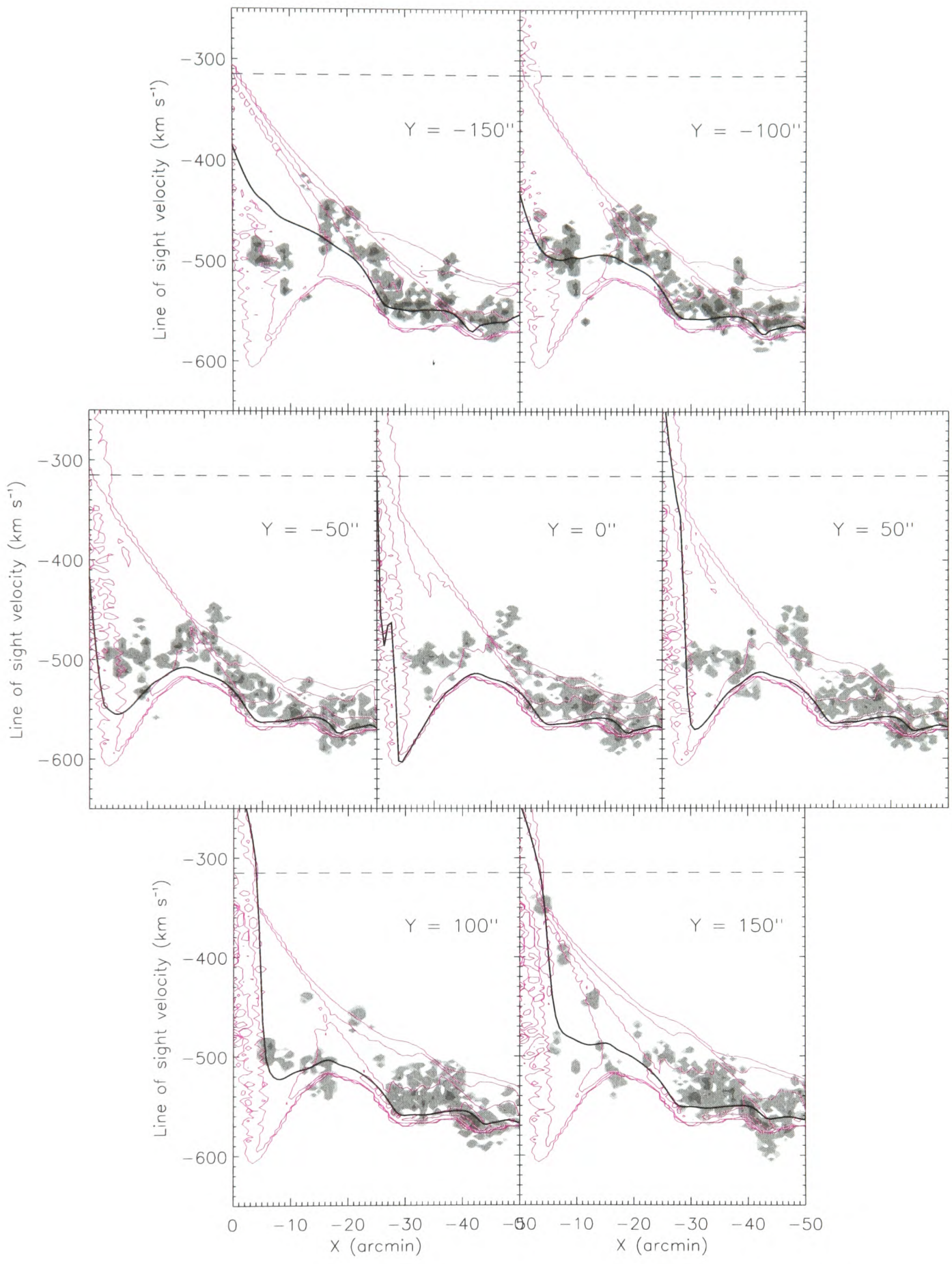


Figure 4.11: Comparison of CO observations and gas velocities from the model of maximum likelihood away from the line of nodes of the disk. The model includes a corotating thick disk of scale height $z_0 = 400$ pc. Observations and model data were integrated between $Y - 25''$ and $Y + 25''$ where Y is indicated on each subplot. Intensity is plotted as a greyscale for observations and contour levels for model data. Contours are given at factors of e, e^3, e^5 and e^7 below the maximum value of each subplot. The dashed line indicates the systemic velocity and the solid line the results at Y'' without the thick disk.

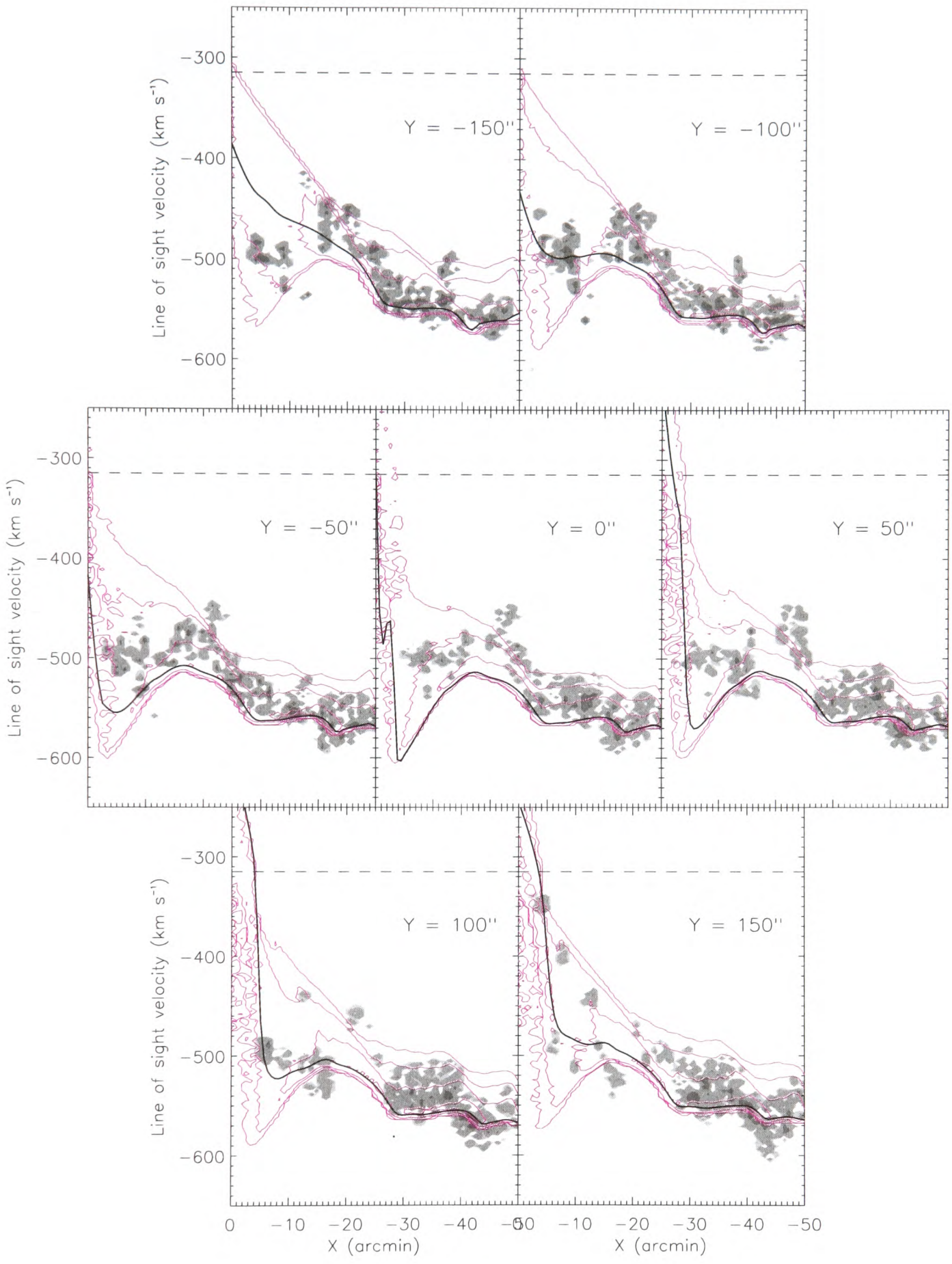


Figure 4.12: Comparison of CO observations and gas velocities from the model of maximum likelihood away from the line of nodes of the disk. The model includes a thick disk of scale height $z_0 = 200$ pc and a loss of velocity of $0.2\,\mathbf{v}_0$ kpc^{-1}. Observations and model data were integrated between $Y - 25''$ and $Y + 25''$ where Y is indicated on each subplot. Intensity is plotted as a greyscale for observations and contour levels for model data. Contours are given at factors of e, e^3, e^5 and e^7 below the maximum value of each subplot. The dashed line indicates the systemic velocity and the solid line the results at Y'' without the thick disk.

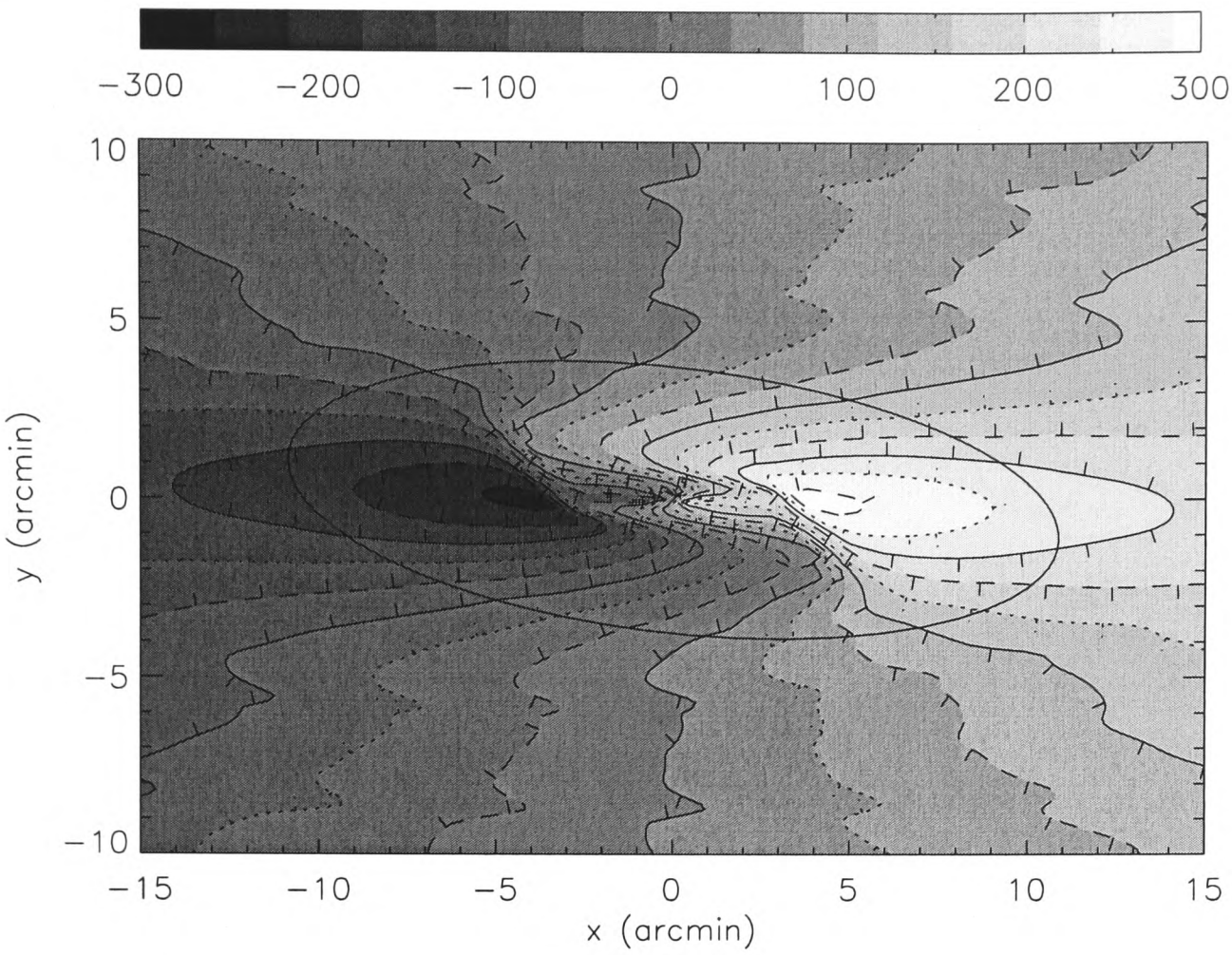


Figure 4.13: Velocity field for the bulge of M31. Obtained by rotating and inclining the velocity field of the model of maximum likelihood to the observer's viewpoint. The contours and greyscale show the line of sight velocity of CO emission. Contours are spaced at intervals of 35 km s^{-1} with the 0 km s^{-1} marked as a solid line. The ellipse outlines the bulge.

disk could then correspond to the thin layer where stars form most actively, whereas the thick disk would correspond to the real total extent of the molecular layer. The existence of large quantities of cold molecular gas in the inner regions of M31 has been proposed by Loinard and Allen (1998).

Martos and Cox (1998) describe numerical simulations which combine magnetic fields with hydrodynamics to create post-shock regions in the galactic disk in which large pressure gradients exist near to the mid–plane, causing the gas layer to expand vertically. The result is a 'hydraulic jump' in which a build up of gas in the vertical direction can be seen with scale heights of hundreds of parsecs. Hence, the multiple velocities at each value of X in Figure 4.10 are probably not only due to the integration over 50″ in Y but also to the motion of CO gas in a thick or warped disk.

4.4.3 Best model dust map

The high density gas appearing in the right hand panel of Figure 4.5 suggests the existence of dust lanes in M31. Using the value for the mean B band absorption in the bulge region of

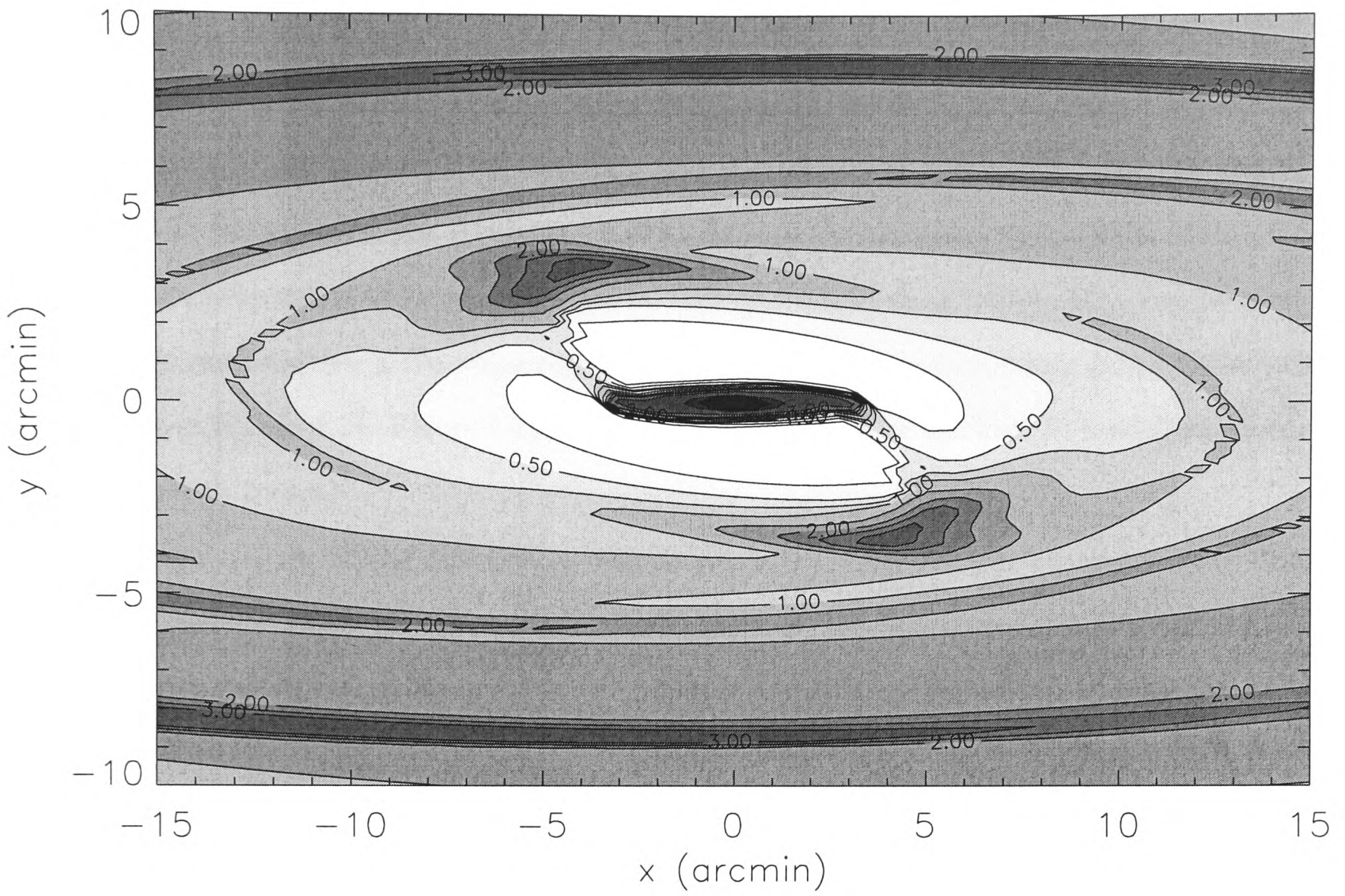


Figure 4.14: Absorption map for the bulge of M31. Derived from the gas density field of the model of maximum likelihood rotated and inclined to the observer's viewpoint. The contours and greyscale show the B band absorption expected from dust. Contours are at 0.25, 0.5, 0.75, 1.0, 1.5, 2.0, 2.5, 3.0, 5.0, 10, 20 and 90 B band magnitudes.

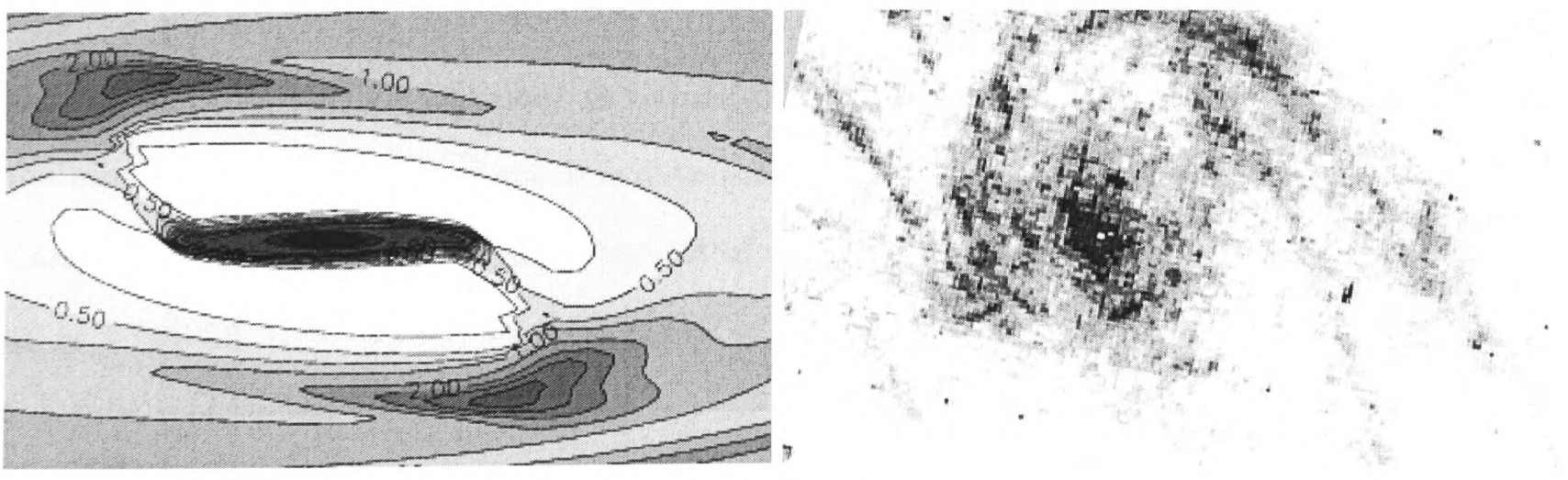


Figure 4.15: Comparison of the central region of the absorption map with 7.7 μm observations of James Lequeux. The X axis extends from $-7'$ to $10'$ and the Y axis extends from $-5'$ to $5'$.

$A_B = 0.98$ mag (van Genderen 1973), it is possible to calibrate the gas density and determine an absorption map for the bulge.

Taking the column density of interstellar hydrogen to be $N(\mathrm{H_{tot}}) = N(\mathrm{H\,I}) + 2N(\mathrm{H_2}) = 1.9 \times 10^{25} A_V$ m^{-2} and $A_B = 1.3\ A_V$ (Binney and Merrifield 1998), it follows that the mean gas density in the bulge region should be $\bar{\Sigma}_{\mathrm{H_{tot}}} = 11\ M_\odot$ pc^{-2}. Assuming that CO is a reasonable tracer of both interstellar hydrogen and dust in M31 (Neininger et al. 1998), $\bar{\Sigma}_{\mathrm{H_{tot}}}$ can be used to calibrate the gas densities in the bulge region and determine the corresponding B band absorption map, shown in Figure 4.14. Figure 4.13 shows the velocity field for the best fit model transformed to the observer's frame.

The central region of the absorption map is compared with 7.7 μm observations of James Lequeux (Lequeux 2000) in Figure 4.15. Both parts of the figure are of the same region of the bulge and are shown to the same scale. Certain features in the model, namely the central overdensity, the dust lanes at X = $\pm 4'$, Y = $\pm 1'$, and part of the ridge extending from X = $-4'$, Y = $3'$ to X = $5'$, Y = $2'$, can be seen in the observations. However, the opposite is not true. The strong spiral structure in the observations is conspicuously absent in the model. This suggests the possibility that, due to the paucity of CO observations in the bulge region, the simple relationship between CO density and dust may not be valid.

4.5 The dark halo of Andromeda

As mentioned above, the axisymmetric component is assumed to comprise a disk and any halo which exists in M31. This was done to reduce the number of free parameters in the model whilst remaining true to the observations. It is extremely important, however, to determine what fraction by mass of the inner region of M31 consists of dark matter. The inner region of M31 is defined here as a sphere inscribed by the semi-major axis of the bulge of the maximum likelihood model i.e. a sphere of radius 3.5 kpc.

To do this, the circular speed curve of the axisymmetric component of the model of maximum likelihood, shown in Figure 4.4, is taken to represent the combined mass distribution of the disk and halo. For a circular velocity of 130 km s^{-1}at a radius of 3.5 kpc, the mass of a sphere would

be $1.4 \times 10^{10} M_{\odot}$, an upper limit since the mass would be lower if the distribution was in a disk. The mass of the bulge has already been deduced to be $2.3 \times 10^{10} M_{\odot}$. Hence the total mass at 3.5 kpc is $M_T = 3.7 \times 10^{10} M_{\odot}$. However there remains a degeneracy between the contributions of the disk and the halo.

The exact value of the halo mass fraction inside 3.5 kpc is dependent on the profile adopted for disk and halo, even for a maximum disk. However, the disk mass can be estimated using the values for the B band exponential disk central surface brightness ($I_d = 21.6$ mag arcsec^{-1}) and scale length ($R_d = 5.8$ kpc) in Table 4.1. When corrected for the values of absorption and inclination from the same Table, the central surface density of the disk becomes Σ_d =113 $L_{\odot}$ pc^{-2}. With a conservative B band disk mass-to-light ratio of 4, this gives a disk mass within 3.5 kpc of $M_d = 1.2 \times 10^{10}\ M_{\odot}$. Hence the maximum mass of the dark halo within 3.5 kpc is just $M_h = 1.9 \times 10^{9} M_{\odot}$ or 5% of the total. If the B band disk mass-to-light ratio is 5, there would be no room for a dark halo at all. In comparison, the maximum disk model of Kent (1989) implies a halo mass fraction at 3.5 kpc of about 1%.

4.5.1 Predictions of Cold Dark Matter

The above estimate is dependent on the uncertain value of the disk mass-to-light ratio. An alternative method is to use the predictions of cosmological cold dark matter simulations which fit galactic halos with NFW profiles (Navarro et al. 1996),

$$\frac{\rho(r)}{\rho_s} = \frac{1}{(r/r_s)(1 + r/r_s)^2}, \tag{4.7}$$

where ρ_s is a characteristic density, $r_s = r_{200}/c$ is a characteristic radius and c is a dimensionless concentration parameter. r_{200} is the radius within which the mean density of the halo is 200 $\rho_{\rm crit}$, where $\rho_{\rm crit} = 3H^2/8\pi G$ is the critical density of the universe, $H = h\,H_0$ is Hubble's constant, $H_0 = 100$ km s^{-1} Mpc^{-1} and h is taken to be 0.75.

The circular velocity curve for an NFW profile is given by

$$\left(\frac{v_{\rm halo}(r)}{v_{200}}\right)^2 = \frac{1}{x}\frac{\ln(1 + cx) - (cx)/(1 + cx)}{\ln(1 + c) - c/(1 + c)}, \tag{4.8}$$

where v_{200} is the velocity at r_{200} and $x = r/r_{200}$. At a redshift $z = 0$,

$$r_{200} = \left(\frac{h\, v_{200}}{\mathrm{km\ s^{-1}}} \right) \mathrm{kpc}. \tag{4.9}$$

Hence, for a given value of Hubble's constant, the halo circular speed curve is fully specified by v_{200} and c. Navarro et al. (1997) find a relationship between the characteristic density and radius (or, equivalently, between c and v_{200}) of a halo in a given cosmology, thereby reducing the NFW profiles to a one parameter family.

If $v_{200} = v_0 = 257$ km s^{-1}, the method of Navarro et al. (1997) gives $c = 7.5$, a halo mass at 3.5 kpc of $1.1 \times 10^{10} M_\odot$ (30% of the total mass) and a disk mass of just $2.6 \times 10^9 M_\odot$. However, Navarro et al. (1996) state that the maximum rotation velocity can be as high as $1.4\, v_{200}$. So if $v_{200} = v_0/1.4 = 184$ km s^{-1}, the halo mass at 3.5 kpc falls to $8.4 \times 10^9 M_\odot$ (22% of the total mass) and the disk mass rises to $5.5 \times 10^9 M_\odot$. In either case, the inner region of M31 would include significant quantities of dark matter.

4.5.2 Implications of fast bars

Debattista and Sellwood (2000) convincingly demonstrate that fast, long lived bars cannot co-exist with massive dark halos. Their argument is based on the idea that dynamical friction from a dense dark matter halo slows down a rapidly rotating bar. Hence, a fast rotating old bar precludes the existence of a dark halo, unless either it has the lowest possible central density consistent with a monotonically decreasing density profile or it is strongly rotating.

'Fast' bars have $1.0 < \mathcal{R} < 1.4$ and as shown in Chapter 1, of the few galaxies where $\mathcal{R}$ has been measured, all have fallen within this range. Debattista and Sellwood (2000) argue that the vast majority of barred galaxies cannot contain strongly rotating halos. Therefore, barred galaxies should all have minimal halos and maximum disks. And unless the distribution of halo densities is bimodal, this conclusion of minimum halos also applies to unbarred spiral galaxies.

The bulge of M31 is older than 6 Gyr (Trager et al. 2000) or 52 bar rotation periods. Hence, there has been plenty of time for the triaxial bulge of M31 to slow down in the presence of a high density inner dark halo. Since the model of maximum likelihood for M31 is fast, with $\mathcal{R} = 1.2$,

M31 must have a low density inner dark matter halo and a maximum disk, or a strongly rotating halo. According to Figure 7 of Debattista and Sellwood (2000), it is only possible to maintain a fast bar with a value of $\mathcal{R} = 1.2$ for longer than ~ 5 Gyr if the angular momentum of the halo is more than 50% of the disk angular momentum to radii past the end of the bar. This corresponds to a value of the dimensionless spin parameter $\lambda \equiv L|E|^{1/2}/(GM^{5/2})$ of 0.12.

Steinmetz and Bartelmann (1995) show analytically that the probability of forming a halo with $\lambda \geq 0.12$ is less than 10%, corroborating the numerical simulations of Barnes and Efstathiou (1987). Spinning up an entire halo since its formation is extremely unlikely, since the total mass of a dark halo is thought to be much greater than that of the baryonic galaxy. And if a large quantity of angular momentum could be given up to the inner halo by the disk, it would seem to be dynamically perverse for that angular momentum to be retained in the inner halo rather than disperse outwards, leaving the inner parts slow again. Lastly, observations of the halo of M31 (Huchra 1993) show that the system of globular clusters within it rotates slowly at a rotation velocity of 70 km s^{-1}. Hence, the dark halo of M31 is probably not a fast rotator. Since the simulations predict a triaxial bulge which is rotating fast, the halo must be minimal in its inner parts. This is clearly at odds with the higher density NFW halo implied by CDM.

4.6 Conclusions

The mass distribution for the bulge of M31 has been determined from observations of the bulge surface brightness profile. The bulge is triaxial and rotates about the same axis as the disk. The DUAL code, based on the FS2 algorithm, has been used to reproduce observations of CO position and velocity in the bulge of the unbarred Andromeda galaxy, particularly the extreme velocities which appear in the central 10′. By matching the model gas velocity field to observations, the B band bulge mass-to-light ratio Υ_B, the bulge pattern speed Ω_p and the angle between the minor axis of the bulge and the line of nodes of the disk ϕ have been constrained at the 95% confidence level to $5.9 < \Upsilon_B < 7.2$, $53 < \Omega_p < 57$ km s^{-1} kpc^{-1} and $9° < \phi < 18°$. For the model of maximum likelihood, $\Upsilon_B = 6.5$, $\mathcal{R} = 1.2$, the semi-major axis $a = 3.5$ kpc and $\phi = 15°$.

This value of the B band mass-to-light ratio is consistent with Kent (1989) who finds that

$\Upsilon_B = 7.1 \pm 0.5$. It does not, however, fit the log-log relationship between the B band mass-to-light ratio and the total B band luminosity of elliptical galaxies described by Gerhard et al. (2001) if this is naively extended to galactic bulges. The integrated B band luminosity for the bulge of M31 (Walterbos and Kennicutt 1988) when corrected for absorption is 1.5×10^9 $L_\odot$. According to Figure 13 of Gerhard et al. (2001), this should correspond to a B band mass-to-light ratio of just $\Upsilon_B = 1.5$. Based on the data of Gerhard et. al., the validity of this log-log relationship seems far from conclusive. It would be interesting to see a selection of galactic bulges added to their data set to test whether their conclusions apply at low luminosities.

This best fit model has been shown to be consistent with further CO observations away from the line of nodes of the disk, if the kinematics of gas away from the plane of the disk is included. To account for the whole velocity range covered by the observations, a finite, fairly large thickness (half width at half maximum intensity of 200–500 pc) must be given to the molecular disk. It is quite unclear, however, whether the inner molecular disk of M31 is truly that thick or whether adding it to the model is merely an effective theoretical trick able to account for a warping of a thinner disk. Other out of plane, or peculiar velocities (such as those associated with hydraulic jumps or turbulence) could also contribute to the wide velocity range covered by the CO observations. This conclusion is consistent with other recent work on central gas disks in spiral galaxies.

These results were calculated using a distance to M31 of 690 kpc for ease of comparison with previous work. Stanek and Garnavich (1998) use recent data from Hipparcos to deduce that the distance is 784 kpc. Adopting this value means that $1' = 230$ pc on the line of nodes of the disk. Observed or deprojected lengths and the mass of the bulge may be easily adjusted by scaling in proportion to the distance, whereas the central density scales as the inverse square of the distance.

The procedure described above improves on previous bulge models by constructing the mass distribution from observations. Using a hydrodynamics code to determine the response of the gas to the underlying potential provides a more accurate representation of the gas velocity field than calculating periodic orbits. However, the oscillations inherent in hydrodynamic models limit the simulations' accuracy and may lead to more slack in the predictions than the formal errors imply.

The absorption map provided in Figure 4.14, shows correlations with the observations of dust

in the bulge region by Lequeux (2000). In particular, the reproduction of the offset dust lanes of the model in the observations provides further evidence for the triaxiality of the bulge. However, the model has been unable to account for the spiral structure seen in these observations.

Since the bulge of M31 is a fast rotator and its halo is likely to be rotating slowly, it follows that M31 must have a maximum disk and a halo mass fraction within 3.5 kpc of just a few percent. This is in stark contrast to CDM predictions which imply that the halo mass fraction within 3.5 kpc is between 22% and 30%.

Triaxiality has now been demonstrated for both our Galaxy and our nearest large neighbour M31, so it seems likely that triaxial bulges are a common feature of both SA and SB spiral galaxies.

Chapter 5

Simulations of dual morphology in spiral galaxies

5.1 Observational and theoretical motivation

Although optical images of spiral galaxies are often beautiful, they can be deceptive, masking the underlying morphology of a galaxy by obscuring the bulk of its mass behind a veil of dust and bright young stars. Most of the mass is in the form of old stars which are best observed in infrared images. Young stars contribute much of the luminosity in optical images, but they make up very little of the overall gravitational potential.

Despite the fact that young and old stars orbit in the same potential, the morphology of images in the infrared and optical are sometimes completely uncorrelated. In this Chapter, the DUAL code is used to discover whether it is possible to recreate these dual morphologies using numerical simulations, and whether predictions can be made about the intrinsic characteristics of spiral galaxies from their morphologies. Of course, crude numerical simulations such as these cannot hope to reproduce the wealth of structure and form of a spiral galaxy. However, it is hoped that these simulations include enough physics to gain some insight into the cause of dual morphologies.

Section 5.1 briefly describes the observations which motivate this project and the astrophysics which is needed to explain them. In Section 5.2, I provide the details of the models used to recreate dual morphologies and in Section 5.3, the various methods used to analyze them. The results are presented in Section 5.4 and they are discussed further in Section 5.5. Conclusions and suggestions for future work are presented in Section 5.6.

5.1.1 Summary of observations

As mentioned in Section 1.5, observations of many spiral galaxies have been made simultaneously in the optical and the infrared. Often the difference in morphology between B and K' band is not very dramatic (Knapen, private communication), but sometimes radical differences are revealed

in the morphologies of the two wavebands. In particular:

- B and K' band images of the same galaxy can be completely decoupled (Block and Puerari 1999), for example NGC 5185 in Figure 5.1;
- spiral pitch angles can be tighter in B band than in K' band by upto 7° (Grosbol and Patsis 1998);
- bars are more prominent in the near infrared than in the optical (Eskridge et al. 2000), for example NGC 309 in Figure 5.2;
- grand design spirals are very common in K' band images whereas most B band images are flocculent (Grosbol and Patsis 1998), for example M63 in Figure 5.3;
- K' band images are more symmetric than B band images (Block and Puerari 1999), for example NGC 5195 in Figure 5.1;
- K' band images are mostly one or two armed whereas B band are frequently multi armed (Block and Puerari 1999), for example NGC 309 in Figure 5.2;
- K' band images are smoother than those in the B band.

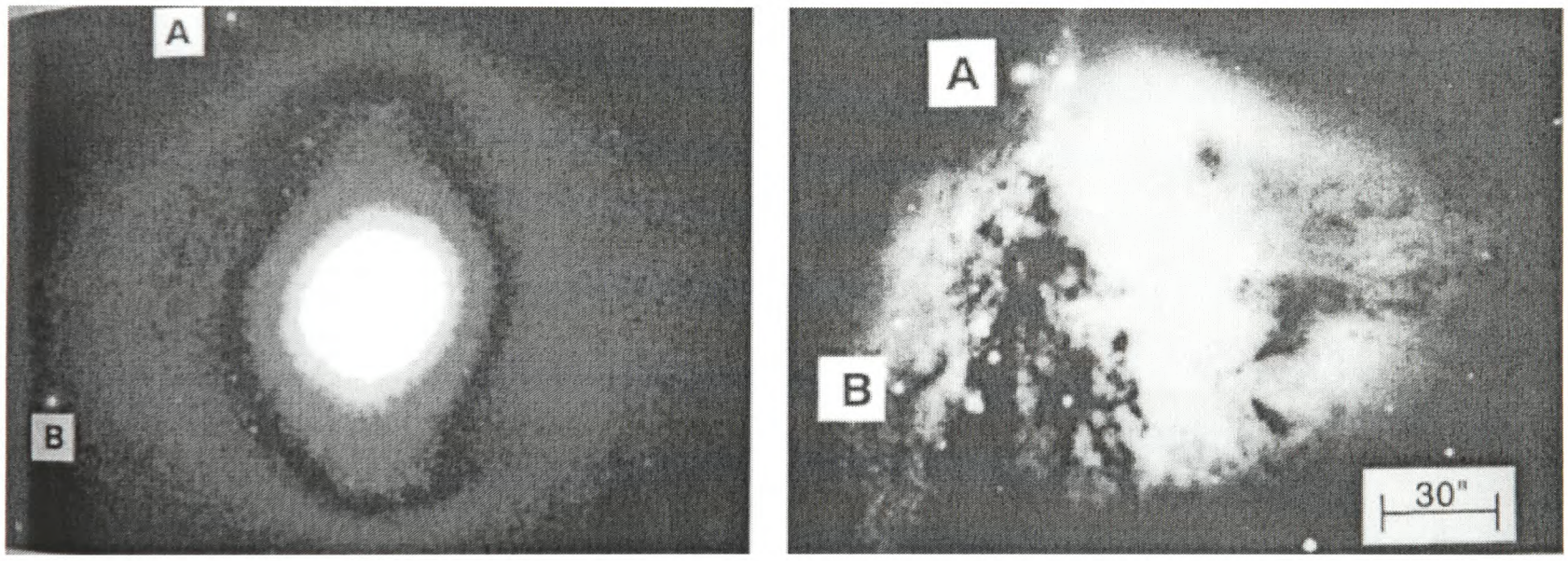


Figure 5.1: K' and B band observations of NGC 5195 (Block et. al. 1994).

5.1.2 Do optical and infrared images trace separate populations?

Spiral galaxy disks can be decomposed into two dynamically distinct components. One forms only a small part ($\sim 5\%$) of the mass of a galaxy disk and is composed of phenomena that evolve fast: young hot O and B stars, H II regions, cold interstellar H I gas, and dust. Dynamically this part of the disk is 'cool', being characterized by small random motions in the disk plane of $\sim 10\,\mathrm{km\ s^{-1}}$. Assuming the disk to be a locally isothermal sheet (see Equation 5.2), with a surface density of

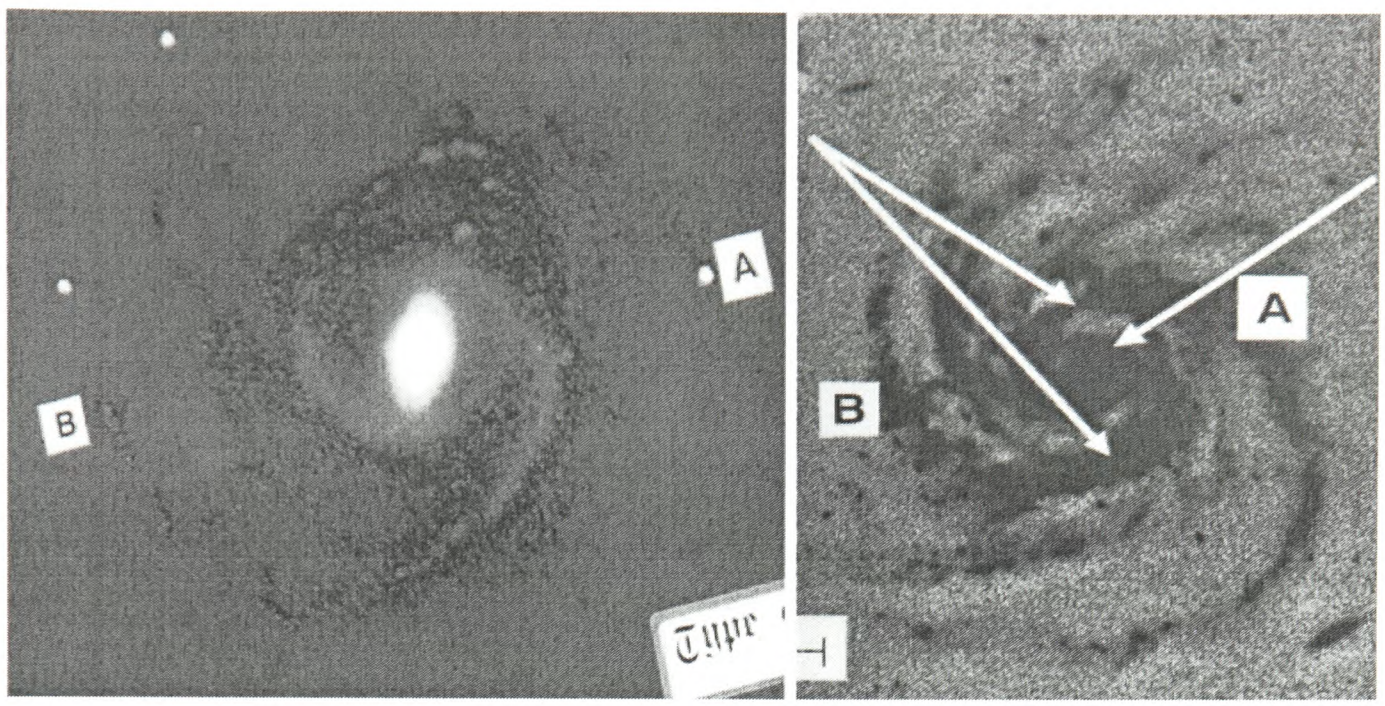


Figure 5.2: K' and B band observations of NGC 309 (Block & Wainscoat 1991).

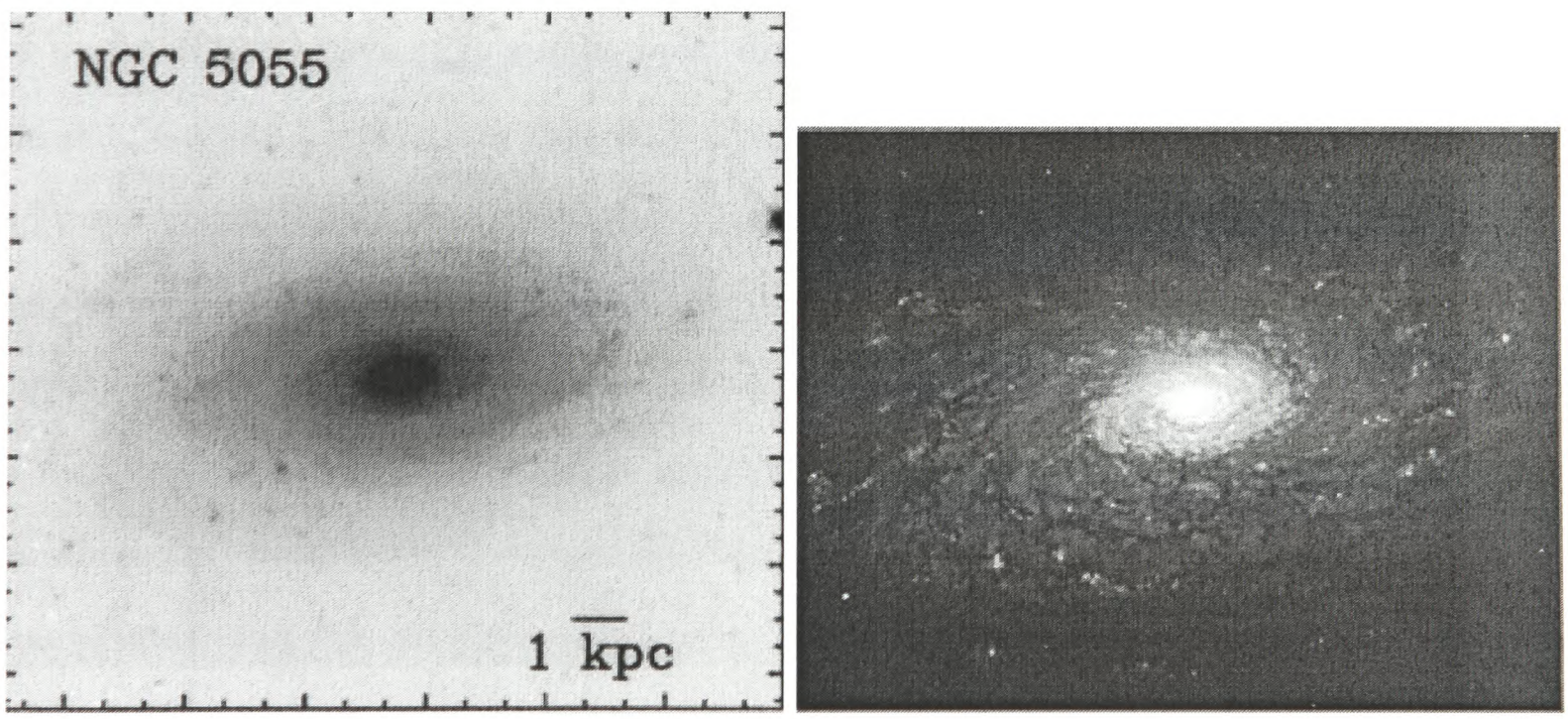


Figure 5.3: K' band (Thornley 1996) and B band (Subaru 2000) observations of NGC 5055.

71 $M_\odot$ pc^{-2} equal to that near to our sun (Kuijken and Gilmore 1991), this leads to an unstable thin disk of scale height ~ 37 pc.

The second component forms the bulk of the mass of the galaxy and is predominantly composed of old K and M dwarf stars. It is dynamically 'warm' and hence stable, and characterized by a vertical velocity dispersion of $\sim 20\,\mathrm{km\ s^{-1}}$ (Dehnen and Binney 1998). Thus it forms a thicker disk of scale height ~ 410 pc.

The dust penetrated classification system of Section 1.5.2 makes the assumption that the light in K' band images faithfully traces the underlying mass distribution of the galaxy being observed. It also makes the implicit assumption that B band observations do not trace the distribution of mass. These two assumptions are the basis for modelling dual morphology galaxies using a combination of N-body and hydrodynamic techniques and so it is necessary to examine how accurate they both are.

B band images trace the cool component

In Section 2.1.1, I justify the use of isothermal hydrodynamics as a model of interstellar gas. I now assume that hydrodynamic simulations provide an accurate representation of spiral galaxies viewed in the B band (445 nm), and so must discuss what we see in B band images.

Rix and Zaritsky (1995) show that 30% of the luminosity emitted by stars in B band images comes from a young stellar population (less than a few dynamical periods or 3×10^8 yrs), and 35% from an old population (more than 3×10^9 yrs). Hence, old stars form a substantial contribution to the overall luminosity of B band images. However, old stars will be more evenly spread around a galaxy disk, having had many dynamical periods in which to disperse away from the star forming regions where they were born; their contribution to B band luminosity is global rather than local. But young stars, and the gas from which they are formed, will be concentrated in star forming regions, and, in the absence of strong shear, these will also be regions of high gas density.

Hence, it is reasonable to assume that high density regions of hydrodynamic simulations reproduce the morphology of high luminosity regions in B band images. This assumption carries the caveat that significant areas of a B band image may be obscured due to absorption by dust.

K' band images trace the warm component

For N-body methods to be an effective representation of infrared (K' band, 2.1 μm) images, those images should trace the underlying, collisionless mass distribution of the galaxy.

Binney and Merrifield (1998) show that 84% of the mass density in the solar neighbourhood is due to old ($> 10^{10}$ yrs or 100 galactic rotation periods) K and M dwarf stars and white dwarfs. These stars follow the underlying mass distribution rather than regions of star formation in which they were born. So the two questions to address are how much of the K' band radiation traces the old star population and how much local contamination is there from dust and young stars?

Johnson (1966) shows that 98% of the radiation in his solely stellar, synthetic K band images is due to K and M stars. Rieke and Lebofsky (1985) detail the extinction law due to dust at infrared wavelengths, concluding that there is a negligible contribution to the K band light from dust.

Rix (1993) and Rix and Rieke (1993) determine that dust reduces the K band flux by only 10%, even at the positions of the dust lanes. Young red supergiants (the most likely contaminators) only contribute significantly to the K band light in localized regions within dust lanes. Away from the dust lanes, supergiants contribute $< 15\%$ of the luminosity. This implies that K band images are good tracers of the old disk population.

More recently, Rhoads (1998) used the 2.3 μm CO absorption index to deduce that young red supergiants contribute $\sim$ 3% to the global K' band flux but upto 33% in local features. And Rix and Zaritsky (1995) cite an upper limit of 20% for the contamination of K' band luminosity by young stars. Hence a possible bias exists in arm-interarm contrast amplitudes and other measurements of galactic substructure.

5.1.3 Dual morphologies within the modal framework

The regular one or two armed spiral structures seen in the infrared and the multi armed nature of optical images, can be partly accounted for within the modal theory of spiral structure of Bertin and Lin (1996). This theory asserts that large scale spiral structure is long lasting and determined by intrinsic properties of galaxies rather than by external effects (mergers, collisions, etc.). Spirals

are associated with a few slowly evolving global modes which form a standing density wave pattern. Modes are produced by the cool excitable outer disk, propagate inwards and are refracted back again at the centre by either a central bulge or a dynamically warm inner disk.

This feedback process is modified by the presence of an Inner Lindblad Radius (ILR). In a pure stellar disk the ILR acts as a perfect wave absorber, destroying the feedback process, whereas in a gaseous disk the ILR is imperfect in its wave absorption. Modes for which the radius of the ILR is greater than the radius of refraction can only make a small contribution to the spiral pattern of a galaxy. For a stellar disk, this means that modes with $m \gtrsim 3$ are unlikely to be present. This is the case for K' band, stellar images which are largely characterized by one or two armed spirals. However, gaseous disks are composed of a number of dynamically different constituents (hot gas, cold gas, OB stars, etc.) and have imperfect ILR's. Hence gaseous B band images exhibit many armed, flocculent spirals.

In one respect it is not surprising that the dynamically distinct gaseous and stellar disks exhibit different morphologies since they are composed of very different astrophysical objects. However, being part of the same galaxy implies that they must be coupled in some way. The most obvious mechanism is the gravitational potential which is largely determined by the old warm disk and responded to by the young cool disk. This situation is not so simple in the outer disks of spiral galaxies, however, where there is a much higher proportion of gas and hence the potential is determined by both stars and gas.

The gas also plays a very important role in the self regulation of the disk. A pure stellar disk gradually converts ordered kinetic energy into random motion, leading to a hot, stable system in which spiral structure could not be present. Cold, dissipative, interstellar gas balances this stellar heating and maintains the disk in a marginally stable state with respect to local Jeans instability. Eventually the stellar disk will become hot enough to decouple from the gas and only small scale spiral structure will survive, dooming all spiral galaxies to eventual flocculence.

Figure 5.4 shows a method of categorizing galaxies in a 3D parameter space on the basis of their intrinsic properties (Block et al. 1994). The barredness of a disk galaxy is determined by the active disk mass relative to the total mass of the galaxy, with heavy disks possessing a bar

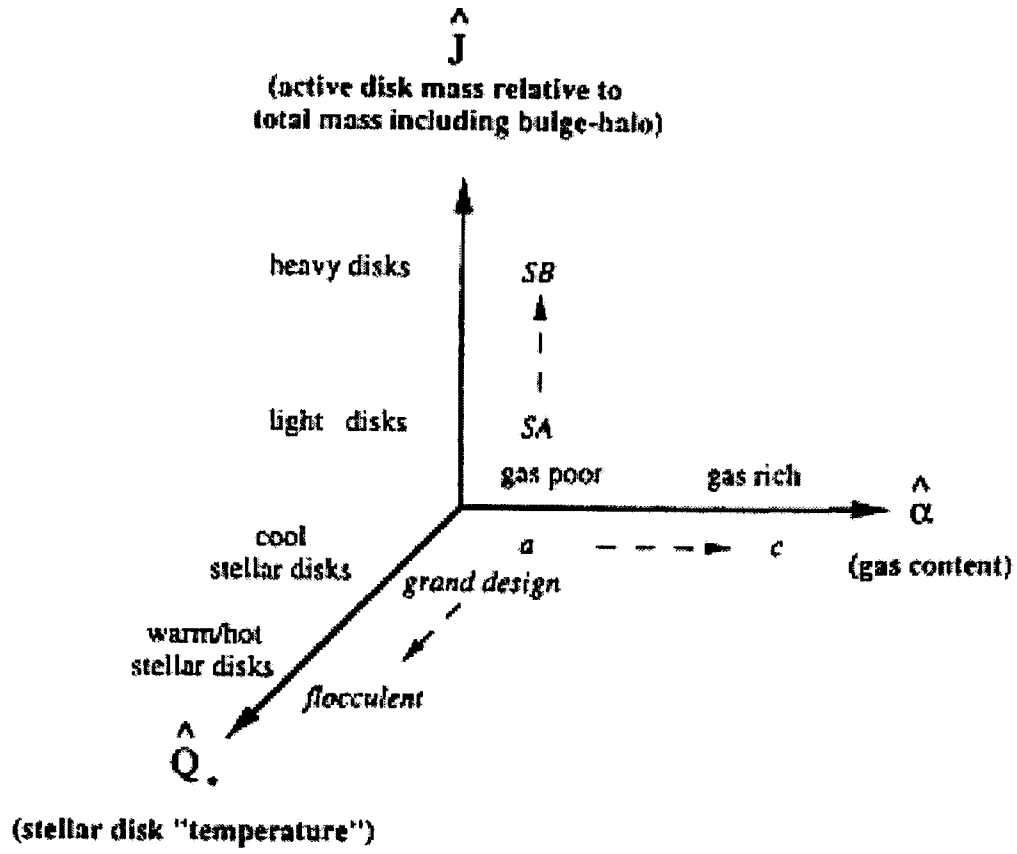


Figure 5.4: Framework for the classification of the morphologies of spiral galaxies on the basis of their intrinsic modal characteristics.

(SB galaxies) and light disks being unbarred (SA galaxies). The transition between the Hubble types (a) to (c) is dictated by the fraction of gas by mass, with (a) galaxies being gas poor and (c) galaxies being gas rich. Finally, grand design galaxies are associated with warm stellar disks (relative to that required for marginal axisymmetric stability) whereas flocculent spirals have cool stellar disks. These relationships will be investigated using the models described below.

5.2 Modelling dual morphology galaxies

In modelling the above observations, the stellar and gaseous components of the galaxy models evolve freely. The gas is represented by an isothermal fluid, the stars by N-body particles and the background dark halo by an analytic potential. The stars and gas both feel the potential of the dark halo and the stars, however the gas is not self gravitating and hence the stars do not feel the gas. No bulge component is used and so the models best represent Sc galaxies.

The units used in this Chapter are the disk scale length R_d and the disk mass M_d with G set equal to unity. Thus the unit of time is $t_d = (R_d^3/GM_d)^{1/2}$. Although all results are presented in these units, it is sometimes helpful to transform to kpc, $M_\odot$ and Gyrs. In these cases, I use the

canonical values of $R_d = 5$ kpc and $M_d = 2 \times 10^{11}\,\mathrm{M}_\odot$ which gives a dynamical time of $t_d \sim 10^7$ yrs and a velocity of $v_d \sim 420$ km s^{-1}.

The grid on which both the hydrodynamic and N-body calculations take place is two dimensional and Cartesian, and the shortest distance from the centre to the edge $R_G = 7.5$ disk scale lengths.

5.2.1 The stellar disk

The stars are modelled as a one component disk with exponential surface density and radial velocity dispersion profiles,

$$\Sigma_d(R) = \Sigma_d \exp(-R/R_d), \tag{5.1}$$
$$\sigma_R(R) = \sigma_R \exp(-R/R_\sigma),$$

where $R_d = 1$ and $\Sigma_d = 0.5/\pi$ to ensure that $M_d = 2\pi\Sigma_d R_d^2 = 1$. For an exponential disk, the maximum circular speed is 0.62 and occurs at $R = 2.15$.

Although simulations all take place in 2D the stellar particles have an effective scale height due to their velocity dispersion, the softened gravitational forces which are used in N-body simulations and the effect of the interpolation of gravitational forces from the grid to the positions of the particles. Following van der Kruit and Searle (1981), I assume the stellar disk to be a locally isothermal sheet, which implies that

$$\rho(z) = \rho_0 \,\mathrm{sech}^2(z/z_0),$$
$$z_0 = \sigma_z^2/\pi\, G\, \Sigma_d, \tag{5.2}$$

where $\sigma_z \sim 0.6\,\sigma_R$ (Dehnen and Binney 1998) is the vertical velocity dispersion. For an exponential radial velocity dispersion with $R_\sigma = 2$, this gives a constant scale height of

$$z_0 = 0.72\,\sigma_R^2, \tag{5.3}$$

where σ_R remains to be determined, consistent with the scale height of late type galaxies being approximately constant (de Grijs and Peletier 1997).

This scale height is equated to the Plummer length $s_{\rm grid}$ of Section 2.3.3 - the effective height of the 2D N-body particle distribution. If $s_{\rm grid}$ is written in terms of grid cells, then $z_0 = (2R_G/I)\, s_{\rm grid}$, where I is the number of cells on each side of the grid. Restricting the Plummer length to $1.2 < s_{\rm grid} < 2.3$ grid cells (see Section 2.3.3) and taking realistic values for the scale height of $0.10 < z_0 < 0.14$ (where $z_0 = 0.14$ is given by van der Kruit and Searle (1981)) limits the number of grid cells to 250×250. Choosing $N = 2 \times 10^5$ particles to model the stellar component gives a mean of 3 particles per grid cell.

5.2.2 The dark halo

The form of the dark halo potential is an isochrone, used because it has a core rather than a cusp to its density profile, falls to zero density at large radii and its potential and circular speed can be defined using just the halo mass M_h and its scale length R_h,

$$\begin{aligned} \phi_h(r) &= -\frac{GM_h}{R_h + a}, \\ v_{c,h}^2(r) &= \frac{GM_h r^2}{a\,(R_h + a)^2}, \end{aligned} \tag{5.4}$$

where $a \equiv \sqrt{R_h^2 + r^2}$. The slope of the total rotation curve at four disk scale lengths is fixed at zero. Since the disk is already determined, the halo mass and scale length are both fixed by choosing the disk fraction at four disk scale lengths f_d.

Figure 5.5 shows the halo mass fraction profile for a variety of values of f_d. When $f_d \geq 0.4$ (true for 9 out of 16 models) the dark matter content within one disk scale length is minimal ($< 15\%$) and hence the galaxy model is dominated by the disk. At the lower value of $f_d = 0.3$, the halo mass fraction at one disk scale length rises to 27%.

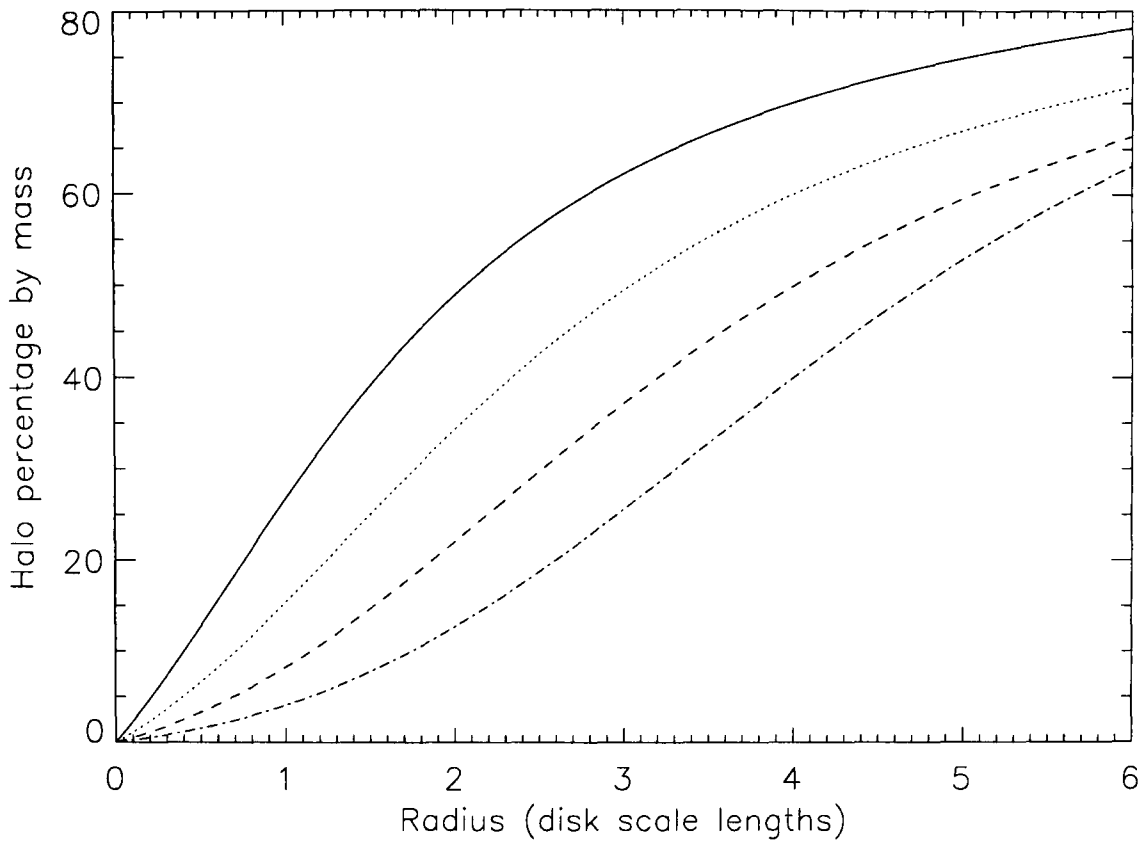


Figure 5.5: Radial profile of the halo mass fraction for four different disk mass fractions $f_d = 0.3$, 0.4, 0.5 and 0.6, where f_d is defined at four disk scale radii.

5.2.3 Stellar distribution function

The initial conditions of the stellar particles are drawn from the warm disk distribution function of Dehnen (1999a),

$$f(E,L) = \frac{\gamma(R_E)\Sigma_d(R_E)}{2\pi\sigma_R^2(R_E)} \exp\left[\frac{\Omega(R_E)[L - L_c(E)]}{\sigma_R^2(R_E)}\right], \tag{5.5}$$

where R_E and $L_c(E)$ are the radius and angular momentum of the circular orbit at energy E, Ω is the angular velocity and $\gamma \equiv 2\Omega/\kappa$. Particles are chosen from this distribution function within radii smaller than $R_{\rm cut} = 5$ to ensure that particles stay within the grid. Since an exponential disk extends to infinity, the mass of the stellar disk on the grid is only 0.96 mass units.

The distribution function $f(E,L)$ is sampled using the following procedure and the accompanying code of Dehnen (1999a) to generate initial positions and velocities of the stellar particles with minimal noise:

1. Choose a random radius R from the surface density distribution based on Σ'_d.
2. Determine $E = E_c(R)$ and $g_1 = \Sigma'_d/\Sigma_d$.

3. Choose $\zeta \in (0,1)$ randomly and determine $L = L_c(E) + \sigma_R^2(R) \ln \zeta / \Omega(R)$.
4. If $E \notin [\phi(0), \phi(\infty)]$ or $L \notin [-L_c(E), L_c(E)]$ go back to step 1.
5. Integrate the orbit (E, L) for one radial period T_R and determine $g_2 = \kappa(R)\, T_R / 2\pi$.
6. Sample $N_{\rm sam}$ phase space points quasi-randomly from $t \in [0, T_R)$. $N_{\rm sam}$ is either of the two integers next to $g_1 g_2$ chosen with probabilities such that their mean is equal to $g_1 g_2$.
7. Repeat steps 1 to 6 until the desired number of particles have been sampled.

Dehnen (1999a) includes an iterative algorithm for getting a better match to the target surface density and radial velocity dispersion profiles. It replaces the input values Σ_d and σ_R by other values Σ_d' and σ_R'. This is followed for just one iteration since further iterations do not significantly improve the distribution function's accuracy.

As mentioned in Section 3.4.4, this process results in a surface density profile which is somewhat lower in the centre than intended. Hence the DUAL subroutine `getacc` is amended to ensure that the forces due to the axisymmetric part of the potential will equal the theoretical rotation curve.

5.2.4 Stability and rotation

As shown in Chapter 2, Q determines the response of the stellar disk to axisymmetric instabilities. According to Fuchs (2000), $1 < Q < 2$ is needed to prevent Jeans instability of the disk and yet allow the disk to develop spiral structure. Since $\kappa(R)$ and $\Sigma_d(R)$ are both determined from the stellar disk and the halo, it can be seen from Equation 2.46 that defining $Q(R)$ fixes $\sigma_R(R)$. In fact it is enough to choose the minimum value of $Q(R)$, $Q_{\rm min}$, and fix σ_R. The left hand panel of Figure 5.6 shows the radial profile of Q at the start of a typical simulation.

Figure 5.7 shows the total rotation curve of the model and the contribution of the stellar and dark components for a typical simulation. The model (dashed) is in reasonable agreement with observed rotation curves of low surface brightness (dotted) and high surface brightness (solid) late type galaxies (normalized to their rotation velocity at two scale lengths) from Swaters et al. (2000). The right hand panel of Figure 5.6 shows the angular velocity profile derived from this rotation curve. From this and the determination of the bar pattern speed from the simulation results (see Section 5.3.8), it can be seen that these models have no Inner Lindblad Radius, since there is

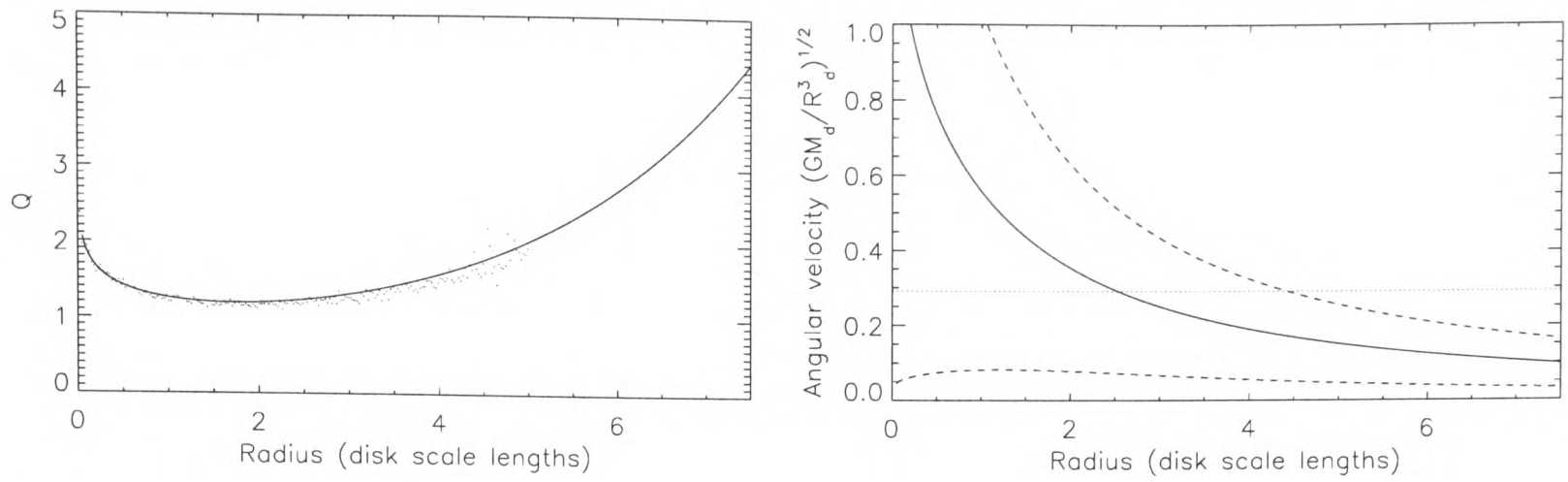


Figure 5.6: Left panel: radial profile of Q at the start of `sta19`. Dots show Q values derived from initial particle positions. The solid line is the theoretical profile based on an exponential disk and isochrone halo. Right panel: angular velocity profiles for `gas19`. The solid line is $\Omega(R)$, the dashed lines are $\Omega(R) \pm \kappa(R)/2$ and the dotted line is the bar pattern speed Ω_p which is determined from the results of the simulations.

no radius at which the bar pattern speed Ω_p is equal to $\Omega - \kappa/2$. The figure can also be used to determine the corotation radius and Outer Lindblad Radius - the radii where the bar pattern speed Ω_p is equal to Ω and $\Omega + \kappa/2$ respectively.

5.2.5 The gaseous disk

The response of the gas is governed by the isothermal Euler equations with gravitational source terms as implemented by the FS2 algorithm. Both the Euler equations and the FS2 algorithm have been described in Chapter 2. In these models there is no fixed pattern speed (i.e. simulations take place in the inertial frame) and gas recycling is often not used (when it is used, it makes little difference to the gas morphology and simply reduces the build up of gas in the centre of the grid).

The sound speed is set to 0.025, equal to 10 km s^{-1} assuming the canonical values of length and mass given in Section 5.2. As mentioned in Chapter 2, Englmaier and Gerhard (1997) find significant differences in gas shock morphology when the sound speed is changed in their simulations. To discover what effect the sound speed has on the simulations of this Chapter, I ran variations of models that turned out to have grand design (run `16`) and multi armed (run `19`) gas spirals using sounds speeds of 0.01 and 0.05. In both models, the smallest scale structures are wiped out as the sound speed rises. Higher sound speeds also lead to lower density contrasts in the gas spirals. However, the values of pitch angle and flocculence (as defined in Section 5.3) are

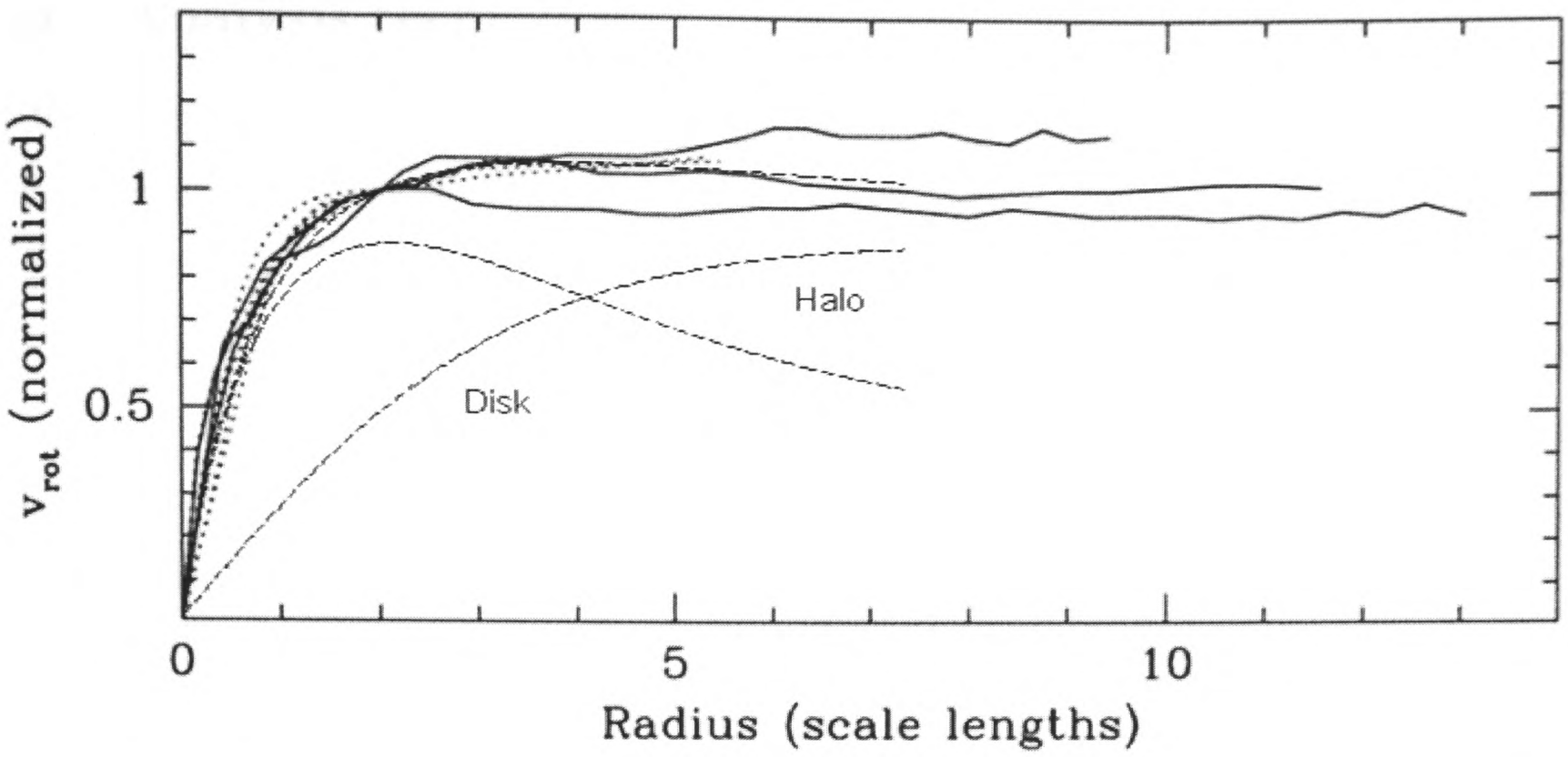


Figure 5.7: Rotation curve for a typical simulation showing the total model and the contribution from the stellar disk and dark halo (dashed lines). Also shown are observed rotation curves from Swaters et al. (2000) for late type HSB and LSB galaxies, normalized to their rotation velocity at two scale lengths.

unchanged. Hence at sound speeds below 0.05, the conclusions of this study would remain valid. But at higher sound speeds, the reduction in small scale structure would bring the stellar and gaseous morphologies closer to one another.

At the start of the simulations, the gas is set on circular orbits at a uniform surface density Σ_g, determined by fixing the total gas mass relative to the disk mass (0.01 for most simulations). Since the distribution of the stellar particles is cut at $R_{\rm cut}$, the gas density is relatively higher in the outer parts of the grid as observed in real galaxies. But without self gravity, the gas is effectively massless and the value of Σ_g is immaterial. This has been verified by performing simulations with total gas mass of 0.1 and recovering the same morphologies and velocities.

5.2.6 Simulation run time

Simulations run for 160 time units. This can be compared to the rotation time for a particle at $R = 2$, $t_{\rm rot}(2) = 2\pi R/v_c(2) = 12.6/v_c(2)$, where $v_c(2)$ is the rotation velocity at $R = 2$ as determined from the rotation curve. For most simulations $t_{\rm rot}(2) \sim 16-20$ and so the simulations run for roughly 8 - 10 rotation periods.

5.3 Analysis techniques

In the following sections, quantitative comparisons are made between the stellar and gaseous morphologies in the same simulation at a single time slice and within a time sequence, between morphologies in different simulations, and between simulation results and real B band and K' band galaxy images. These comparisons are made by defining variables which quantify the morphology of a galaxy and which can address the discrepancies between B and K' band observations. In particular, the size of the bar, number of spiral arms, spiral arm pitch angle, arm-interarm contrast, flocculence and asymmetry are used.

5.3.1 Animating the results

To view the simulations, I construct animations consisting of 80 time slices taken at intervals of 2 time units. IRAF is used to convert the three dimensional ASCII data files into FITS format, which are then animated using AIPSVIEW[1]. The animations are used to determine the largest set of contiguous time slices exhibiting two armed spirality within each simulation. Many of the mophological characteristics defined below are calculated for each time slice and the mean value over the set of time slices gives the 'time average' for the simulation as a whole.

5.3.2 Harmonic analysis

In order to highlight the underlying spiral structure in a galaxy, the images are decomposed into logarithmic spirals of the form

$$R = R(\theta) = R_0 \exp(-\theta \tan\psi), \tag{5.6}$$

where ψ is the pitch angle, the angle between the normal to the radius vector and the ridge of the spiral arm and a measure of the tightness of the spiral, and R_0 is any of the radii at which the spiral passes through $\theta = 0$. An example is shown in Figure 5.8.

Danver (1942) and Kennicutt (1981) show that logarithmic spirals can crudely represent real spiral arms. Hence, if a galaxy image is subject to a 2D Fourier transform using logarithmic spirals

[1] available to download at `monet.astro.uiuc.edu/aipsview/av.html`

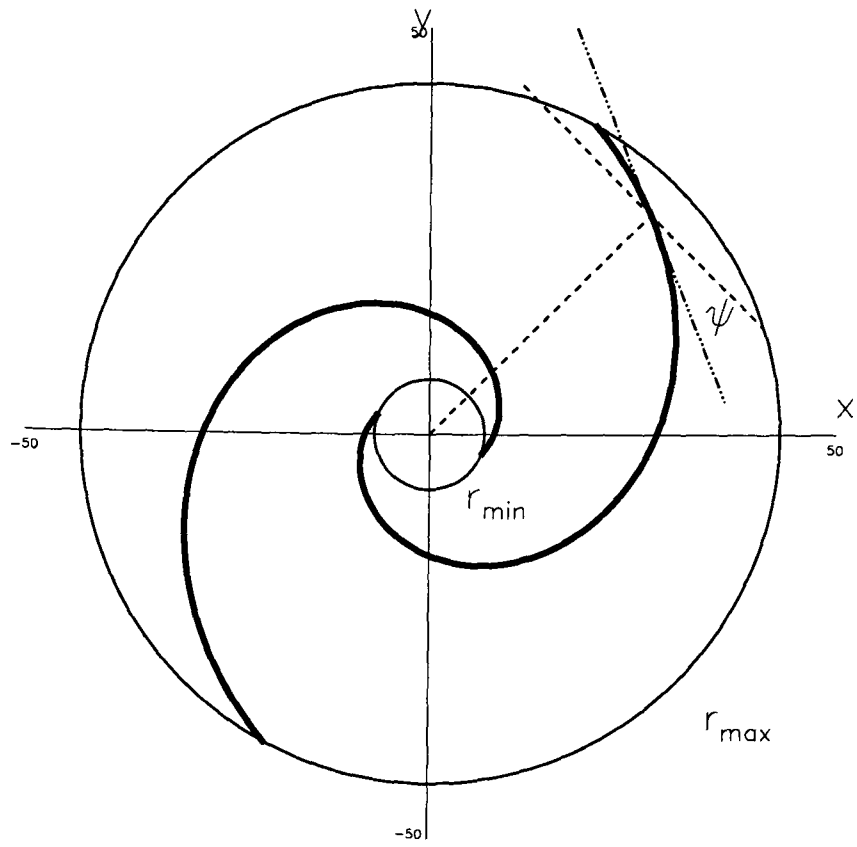


Figure 5.8: Example of a logarithmic spiral, highlighting the pitch angle ψ and the minimum and maximum radii over which it is defined $R_{\mathbf{min}}$ and $R_{\mathbf{max}}$.

as the basis states, it should be possible to reconstruct much of the structure in that image using a small number of Fourier components. For a galaxy image which has been transformed onto a regular grid in a $(\ln R, \theta)$ coordinate system, the complex amplitude of each Fourier component is

$$A(m,\alpha) = \frac{\sum_k I_k(\ln R,\theta)\, \exp[-i(m\,\theta_k + \alpha \ln R_k)]}{\sum_k I_k(\ln R,\theta)}, \tag{5.7}$$

where $I_k(\ln R, \theta)$ is the intensity of the kth pixel. The azimuthal wavenumber or number of spiral arms m and the logarithmic radial wavenumber α are related to the pitch angle by

$$\tan\psi = -m/\alpha. \tag{5.8}$$

The Fourier amplitudes are then converted into a normalized power spectrum for the non-axisymmetric ($m \neq 0$) part as follows:

$$P_{m,\alpha} = \frac{|A(m,\alpha)|^2}{\sum_{m\neq 0} |A(m,\alpha)|^2}. \tag{5.9}$$

Since it is unlikely that a single Fourier component will reproduce all of the spiral structure in a galaxy, even excluding any central bar, the pitch angle of a particular spiral pattern is generally given as the pitch angle of the Fourier component with the largest power.

In this thesis, Fourier amplitudes are calculated using 2DFFT (Saraiva Schroeder et al. 1994), a code which utilizes a 2D discrete Fourier transform in which the amplitudes are produced for -50 $\leq \alpha \leq$ 50 and $0 \leq m \leq 6$ where $\Delta\alpha = 0.25$ and $\Delta m = 1$.

Minimum and maximum radii

To perform a Fourier decomposition in $(\ln R, \theta)$ space a minimum and a maximum radius, $R_{\rm min}$ and $R_{\rm max}$, must be defined. These radii need to be chosen with care because they can have a significant effect on the resulting Fourier spectra. Reducing the minimum radius will increase the value of the calculated pitch angle, particularly when a bar is present.

In the analysis of these simulations, the maximum radius is fixed at 6 scale lengths. The definition of the minimum radius aims to exclude the bar from the calculations involving the spiral patterns whilst including as much of the data set as possible. Since the Fourier component $P_{2,0}$ picks out the bar from a galaxy image, the radius at which the power of the $P_{2,0}$ component first drops below 0.005 is used as $R_{\rm min}$. The value of 0.005 is an arbitrary yet consistent one which produces sensible results roughly in line with 'by eye' expectations.

Example power spectra

Figure 5.9 shows images and spectra of a grand design spiral in which both the stars and the gas follow a regular two armed structure. The spectra corroborate this since the stars and the gas are strongly peaked in the $m = 2$ component. The gas also possess significant power in the $m = 4$ and $m = 6$ components which makes the arms narrower in the gas image.

The stellar image of Figure 5.10 is also strongly two armed, and this is verified by the corresponding spectra. Although the gaseous image may seem to be a completely flocculent mess, the spectra show that there is a strong $m = 4$ structure present with a similar, but smaller, pitch angle to the stars. There is also a lot of other structure in the gas image that is not seen in stars.

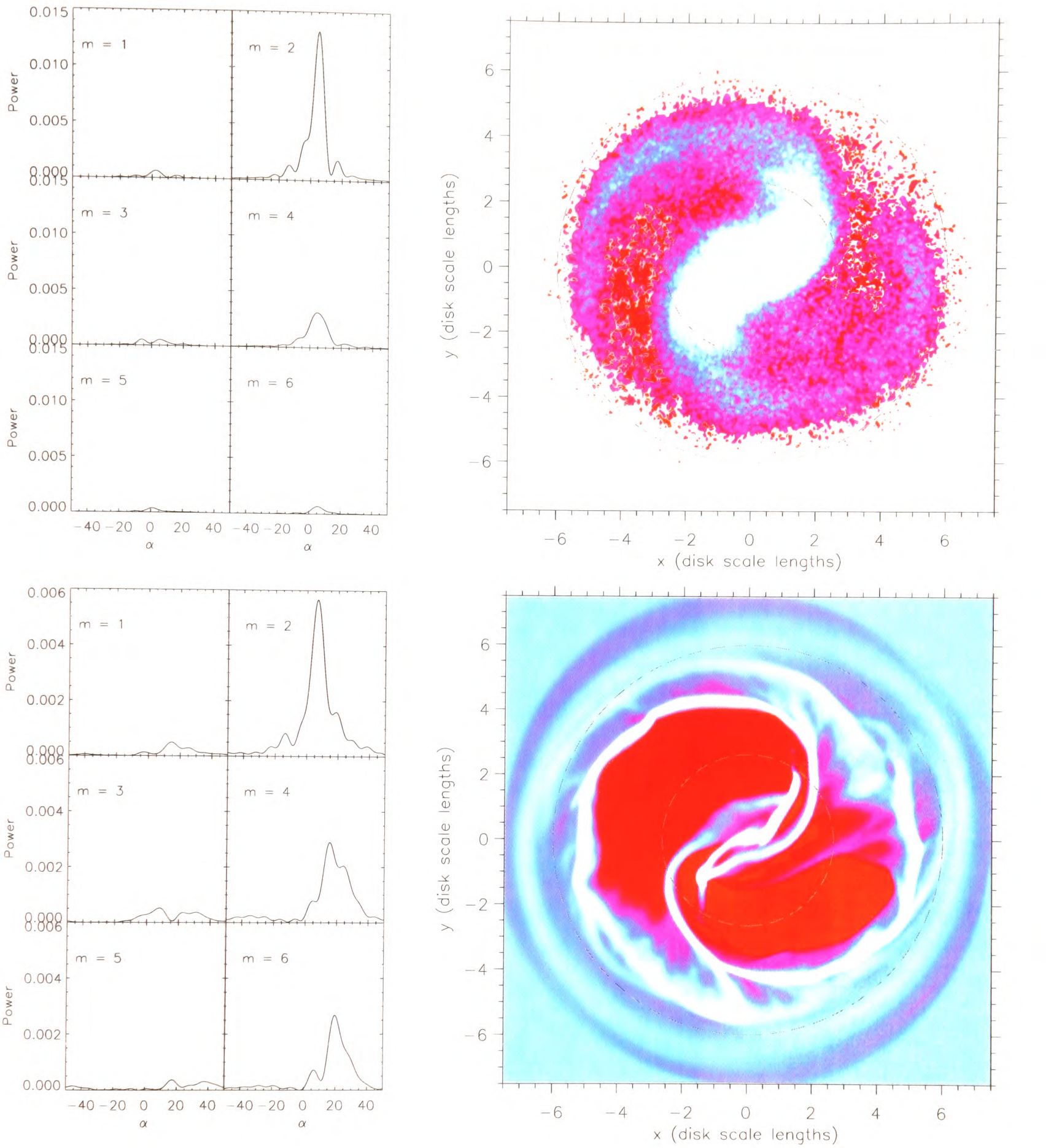


Figure 5.9: Stellar and gaseous density fields and harmonic spectra for the grand design spiral **sta19** and **gas19**. The dashed lines in the galaxy images indicate the minimum and maximum radii used in the Fourier analysis. The flocculence, asymmetry and two armed pitch angles are $\mathcal{F} = 0.51$, $\mathcal{A} = 0.10$ and $\psi = 20.9°$ for the stars and $\mathcal{F} = 0.53$, $\mathcal{A} = 0.13$ and $\psi = 16.5°$ for the gas. The stellar arm-interarm contrast is $\mathcal{C} = 2.48$.

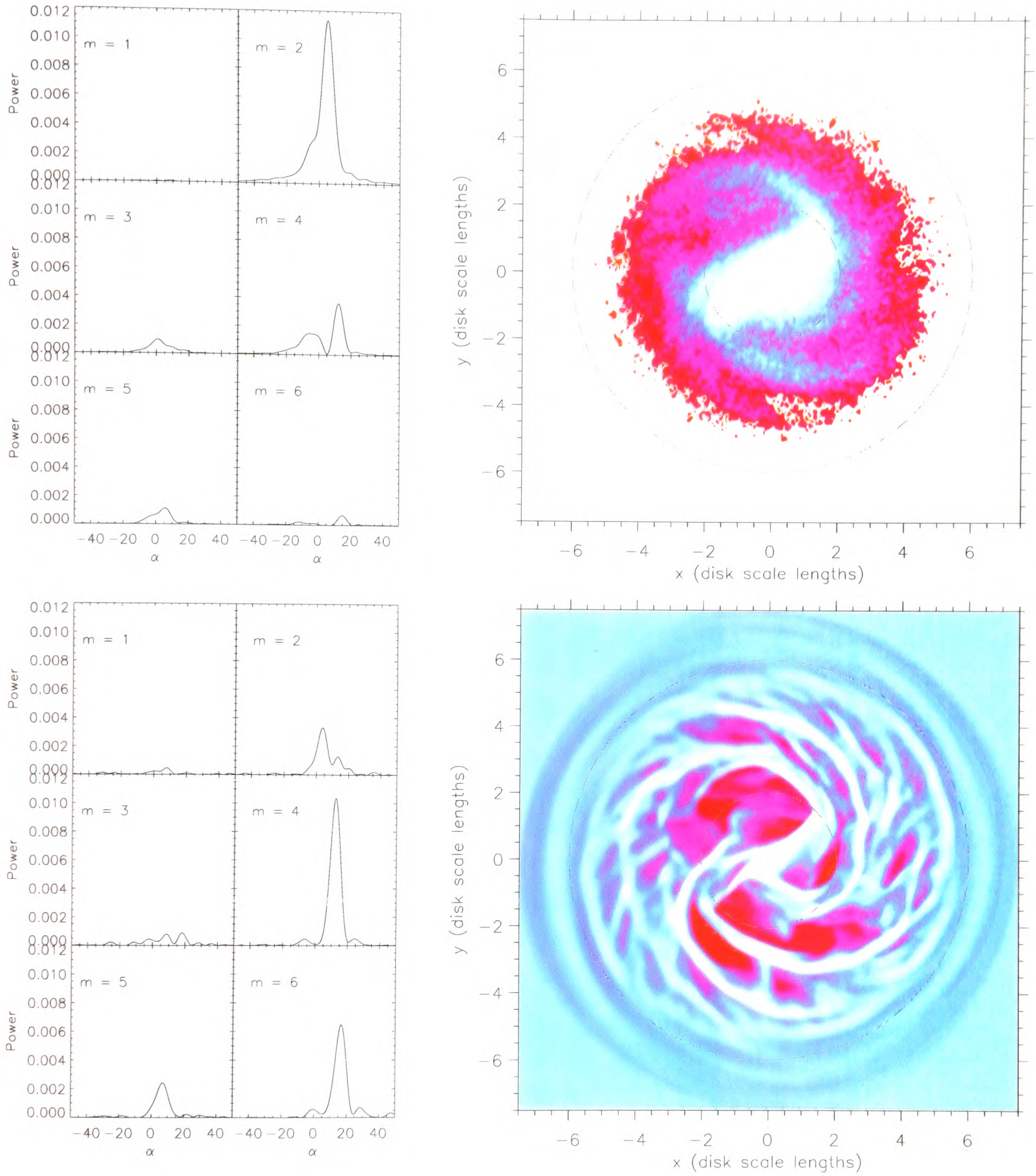


Figure 5.10: Stellar and gaseous density fields and harmonic spectra for the multi armed spiral **sta16** and **gas16**. The dashed lines in the galaxy image indicate the minimum and maximum radii used in the Fourier analysis. The flocculence, asymmetry and pitch angles are $\mathcal{F} = 0.53$, $\mathcal{A} = 0.13$ and $\psi = 20.9°$ (two arms) for the stars and $\mathcal{F} = 0.61$, $\mathcal{A} = 0.22$ and $\psi = 17.7°$ (four arms). The stellar arm-interarm contrast is $\mathcal{C} = 1.71$.

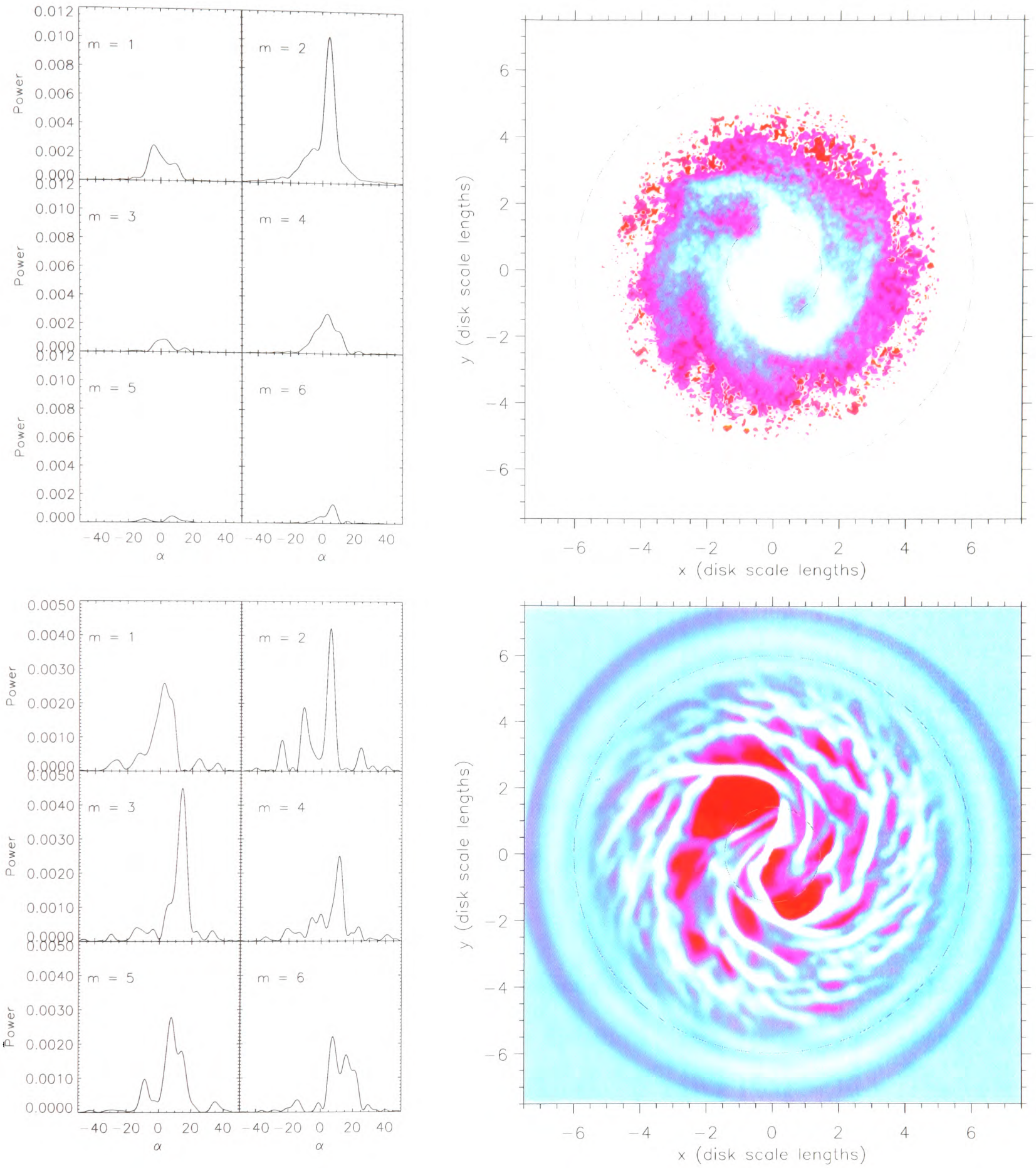


Figure 5.11: Stellar and gaseous density fields and harmonic spectra for the flocculent spiral **sta44** and **gas44**. The dashed lines in the galaxy image indicate the minimum and maximum radii used in the Fourier analysis. The flocculence, asymmetry and pitch angles are $\mathcal{F} = 0.47$, $\mathcal{A} = 0.23$ and $\psi = 24.0°$ (two arms) for the stars and $\mathcal{F} = 0.70$ and $\mathcal{A} = 0.54$ for the gas. The arm-interarm contrast is $\mathcal{C} = 1.58$.

Lastly, the stellar structure of Figure 5.11 is still two armed, as can be seen from its spectra. However, the gaseous image is now completely flocculent and its spectrum is strongly peaked at all values of m. The many interfering spiral waves prevent any single spiral dominating the image.

All of the gaseous images exhibit ring structures near to and outside of $R_{\mathbf{max}}$ due to boundary effects. Rerunning `gas19` with the gas initially in an exponential disk, removes almost all of the ring structures but leaves the rest of the gaseous structure unchanged.

5.3.3 Flocculence

The derivation of a numerical value to characterize flocculence is based on the idea that grand design spirals have one strong peak in their power spectrum whereas flocculent spirals have multiple peaks. Hence, flocculent power spectra have a more even distribution than grand design power spectra. The best measure of the uniformity of a distribution is the entropy, $S = -\sum_{m,\alpha} P_{m,\alpha} \ln P_{m,\alpha}$ (Jaynes 1958). However, before using this to characterize the flocculence of an image, a couple of issues must be considered:

1. If a galaxy image consists solely of a strong two armed spiral, there will obviously be large contributions from the $m = 2$ spectrum. However, the $m = 4$ and 6 spectra may also be strongly peaked, creating a narrower spiral arm, but one which would not be flocculent despite its spectra begin multiply peaked. Hence the $m = 4$ and 6 spectra are ignored in calculating the flocculence.

2. Almost all galaxies have many weak components in their power spectra - the noise. The existence of these components makes the distribution of power less concentrated and tends to blur the distinction between grand design and flocculent images. Hence only components with power $P_{m,\alpha} > 0.25 \max(P_{m,\alpha})$ are considered.

After these considerations, there are still a viable number of components remaining - in the simulations, an average of 123 components are left. Since not all of the components are used in the flocculence calculation, the remaining components are renormalized before calculating the entropy to ensure that $\sum_{m,\alpha} P_{m,\alpha} = 1$.

At this point the entropy S could be transformed into a value of flocculence $0 \geq \mathcal{F} \geq 1$ by dividing by the maximum possible entropy (that due to a flat distribution) $S_{\max} = \ln N_{\max}$. Here $N_{\max} = 1604$ is the maximum number of components that could have been used in the calculation: the total number of components in the $m = 1, 2, 3$ and 5 spectra. This would, however, give a flocculence which depends on $N_{\max}$, making it difficult to compare to other experiments that use a different value of $N_{\max}$. Instead I calculate the standardized value S_{1000}. This is the entropy that the distribution would have if the maximum number of components was 1000:

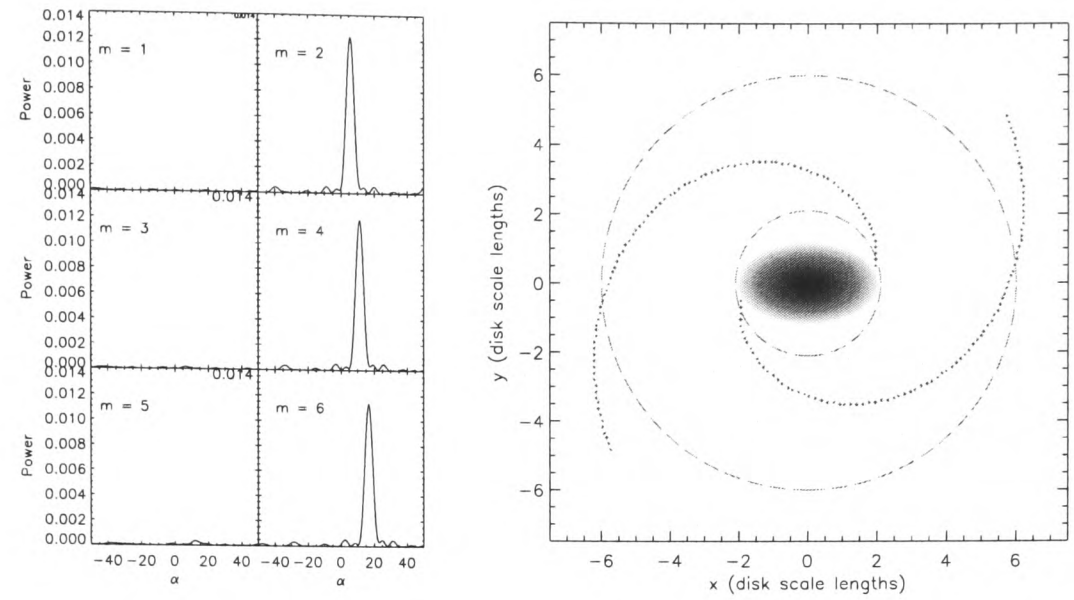
$$S_{1000} = S + \ln 1000 - \ln N_{\max}, \tag{5.10}$$

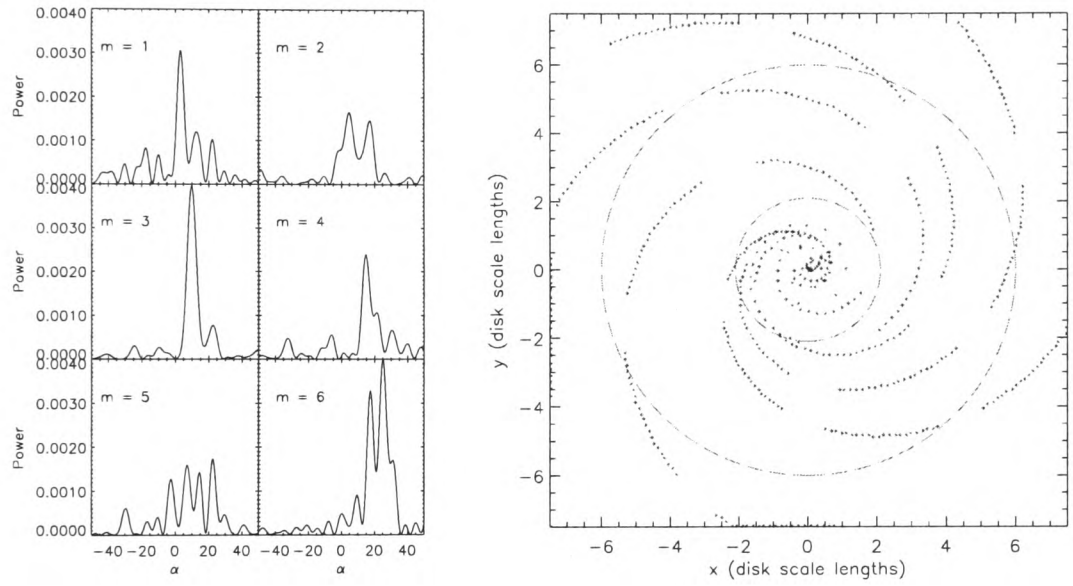
with the condition that $S \geq \ln(N_{\max}/1000)$, which ensures that at least one component would have a non-zero power in the 1000 component distribution and hence that $S_{1000} \geq 0$. This can now be transformed into a value of flocculence $0 \geq \mathcal{F} \geq 1$ by setting $\mathcal{F} = S_{1000}/\ln 1000$.

In the same way that the maximum entropy is due to a completely flat distribution, each image can be associated with a Fourier spectrum in which a number of the components n are equal in power and the rest have zero power. Since the form of the flocculence function is logarithmic rather than linear, an arithmetic increase in $\mathcal{F}$ leads to a geometric increase in n, with $\mathcal{F} = 1$ implying that $n = N_{\max}$. For 1000 components, n doubles when $\mathcal{F}$ increases by 0.1. In this sense, an image is twice as flocculent as another if its flocculence $\mathcal{F}$ is higher by 0.1.

The above algorithm produces values of $\mathcal{F}$ which are roughly in line with 'by eye' expectations for the simulation images and for the test cases shown in Figure 5.12. Values of flocculence derived from the simulations can be compared to these test values of $\mathcal{F} = 0.41$ for a grand design spiral, $\mathcal{F} = 0.74$ for a flocculent galaxy and $\mathcal{F} = 0.80$ for an image composed of pure noise.

Elmegreen and Elmegreen (1982) define a classification system for the regularity of spiral arm structure. Flocculent galaxies have arm classes of 1 to 4 and 'patchy', 'ragged', 'fragmented' or 'fleece-like' spiral arms. Using this definition, `gas44` would certainly be called flocculent. However, many other images with medium to high values of $\mathcal{F}$ would probably be classified as arm class 5 or 6: two symmetric inner arms and irregular outer arms.





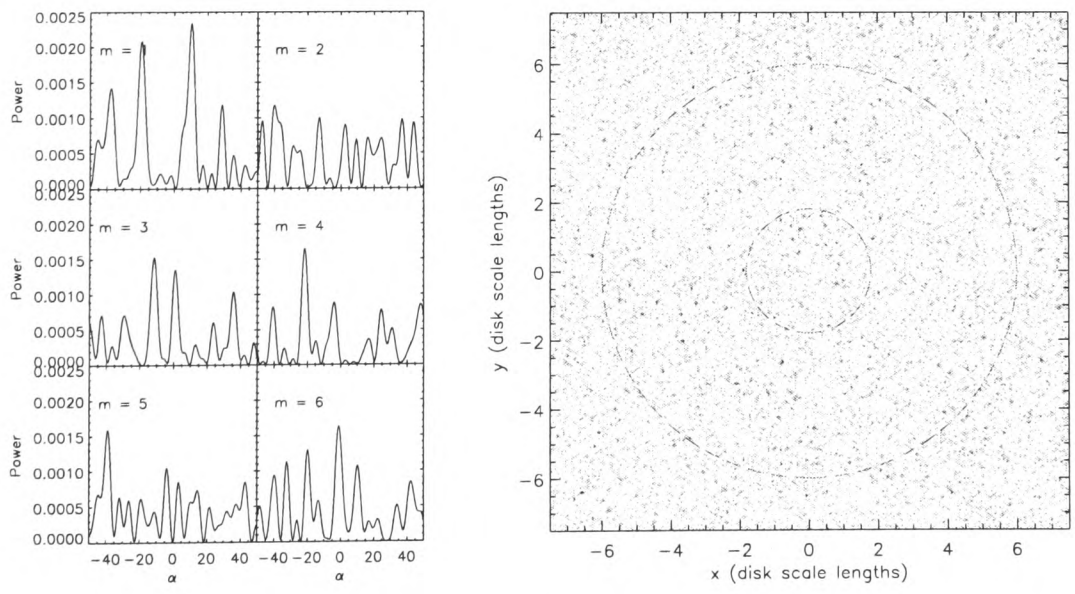


Figure 5.12: Harmonic analysis for tests: grand design and flocculent spirals and noisy image. The dashed lines in the image indicate the minimum and maximum radii used in the Fourier analysis.

5.3.4 Asymmetry

Another useful morphological parameter is the asymmetry $\mathcal{A}$ of an image in the azimuthal direction, defined by Iye et al. (1982) to be the fraction of power in odd-m components,

$$\mathcal{A} = \sum_{m=\text{odd}} P_{m,\alpha} \Big/ \sum_{m} P_{m,\alpha} \qquad 0 \leq \mathcal{A} \leq 1. \tag{5.11}$$

For the grand design test case $\mathcal{A} = 0.06$ whereas for the flocculent test image $\mathcal{A} = 0.49$. Hence, the asymmetry of an image may also be an indication of its flocculence. Asymmetries are calculated for all simulations and they fall roughly within this range.

5.3.5 Radial Fourier analysis

The disadvantage of decomposing an image into logarithmic spirals is that no radial information survives the decomposition. To derive radial variations in morphological characteristics, the data are transformed and interpolated from 250 $\times$ 250 (x, y) coordinates to 256 $\times$ 128 (R, θ) coordinates and one dimensional Fourier transforms are performed on the data at each radius. The result is the radial variation of the amplitude and phase of the $m = 1 \rightarrow 6$ Fourier components. This information is used in Section 5.3.6 to calculate the arm-interarm contrast and was used in Section 5.3.2 to calculate the minimum radius.

5.3.6 Arm-interarm contrast

The contrast between the spiral arms and the space in between the arms is not easy to define because it is difficult to determine where the centre of the arm is and what to take as the interarm region. However, it is worth persevering since high contrast stellar spirals are good candidates for producing grand design gaseous spirals whereas low contrast stellar spirals often result in flocculent gaseous spirals. Following Rix (1993), the simulation images are subjected to radial Fourier analysis and the arm-interarm contrast $\mathcal{C}$ is defined as the ratio of the maximum to the minimum of the even m components. The even components are

$$I_R(\theta) = |a_{0,R}| + \sum_{m=2,4,6,8} |a_{m,R}| \cos(m\theta + \phi_{m,R}), \tag{5.12}$$

where $a_{m,R}$ is the mth Fourier component at radius R and $\phi_{m,R}$ is its phase. At each radius,

$$\mathcal{C}_R = \max(I_R(\theta))/\min(I_R(\theta)),$$

and averaged over the disk,

$$\mathcal{C} = \text{mean}(\mathcal{C}_R) \qquad R_{\text{min}} < R < 5. \qquad (5.13)$$

The resulting values of $\mathcal{C}$ are generally between 1.5 and 3.0. Figure 5.13 shows representative radial variations of arm-interarm contrasts in which $\mathcal{C}_{19} = 2.5$ and $\mathcal{C}_{16} = 1.7$.

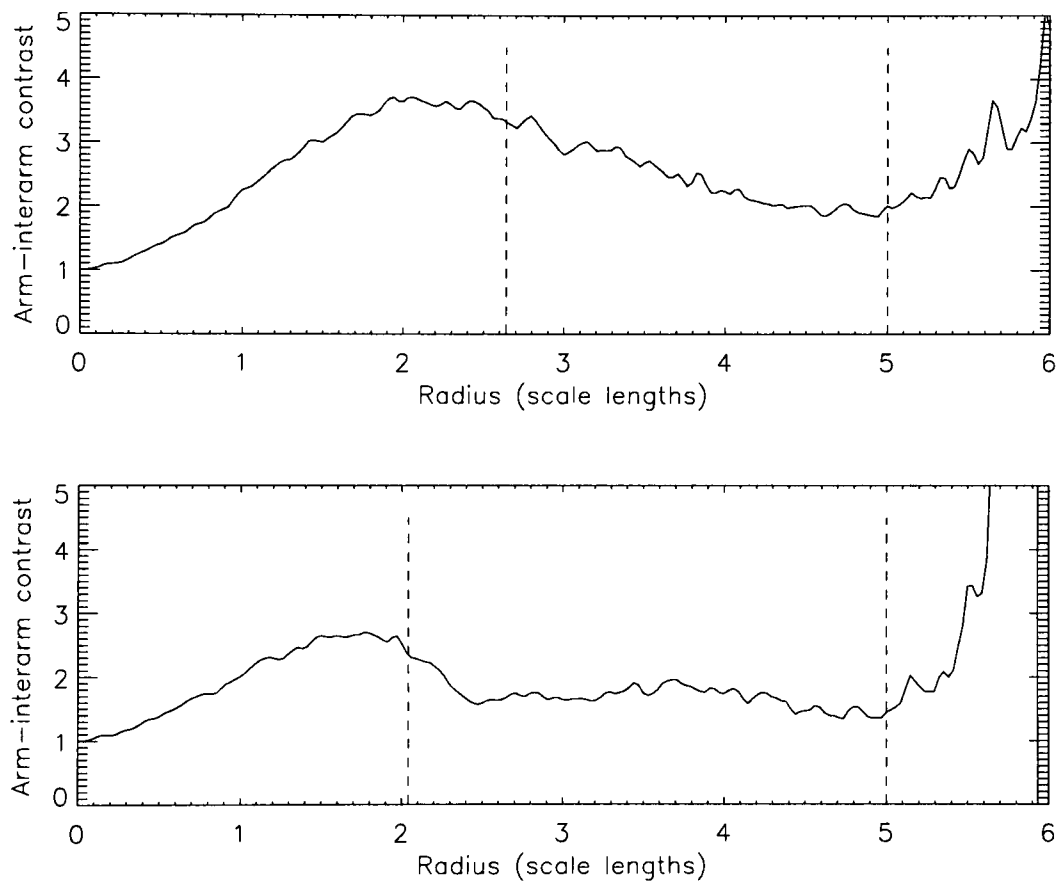


Figure 5.13: Radial variation of arm-interarm contrast $\mathcal{C}$ for `sta19` (top) and `sta16` (bottom). The dashed lines show the minimum radii $R_{\text{min}} = 2.64$ and $R_{\text{min}} = 2.04$ respectively and the maximum radius $R_{\text{max}} = 5.00$. The arm-interarm contrasts for the two simulations are $\mathcal{C}_{19} = 2.5$ and $\mathcal{C}_{16} = 1.7$.

5.3.7 Monitoring $Q(t)$

Although Q_{min} is one of the input parameters to the models, the way in which Q varies with time is not known at the outset of a simulation. It is important to understand the evolution of $Q(t)$ as

it is intimately linked to the formation and maintenance of spiral structure. It might be assumed that the value of $Q_{\rm min}$ at the start of the simulation correlates positively with the mean value of Q at the end of the run. Not so. Figure 5.14 shows $Q(t)$ for two simulations and is illustrative of the situation in all of the simulations: higher values of $Q_{\rm min}$ lead to $Q(t)$ which rises slowly whereas lower values of $Q_{\rm min}$ lead to a rapidly rising and fluctuating $Q(t)$.

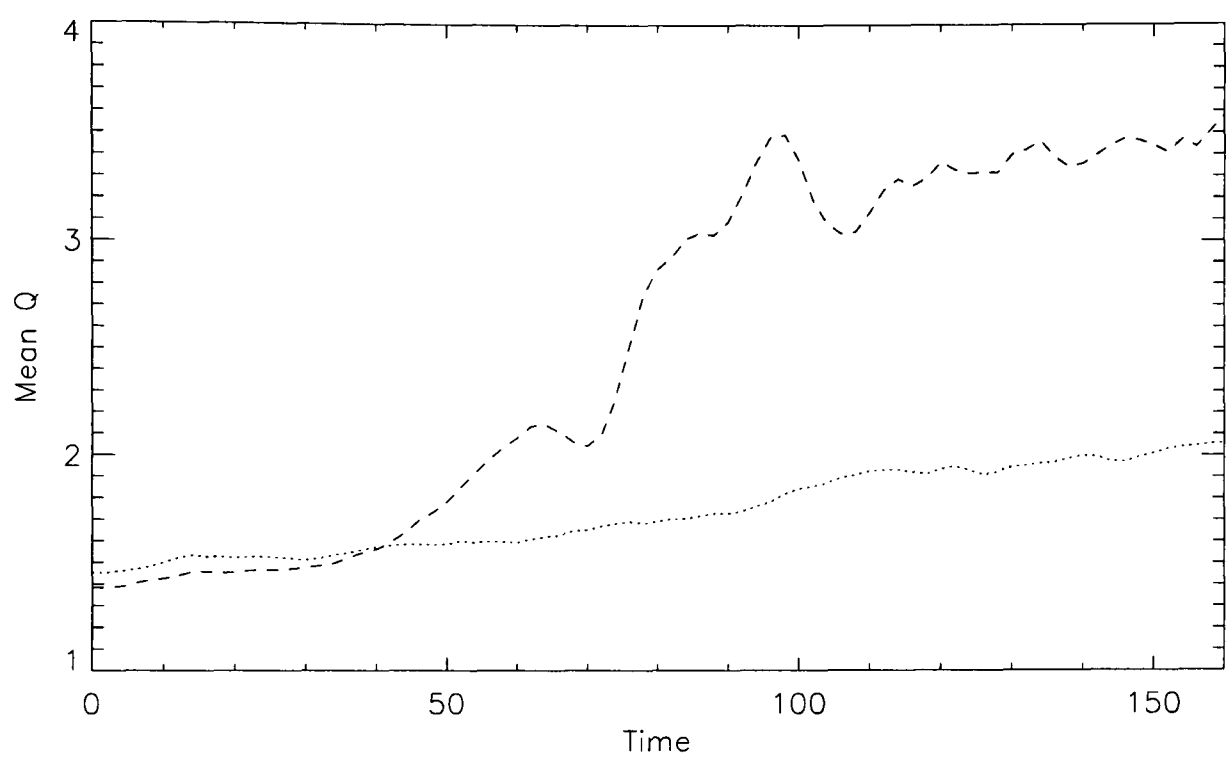


Figure 5.14: Evolution of $Q(t)$. The long dashed line shows run `sta19` which begins with a low Q value but ends up high. The short dashed line shows `sta16` which starts higher but rises only gradually.

5.3.8 Pattern speed and resonances

As mentioned in Section 5.3.2, the Fourier component $P_{2,0}$ picks up any bar lurking within a galaxy image. The time averaged pattern speed is calculated as the average rate of change of the phase of the bar ϕ_B,

$$\Omega_p = [\phi_B(t_{\rm max}) - \phi_B(t_{\rm min})]/(t_{\rm max} - t_{\rm min}), \tag{5.14}$$

where $t_{\rm max}$ and $t_{\rm min}$ are the times of the maximum and minimum time slices used.

The values of the pattern speed Ω_p and the bar length $R_{\rm min}$ as defined above, are used to calculate the rotation ratio, $\mathcal{R} = R_{\rm co}/R_{\rm min}$, where $R_{\rm co}$ is the corotation radius. All of the models for

which this has been calculated fall into the generally accepted region for fast bars of $\mathcal{R} = 1.2 \pm 0.2$ except for $\mathcal{R}_{16} = 1.5$. However, there are large errors on the calculations of $\mathcal{R}$, mostly due to the uncertainty in the derivation of $R_{\rm min}$.

5.4 Simulation results

Sixteen simulations have been run and their positions within the parameter space defined by $1.12 < Q_{\rm min} < 1.26$ and $0.3 < f_d < 0.5$ are shown in Figure 5.15. Outside of this parameter space it was not possible to generate two armed stellar spirals. The dashed line in Figure 5.15 shows the additional constraint imposed by demanding that the softening length s, measured in cell lengths, satisfy $0.5 < s < 2.0$. Other parameters have a changeable canonical value, listed in Table 5.1. Data describing the set up and results of and representative images during each run are included in Appendix C.

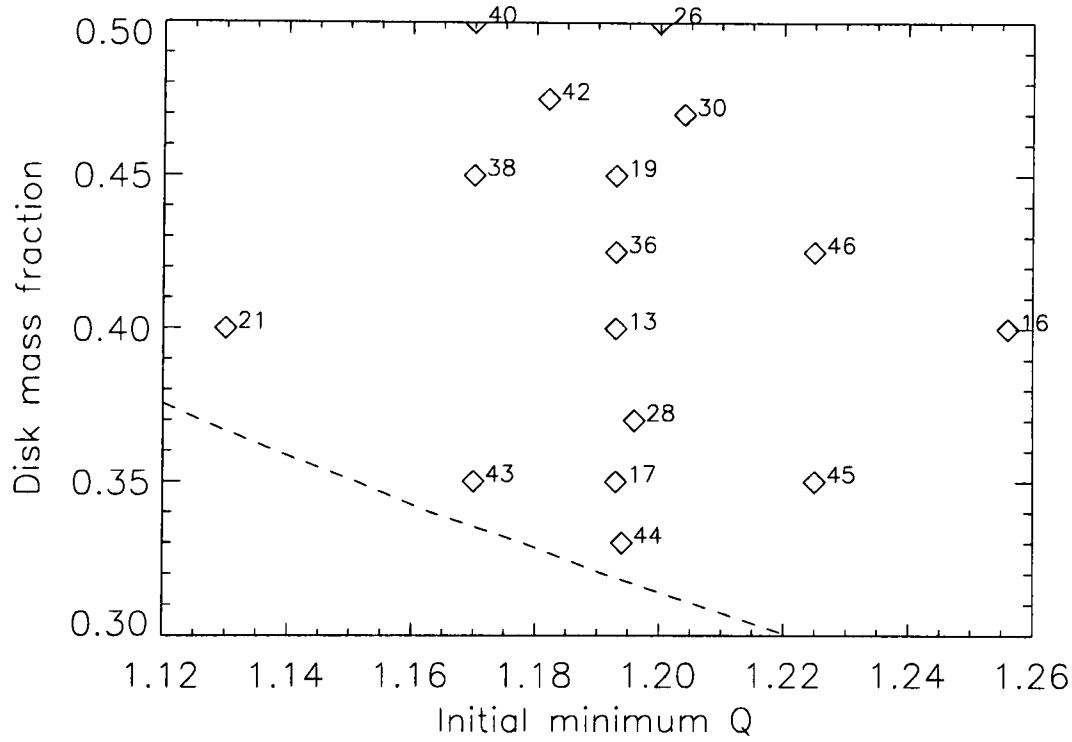


Figure 5.15: Position of all simulations within the parameter space. The dashed line marks the line of softening length $s = 0.5$ grid cells.

5.4.1 Computational checks

Two checks ensure that each simulation behaves reasonably sensibly. The number of particles lost during the simulation $N_{\rm lost}$ is always less than 0.5% of the total (1000) in all models. The position

Table 5.1: Values of canonical parameters in dual morphology models

Number of particles	N	2×10^5
Grid size	$I \times I$	250×250
Simulation run time	t_{end}	160
Time between data dumps	t_{dump}	2
Courant number	C	0.75
N-body time coefficient	φ	0.6
Galaxy radius	R_G	7.5
Sound speed†	c	0.025
Gas mass	M_g	0.01
Cut off radius for particle distribution	R_{cut}	5.0
Rotation curve slope at four scale radii	g_4	0.0
Gas recycling coefficient	α	0.0

† Cowie (1980)

of the centre of mass of the stellar particles with respect to the grid centre is drifts by less than 0.18 R_d. The values for each simulation are detailed in Appendix C.

Of the five runs in which no particles left the grid, angular momentum was conserved to better than 0.002%, and the ratio of the kinetic energy to the virial of Clausius[2] was unchanged from its initial value of 0.5 to better than 1%.

5.4.2 Time evolution of simulations

Figure 5.16 shows representative images and spectra at increasing times for `sta19` and `gas19`. After a couple of bar rotation periods, the stars settle into a two armed spiral pattern. A similar pattern can be seen in the gas spectra, in addition to strong $m = 3$ and $m = 5$ patterns left over from the initial formation process. The stellar pattern does not change much over the next couple of rotation periods, while the higher order gaseous harmonics gradually disappear to leave a grand design spiral in the stars and the gas. At later times, the stars become dynamically hotter causing the spirals to fade somewhat and resulting in a more flocculent stellar distribution. However, the gas has a stronger memory than the stars and retains its grand design nature for longer.

This overall scenario is repeated in most simulations. However, the details vary, in particular

[2]For a system of i particles of mass m at positions $\mathbf{r}_i$ and potential ϕ_i, the virial of Clausius is $m \sum_i \mathbf{r}_i \cdot \nabla\phi_i$.

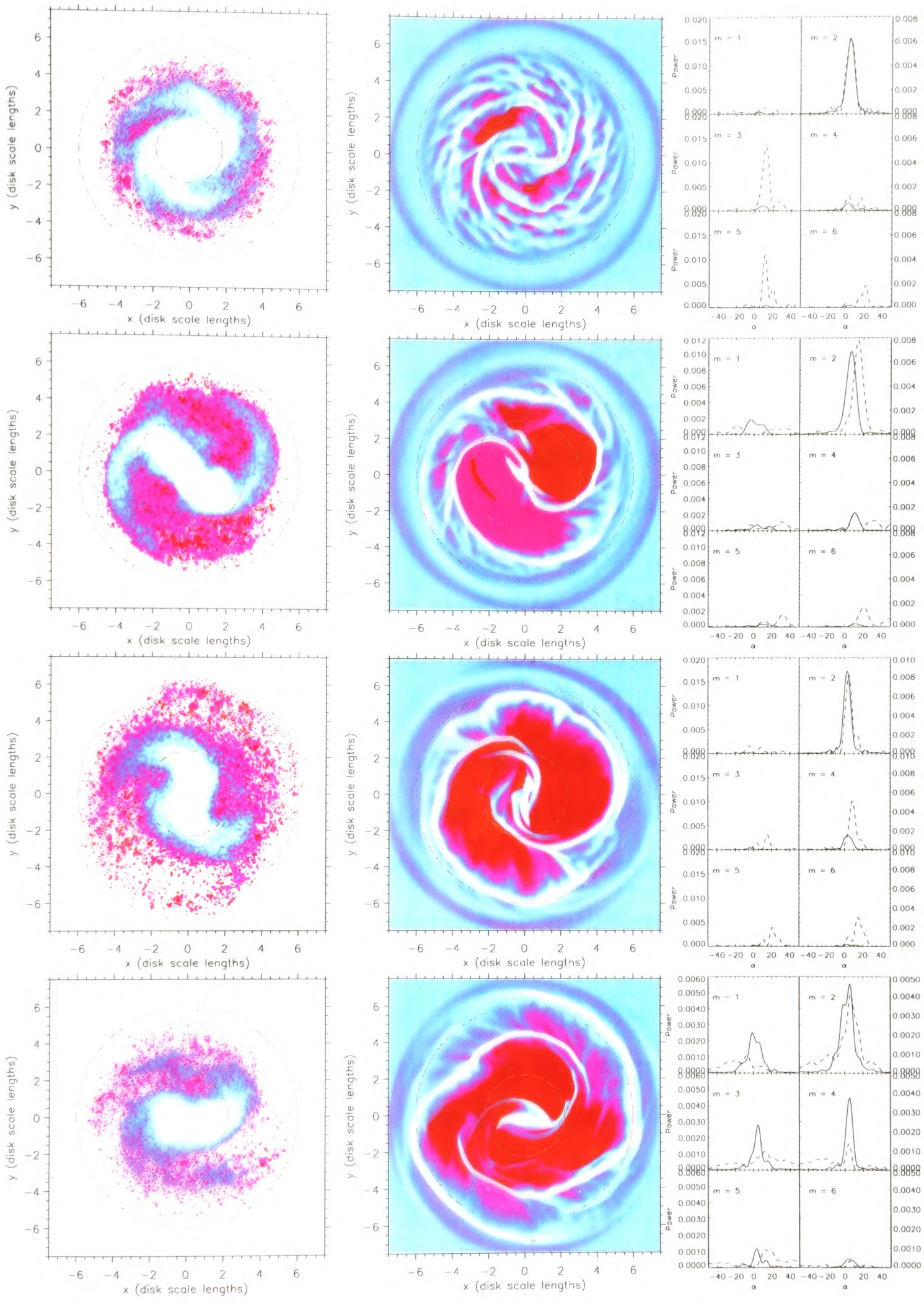


Figure 5.16: Density fields and Fourier spectra for **sta19** and **gas19** at 44, 80, 94 and 110 time units or 2.0, 3.7, 4.4 and 5.1 bar rotation periods. In the Fourier spectra, solid lines indicate the stars and dashed lines the gas.

how long it takes for the grand design stellar spirals to form and how soon afterwards the gas spirals lose their initial flocculence. There are also differences in how far the gas moves along the path between flocculence and grand design structure. In some simulations, multi armed gaseous structures can be observed at points along this path.

5.4.3 Do these models reproduce the observations?

i. Stellar and gaseous images can be decoupled Numerous examples are shown of images and spectra to establish the truth of this proposition. Every simulation has differing gaseous and stellar spectra, some significantly so, and these morphological differences are borne out in simulation images. Hence, some of the dual morphologies seen in many real galaxies could simply be due to the difference in the dynamics of the stellar and gaseous components of those galaxies.

ii. Gaseous pitch angles are either similar to or tighter than stellar images For simulations in which the gaseous images are two armed, the time averaged gaseous pitch angles are almost always smaller than those of the stellar images. To illustrate this, Figure 5.17 shows the time averaged gaseous and stellar pitch angles for all simulations. This is in contrast to Block and Puerari (1999) who find examples of Sc galaxies with tight ($\psi < 20°$) K' band spiral arms. They do not, however, quote B band pitch angles which makes comparison difficult, and any calculated angles would result from a combination of both inherent dynamics and dust contamination in the images.

iii. Bars are more prominent in stellar than in gaseous images From the images, the stellar bar is obviously much more prominent than the gaseous bar. This is to be expected since the shocked gas loses angular momentum and falls into the centre of the stellar bar region. Hence, over the majority of the bar region, the gas is low density, as observed in real galaxies.

iv. Stellar images are always grand design, whereas gas images can be flocculent For almost every case in Figure 5.17 the stellar image has a lower value of the flocculence $\mathcal{F}$ than the gaseous image. No stellar image has $\mathcal{F} > 0.63$ whereas the gas images have $\mathcal{F}$ up to 0.68 or

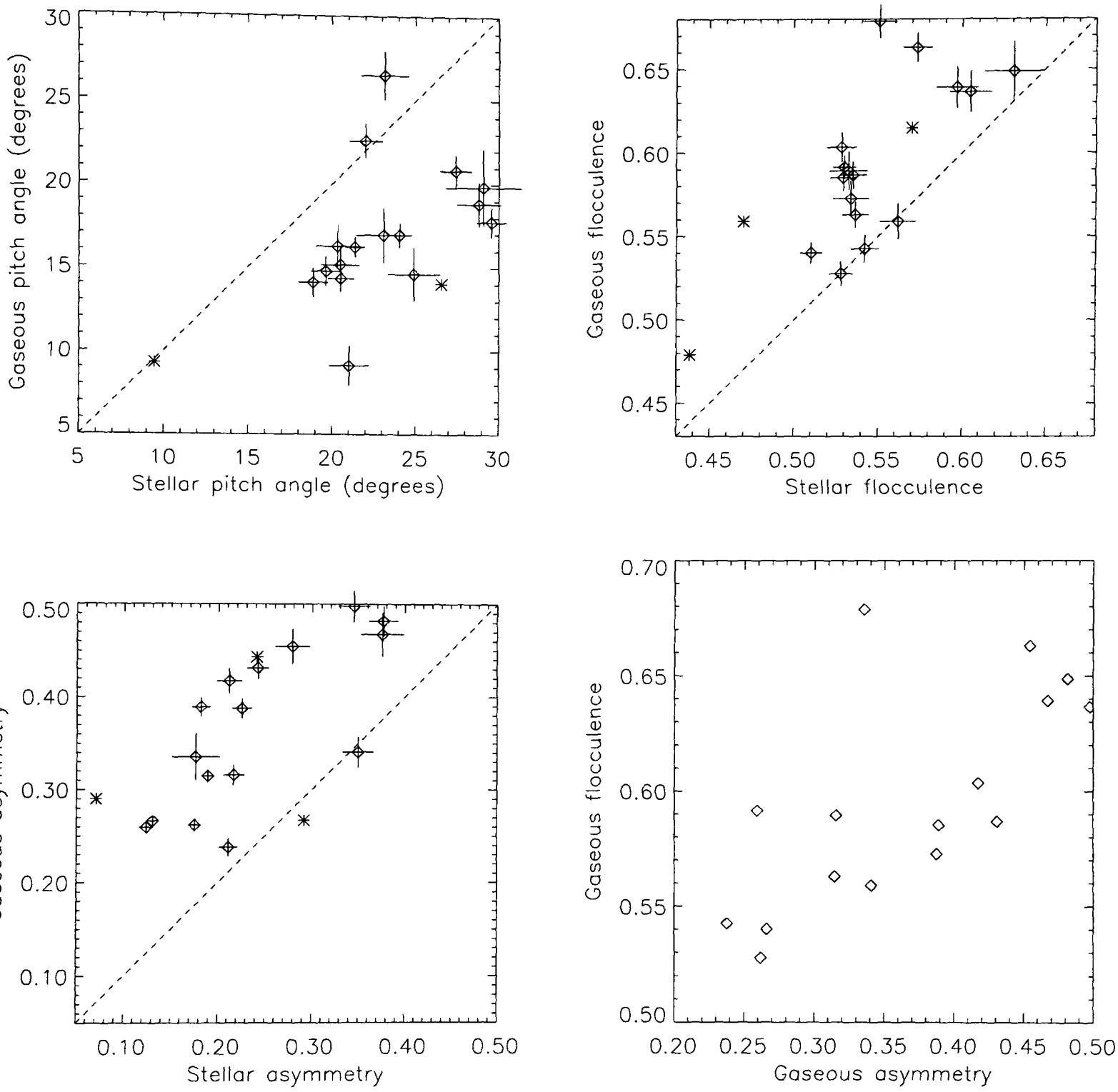


Figure 5.17: The first three panels show the time averaged gaseous and stellar pitch angle ψ, flocculence $\mathcal{F}$ and asymmetry $\mathcal{A}$ for all simulations. The dashed line is the line of equal pitch angle, flocculence or asymmetry. Error bars show parameters' values if R_{min} was changed by one increment. Asterisks indicate values for real galaxy images. The final (bottom right) panel shows the correlation between the gaseous flocculence and asymmetry parameters.

1.4 times as flocculent in the sense of Section 5.3.3. This is borne out in the images, in which almost all of the stellar models have an obvious two armed grand design structure. Although few of the gaseous models are flocculent in the sense of Elmegreen and Elmegreen (1982) i.e. many small scale arms, and many models possess an underlying two armed spiral, they also exhibit extra structures that are not seen in the relevant stellar model. These extra structures take the form of fronds leaving the main spiral or bridges between spiral arms. Sometimes, extra partial spiral arms can also be seen. And in some instances, grand design stellar models do produce gaseous images that are truly multi armed or flocculent. These cases are rare but do arise.

v. Stellar images are more symmetric than gaseous images Using the asymmetry number $\mathcal{A}$, it can be seen in Figure 5.17 that the gaseous images are more asymmetric than the stellar images in almost every case. For the stellar images $0.12 < \mathcal{A} < 0.38$ whereas in the gaseous images $0.23 < \mathcal{A} < 0.50$. It is interesting to note that the asymmetry number roughly correlates with the flocculence of the images, particularly for the gaseous images (as shown in Figure 5.17). It may be useful to repeat this analysis using the Conselice number (Conselice 1997).

vi. Stellar images are one or two armed, gaseous images can be multi armed This is seen to be true when looking at the images, but it is also apparent when comparing the power of the strongest component from each m spectrum for every time slice which has been analyzed. Almost all of the stellar spectra are dominated by the strongest $m = 2$ component - the strongest component is $m \neq 2$ in only 5 out of the 311 time slices analyzed. Many gas spectra are also $m = 2$ dominated but it is less common and the degree of dominance is generally much less - there are 32 gaseous time slices in which the strongest component is $m \neq 2$.

However the only model which can truly be called multi armed is `gas16`. For the majority of the time that the stellar image exhibits a two armed spiral, the gas spectrum is dominated by an $m = 4$ component. Interestingly, this model had the highest value of $\mathcal{R}$, the highest value of $Q_{\rm min}$ and the lowest value of the time averaged $\bar{Q}(t)$. It has a similar morphology to our Galaxy as deduced from K band and 240 μm observations by Drimmel and Spergel (2001) i.e. a two armed stellar spiral and a four armed gaseous spiral. SPH simulations of Englmaier and Gerhard (1999)

and Fux (1999) have also reproduced multiple arm gaseous spirals for our Galaxy.

vii. Decoupling correlates with arm-interarm contrast In order to verify this, it is necessary to define what is meant by decoupling. In the context of the observations, it is used to describe the independence of the morphology in the B band from that in the K' band. Using the above parameters, this could mean $\mathcal{F}_{\rm gas} - \mathcal{F}_{\rm sta}$ or $\mathcal{A}_{\rm gas} - \mathcal{A}_{\rm sta}$. Neither of these choices produces a significant correlation with the time averaged arm-interarm contrast $\mathcal{C}$, (correlation coefficients of -0.54 and 0.35), possibly because the errors on the gaseous and stellar parameters become overwhelming when the subtraction is performed.

If the gaseous flocculence or asymmetry are considered, there is a significant correlation with the arm-interarm contrast, (see Figure 5.18). The correlation coefficient for gaseous flocculence is -0.77 and -0.62 for gaseous asymmetry. Since the observations and the simulations show that K' band and stellar images are generally not flocculent, there is a good case for using the flocculence of the gas, and hence the arm-interarm contrast, as a guide to the amount of decoupling.

It is reasonable to assume that the stellar arm-interarm contrast is not only an indicator of decoupling between stellar and gaseous morphologies but also its cause. This is particularly true of the non self gravitating case, where the driver of morphological structure within the gas is the non-axisymmetric part of the stellar potential. Hence when the arm-interam contrast is low, the noise in the stellar component can make a significant contribution to the driving force, leading to many low amplitude features in the gas and an overall flocculent structure. But when the contrast is high, the stellar spiral dominates the driving force, creating a single high amplitude wave in the gas and hence a grand design spiral. This can be seen in the *gaseous* arm-interarm contrasts: $\mathcal{C}_{16} = 3.08$ for the flocculent simulation and $\mathcal{C}_{19} = 6.13$ for the grand design simulation.

This relationship has also been noticed in real galaxies. Four galaxies (including M63 in Figure 5.3), that are obviously flocculent in the optical, were observed by Thornley (1996) in the infrared. He found good evidence for low amplitude grand design structures with K' band arm-interarm contrasts of just 1.1 to 1.4. In real galaxies, however, a combination of supernovae and differential rotation can give rise to self-propagating star formation, causing flocculence in optical

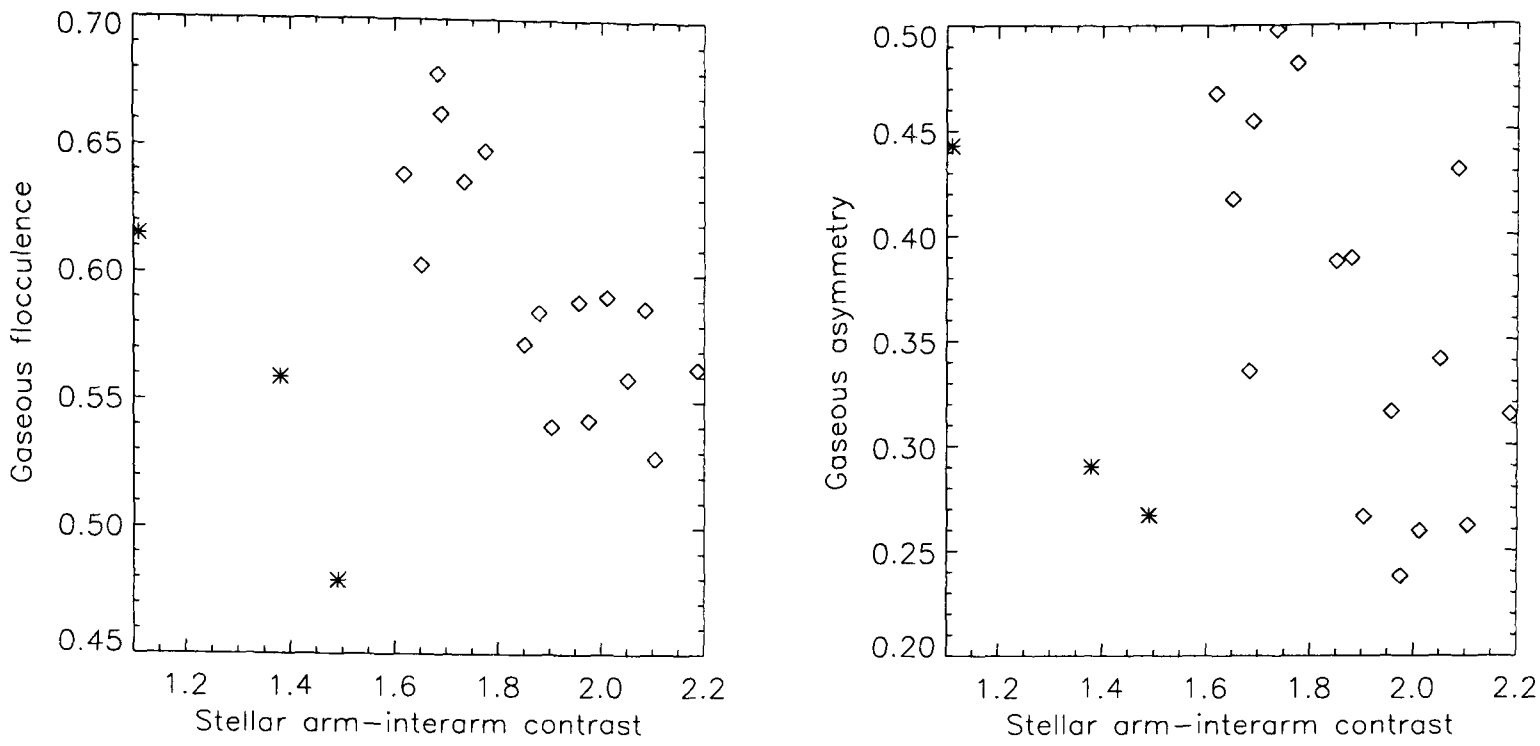


Figure 5.18: Correlation of time averaged arm-interarm contrast $\mathcal{C}$ with gaseous flocculence $\mathcal{F}_{\mathbf{gas}}$ (left panel) and gaseous asymmetry $\mathcal{A}_{\mathbf{gas}}$ (right panel). Asterisks indicate values for real galaxy images.

images irrespective of the underlying potential (Mueller and Arnett 1976; Gerola and Seiden 1978).

5.4.4 Can these models predict galactic structure parameters?

The modal theory of spiral structure predicts that galaxy morphology should be directly related to the underlying structure of the galaxy (Bertin et al. 1989), as shown in Figure 5.4. I now examine whether this is true in my models.

i. Galaxy disk mass correlates with bar size Figure 5.19 shows the strong relationship between the disk mass fraction at four scale lengths f_d and the size of the bar as defined by the parameter $R_{\min}$. The correlation coefficient is 0.80.

ii. Stellar Q correlates with flocculence There are rather weak correlations between the time averaged $\bar{Q}(t)$ for the stars and the gaseous asymmetry, stellar flocculence and stellar asymmetry. The correlation coefficients are -0.26, -0.43 and -0.20 respectively. However, $\bar{Q}(t)$ and gaseous flocculence correlate with a coefficient of -0.62. This is in the opposite direction to the correlation in Figure 5.4 and in Elmegreen and Thomasson (1993) who state that flocculent spiral

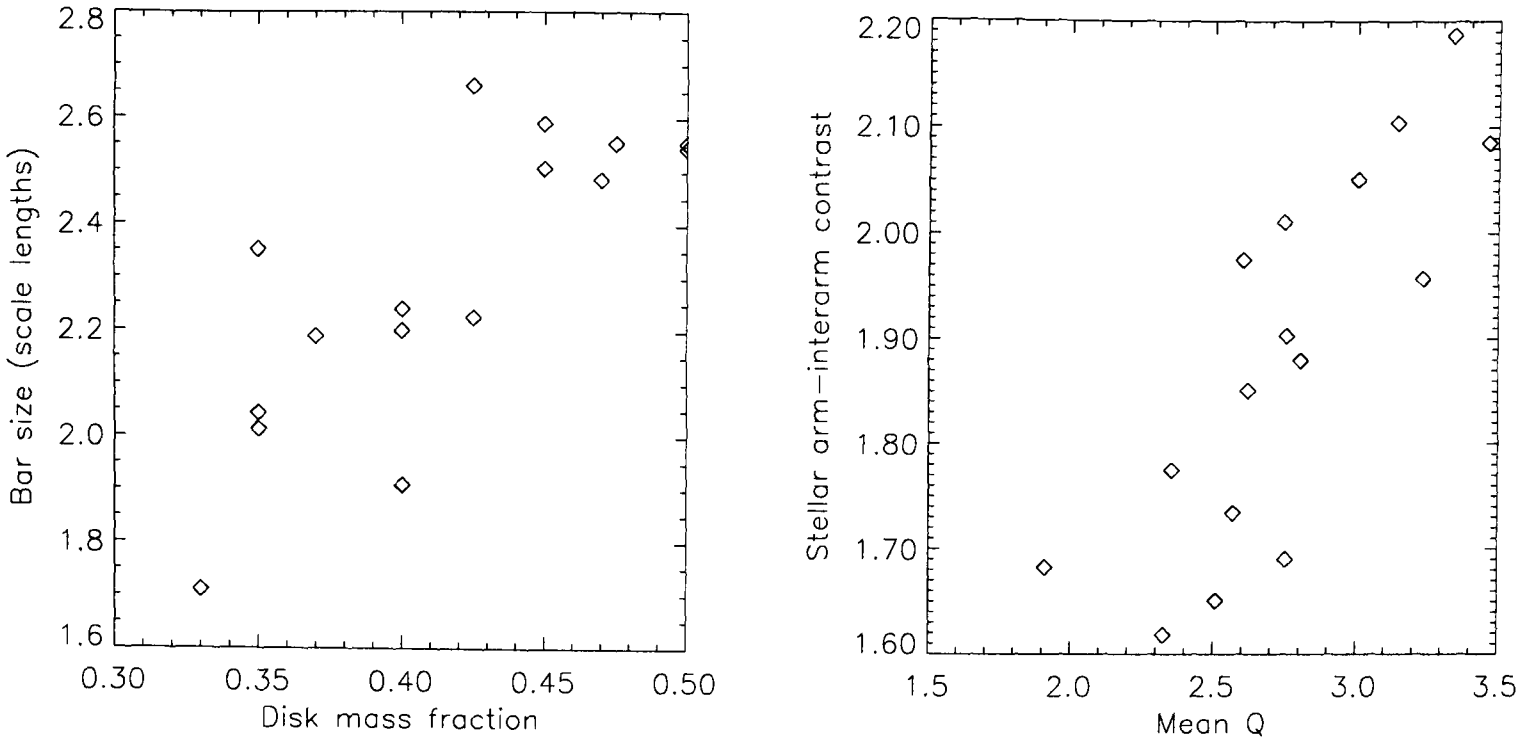


Figure 5.19: Left panel: correlation of disk mass fraction at four scale lengths f_d with the time averaged bar size as defined by the parameter $R_{\rm min}$. Right panel: correlation of the time averaged mean value of Q with time averaged stellar arm-interarm contrast $\mathcal{C}$.

structure arises when $Q > 2.5$. My results are probably due to the fact that, unlike Elmegreen and Thomasson (1993), I do not have any stellar cooling algorithms in my code. It would be interesting to compare my results with S0 galaxies, since they have very little gas and hence little chance to dynamically cool their stars.

$\bar{Q}(t)$ also correlates strongly with the time averaged stellar arm-interarm contrast (coefficient of 0.81) as shown in Figure 5.19. This probably results from the scattering of stellar particles from the spiral arms, with higher contrast leading to greater scattering and hence dynamically hotter disks. Therefore the correlation between $\bar{Q}(t)$ and $\mathcal{F}_{\rm gas}$ is a result of the correlation between gas flocculence and arm-interarm contrast.

iii. Gas mass correlates with pitch angle Since the gas is effectively massless in these models, it is not possible to make any statements concerning this hypothesis.

5.4.5 Comparison with observations

It is important to compare the results of the simulations with observations of real galaxies to ensure that they are not just interesting computer simulations but that any conclusions drawn

from them are relevant to the real universe.

Arm-interarm contrast The magnitude of the simulated time averaged arm-interarm contrasts are similar to those found in real galaxies. In their sample of 18 spiral galaxies, Rix and Zaritsky (1995) find that $\mathcal{C}$ has an average value of ~ 2, with large variations above and below this value. Rix (1993) finds that $1.8 < \mathcal{C} < 3.0$ for the grand design galaxy M51. The arm-interarm contrast is found to be somewhat lower for the flocculent galaxies studied by Thornley (1996) who concludes that $1.1 < \mathcal{C} < 1.4$. My models fall towards the top end of these ranges.

Pitch angles Data from Kennicutt (1981) shows that the gaseous stellar pitch angles in Sc galaxies occupy a similar range to that of the models, $10° < \psi_{\rm gas} < 30°$.

Rotation curve Figure 5.7 has already shown that the models have similar mass distributions to that of both disk and halo dominated Sc galaxies.

Harmonic analysis of real galaxies Many of the Figures presenting correlations amongst simulation parameters include observations of real galaxies in the K' and B bands. The parameters associated with these observations have been derived by subjecting the data to a similar procedure to that used on the models. The galaxies used were NGC 2997, NGC 3223 and NGC 5085 and the data were kindly donated by David Block. These data show similar relationships for pitch angle, asymmetry and flocculence (Figure 5.17) to that of the models when the K' band and B band data are used in place of N-body and hydrodynamical results. The arm-interarm contrasts for the real galaxies show a correlation with flocculence similar to that of the models (Figure 5.18) but with the arm-interarm contrasts offset by $\Delta\mathcal{C} \sim 0.6$. The three galaxies are, however, correctly ranked in order of B band flocculence.

5.5 Discussion

The simple N-body and hydrodynamical simulations described in this Chapter have reproduced many of the observed differences between B and K' band images of galaxies. At first glance this

seems slightly bizarre, particularly since the simulations neglect the self gravity of the gas. The motion of the gas should just follow that of the underlying stellar potential. But, due to the long range nature of gravity, the potential generally does not follow the density which generates it (as can be seen in Figure 5.20). As such, there are questions to be addressed: why is the pitch angle of the gas component less than that of the stars? Why are images dominated by gas more flocculent than those dominated by stars? And can predictions be made about the intrinsic characteristics of spiral galaxies from their morphologies?

5.5.1 The pitch angles of gaseous and stellar spirals

Before discussing the difference between the pitch angles in the gaseous and stellar components, the very existence of the gaseous spirals must be accounted for. Englmaier and Shlosman (2000) show that the dispersion relation for non self gravitating gas linearly forced by a stellar potential only admits gaseous waves outside the Outer Lindblad Resonance and inside any Inner Lindblad Resonance. The simulations of Sanders and Tubbs (1980) and Lindblad et al. (1996) reached similar conclusions: they found that a barred stellar potential was not enough to generate gaseous spirals outside of corotation. Only when Lindblad et al. (1996) added a spiral potential were they able to reproduce the spirals of NGC 1365, stretching beyond the OLR. As illustrated in Sellwood and Sparke (1988), the field of a finite bar decays quickly with radius, exerting little influence on the outer regions of the disk, which may even have a different pattern speed. Hence, my gaseous spirals are probably forced by the stellar *spirals* rather than the bars.

In order to explain the secular evolution of spiral galaxies, Zhang (1996) presents a derivation of a phase shifted, self consistent, stellar density and potential pair. She showed that such phase shifts create a torque applied by the spiral arms on the stars and lead to the disk matter inside corotation accreting inwards and that outside corotation excreting outwards. Figure 5.20, showing the maxima of the $m = 2$ Fourier components of the gas density, stellar density and stellar potential, indicates that this type of effect may be at work in my simulations. The gas follows the displaced stellar potential rather than the stellar density, particularly outside of the bar region. The gas maxima turn from the bar region into spirals much later than the stars, possibly due to

x_1 orbits inside the bar region, and this results in a lower gaseous pitch angle.

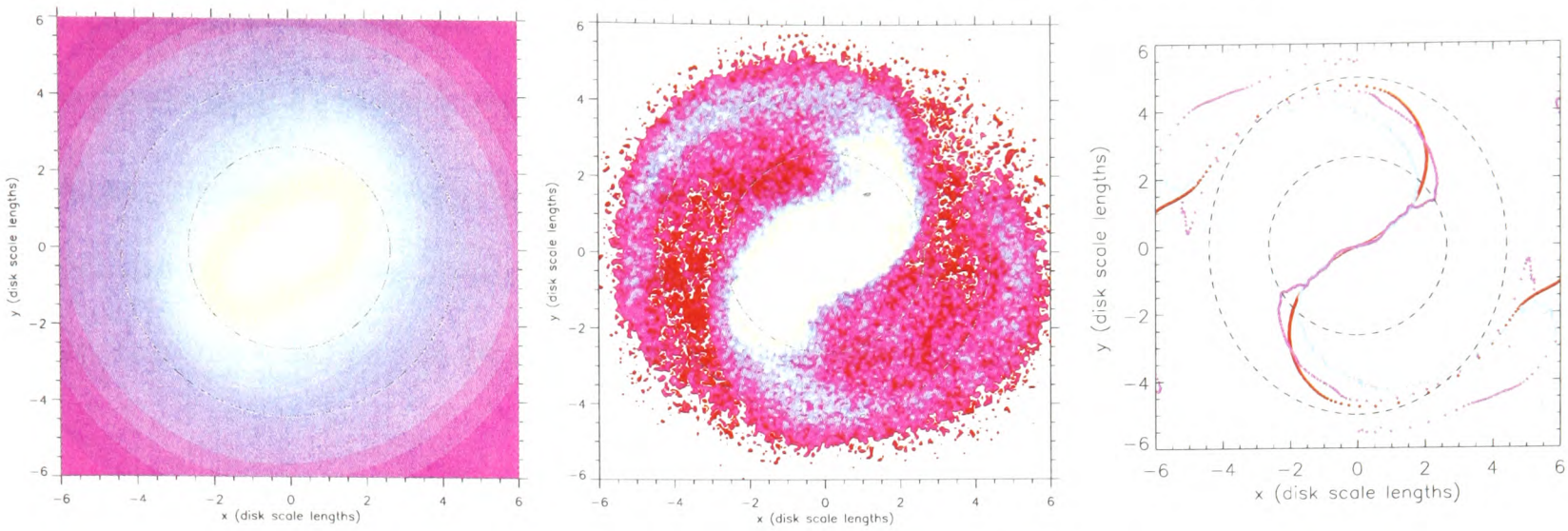


Figure 5.20: Left panel: Potential for grand design model `sta19`. Central panel: Stellar density for grand design model. Right panel: Position of the $m = 2$ maxima for grand design model. Purple diamonds mark the gaseous maxima, blue diamonds mark the stellar maxima and red diamonds mark the minima of the stellar potential. Subsequent diamonds represent equal steps in radius. In all panels, the dashed lines mark $R_{\rm min} = 2.6$ and $R_{\rm OLR} = 4.4$.

5.5.2 The flocculence of gaseous and stellar spirals

Much of the difference in the smaller scale structure between the gaseous and stellar images can be explained in terms of the dynamics of the two components. Structures in the stellar disk get strongly amplified in the dynamically cold gas, leading to shocks. Since the sound speed in the gas is an order of magnitude lower than the velocity dispersion of the stars, gaseous structures are much sharper and longer lasting than stellar ones. Hence the history of the simulation is important as well as any particular snapshot in explaining differences in morphology. This would not have been apparent if an analytical potential was used in place of an evolving disk of particles for the stellar component of the models.

Eventually a two armed density wave forms in the stellar disk and then in the gaseous disk. If the wave is fast and strong enough it will soon wipe away the smaller amplitude features in the gas, particularly in the inner regions where it is strongest. So low amplitude stellar spirals lead to flocculent gas images. This can be seen in the gas velocity fields where flocculent models have much weaker gas velocity divergences (and hence shocks), particularly outside corotation.

The systematically high values of stellar arm-interarm contrast seen in Figure 5.18 may be the reason why truly flocculent models have been hard to find. Flocculent structures in real galaxies

are thought to be due to coherent star forming regions (Mueller and Arnett 1976; Gerola and Seiden 1978). Hence, flocculent gaseous images may be inherently difficult to recreate with these models unless star formation algorithms are added.

Overall, the stellar images are much smoother because a single broad stellar feature becomes resolved into many smaller features in the gaseous case. Similarly, small asymmetries in the stellar disk are amplified in the gas to create more asymmetric structures.

5.5.3 Intrinsic characteristics of galaxies

It has been shown that the decoupling between stellar and gaseous morphologies is a function of the arm-interarm contrast in the stars. The arm-interarm contrast is an observable quantity which, using Figure 5.19, yields a rough value for $\bar{Q}(t)$. Similarly, the bar length R_{min} correlates with the disk mass f_d.

Before these trends can be applied to real galaxies, it is worth considering whether they would be affected by changes in the models. Many N-body simulations include techniques to keep the stellar particles dynamically cooled and hence prolong the life of the spiral structure. From the discussion in Section 5.4.2, it is likely that this would lead to more grand design spiral structure in the gas. A similar effect may happen if an ILR were included in the models, perhaps by adding a bulge component.

Adding self gravity to the gas could change the models in a number of different ways. Gas with just a few per cent of the total mass would aid the stellar cooling process since the gas is always dynamically cool. Any gravitational coupling between the two components could lead to greater similarity in form. However, local gaseous features would almost certainly be enhanced by gaseous self gravity, leading to a more decoupled morphology. Gas with too much self gravity (above $\sim$ 10% of the total mass) could collapse into lumps, scattering the stars and actually heating the disk, reversing the conclusion above (Junqueira and Combes 1996). Thus, to determine the true effect of changing the mass of gas present, gaseous self gravity must be added to the models.

Finally, many of the effects seen in real galaxies may only be able to be reproduced when star formation algorithms are added to the models.

5.6 Conclusions

Many spiral galaxies exhibit significant differences in morphology when observed in the K' or B bands. Using the DUAL code to combine N-body and finite-difference hydrodynamic techniques, simple simulations have reproduced some of these differences. It has been assumed that N-body simulations can mimic the dynamics of the underlying stellar population seen in the K' band and that the morphologies produced by hydrodynamic modelling can be directly compared to B band images. These assumptions are tempered by the caveats that K' band images can be substantially contaminated by the light of young stars and that, particularly outside high luminosity regions, B band images are a combination of light from star formation, dust and older stars.

With these assumptions, the simulation results imply that some aspects of dual morphology can be explained by the dynamics of stars and gas. In particular, the general principle of morphological decoupling has been verified as has the tighter nature of gaseous spiral arms (possibly due to a phase shift between the stellar density and potential). Gaseous images have been shown to be more asymmetric and less smooth than their stellar counterparts and more likely to have multiple arms. These results are very similar to those found in K' and B band galaxy observations.

Furthermore, I have shown that the morphological decoupling increases as the stellar arm-interarm contrast falls. Hence flocculent galaxies contain low amplitude stellar arms. High arm-interarm contrasts lead to a hotter disk, so it is possible to make predictions about Q from the strength of the arms. The second parameter controlling the models, the disk mass, can be predicted from the size of the stellar bar. However, the N-body simulations proceed in a very non-linear fashion and so it is not possible to draw any conclusions about the initial value of Q from the end state of the run.

The history of the simulations has a large effect on the morphology at any one time. The discrete nature of the stellar distribution allows for small scale structure to develop. The low sound speed and the appearance of shocks in the gas amplifies these structures and prevents them from dissipating quickly.

There is a lot still to be done. It remains to be seen whether the lack of truly flocculent and multi armed model galaxies is because the stellar arm-interarm contrasts are too high or

because it is not possible to produce them without adding new physics such as star formation to the models. It has not been possible to reproduce some of the gross decoupling in morphology seen in galaxies such as NGC 5195 seen in Figure 5.1. These probably result from a combination of interactions with nearby companions and dust obscuration. And many of the galaxies along the Hubble sequence remain to be modelled: adding a bulge component, removing the bar and including gaseous self gravity are all necessary to ensure that the full range of real galaxies are modelled. The self gravity of the gas will also enable these galaxies to be modelled self-consistently.

Dust and star formation both have a large role to play in defining the morphology of B band images. Hence it was expected that much of the morphological complexity and variety of spiral galaxies in the B band would not be explained by these purely dynamical models. However, enough similarities do exist between the models and the observations to imply that a significant part of the story of dual morphologies in spiral galaxies is due to the difference in the dynamics of stars and gas.

Chapter 6

Conclusion

Galaxies are very complex astronomical systems. Like the human body or an ecosystem, galaxies consist of many subsystems: old disk stars, cold disk gas, dark matter halo, globular clusters, etc. Each of these subsystems contain large numbers of individual elements, evolve on many time scales, involve different areas of physics and interact with one another through a host of mechanisms. Galaxies are dynamic, evolving in less than a Hubble time, and non-linear, making them very difficult to describe analytically. Lastly, no galaxy is an island. Interactions and mergers form a fundamental part of the history of any galaxy.

Astronomers have tried to understand galaxies by solving one piece of the puzzle at a time, and have made great strides in many areas of galactic astrophysics. Due to the non-linear nature of many galactic phenomena: star formation, stellar dynamics, interstellar hydrodynamics, galaxy formation, for example, computer simulations have been an essential part of this endeavour, complementing astronomical observations and theories.

The exponential rise in computing power and the improvement in our understanding of numerical algorithms, has enabled much progress in numerical modelling over the last sixty years. The resolution of simulations in both time and space has become ever higher, their dimensionality has increased from 1D to 2D to 3D, and extra physics has been added step by step. Initially, this physics described the dynamics of stars using N-body simulations, or gas using hydrodynamics (finite-difference or SPH). These simulations often included an additional background potential and rudimentary star formation and supernovae algorithms. More recently, simulations have included magnetohydrodynamics, stellar evolution, accretion, jets, multiphase ISM, galaxy formation, galaxy interactions, etc.

In this thesis, I have attempted to make a significant contribution to our understanding of how galaxies work by developing a new code, capable of modelling the dynamics of both gas and stars in

a spiral galaxy. The code, DUAL, has been used to investigate the important question of whether mass follows light in spiral galaxies. In the spirit of many previous simulations, hydrodynamics alone was used to constrain the shape and dark matter content of the centre of M31, by reproducing observations of molecular gas kinematics. Then, hydrodynamic and N-body techniques were combined to predict intrinsic characteristics of spiral galaxies, by reproducing observations of their dual morphology in the optical and infrared regimes. This final chapter summarizes the content and conclusions of the thesis and provides suggestions for extending the work therein.

The dynamics of gas: modelling the bulge of M31

The model of the bulge of M31 presented in Chapter 4 improves on previous ones. The mass distribution is constructed from observations, and the use of a hydrodynamics code to determine the response of the gas to the underlying potential provides an accurate representation of the gas velocity field. Although the model is reasonably robust, the hydrodynamic oscillations mentioned in Section 4.4.2 warrant further investigation to determine whether they are numerical artifacts or real features in the predicted gas flow of the galaxy. If the latter is true, this would place limits on the ability of hydrodynamical simulations to predict the true behaviour of an isothermal ISM, given that astronomical observations only capture a single snapshot in time.

The hydrodynamic theory underlying the part of the DUAL code used to model M31 was described in Section 2.1. I detailed the 2D isothermal Euler equations and the Rankine-Hugoniot jump conditions that enable the dynamics of gas to be followed both in smooth regions of flow and across discontinuities. I then introduced some of the concepts necessary to explain the physics behind the FS2 algorithm used in the DUAL code. In particular, I showed that flux splitting was based on the theory of characteristics and that slope limiting followed from imposing the Rankine-Hugoniot conditions. The FS2 algorithm was shown to have very little numerical viscosity and to reproduce the solution of the 1D isothermal shocktube problem. This, together with its ability to produce realistic simulations of barred spiral galaxies and the fact that it is computationally inexpensive, makes FS2 a reasonable choice for galaxy simulations. There is, however, a case for running the same model using FS2 and other codes such as CMHOG and ZEUS to compare the

details of their results. A similar comparison was made by van Albada et al. (1982), in which FS2 was shown to be the code of choice, but it would be useful to have a modern rematch.

It may also be worthwhile including SPH codes in such a comparison. However, SPH codes have a hard time modelling low gas density regions, and this is a serious disadvantage when compared to grid-based codes. In this work, the use of grid-based hydrodynamics has been instrumental in allowing the inclusion of N-body forces into the code. The use of a Cartesian geometry has prevented criticism of circular features in the results as artifacts of the grid. Of course, when modelling a large range of length scales, simple Cartesian geometry is impractical, but utilizing adaptive grids may still be preferable to a polar geometry.

In using the DUAL code to determine the mass distribution of the bulge of M31, the bulge was assumed to be triaxial and rotating in the plane of the disk. This assumption was justified by finding the model that best fits the CO observations of Loinard et al. (1995) along the line of nodes of the disk. The parameter space was defined by the B band bulge mass-to-light ratio Υ_B, the bulge pattern speed Ω_p, and the angle between the projected minor axis of the bulge and the line of nodes of the disk ϕ. This model showed that the bulge rotates fast, with $\Omega_p = 54$ km s^{-1} kpc^{-1}, and the ratio of bulge semi-major axis to corotation radius of $\mathcal{R} = 1.2$. Further, $\Upsilon_B = 6.5$, $\phi = 15°$, the semi-major axis $a = 3.5$ kpc and the mass of the bulge $M_{\rm bulge} = 2.3 \times 10^{10}$ M$_\odot$. Given that triaxiality has now been demonstrated for both our Galaxy and our nearest large neighbour M31, it is likely that triaxial bulges are a common feature of all spiral galaxies. This is, however, based on small number statistics and it would be worthwhile modelling larger numbers of SA and SA(B) galaxies to see whether they too harbour triaxial bulges.

This bulge model was also shown to be consistent with CO observations parallel to the line of nodes of the disk. Since these observations covered a larger range in velocity than predicted by the models described above, the kinematics of gas away from the plane of the disk had to be included. In fact, a fairly large thickness (half width at half maximum intensity of 200–500 pc) had to be given to the molecular gas disk. The lower end of this range is attainable if the gas loses about one fifth of its velocity per kpc above the plane. It not clear whether the inner molecular disk of M31 is truly thick or whether adding a thick gas disk to the model is just a way to account for a

warped thin disk. Other out of plane or peculiar velocities (such as those due to hydraulic jumps or turbulence) could also contribute to the wide velocity range covered by the CO observations. The deduced thickness of the gas disk is, however, consistent with other recent work on central gas disks in spiral galaxies.

The value of the B band mass-to-light ratio for the best fitting model is consistent with Kent (1989) who finds that $\Upsilon_B = 7.1 \pm 0.5$. It does not, however, fit the log-log relationship between the B band mass-to-light ratio and the total B band luminosity of elliptical galaxies described by Gerhard et al. (2001). If this is naively extended to galactic bulges, it predicts an extremely low value of $\Upsilon_B = 1.5$. Based on the data of Gerhard et. al., the validity of this relationship seems far from conclusive. It would be very interesting to see a selection of galactic bulges added to this data set to test their conclusions at low luminosities.

Section 4.4.3 provided an absorption map of the bulge region. This shows a number of similarities with the 7.7 μm observations of dust in the bulge region from Lequeux (2000). In particular, the reproduction of the offset dust lanes of the model in the observations provides further evidence for the triaxiality of the bulge. However, the spiral structure seen in these dust observations was not accounted for by the model. This may be due to the lack of self gravity in the gas. There have been a number of recent observational and computational studies of nuclear bars and spirals in disk galaxies. Hence, it may prove fruitful to perform much higher resolution simulations of the bulge of M31 to determine whether such structures are likely to exist and what size of gas inflow the model predicts. This may also help to explain the 7.7 μm spiral structure.

A major conclusion from this investigation concerns the distribution of dark matter within the inner 3.5 kpc of M31. The bulge of M31 has been shown to be old and a fast rotator and, based on arguments from galactic formation, galactic dynamics and observations of globular clusters, it is likely that the dark halo of M31 rotates slowly. Hence, from the work of Debattista and Sellwood (2000), dynamical friction between the bulge and the halo implies that M31 must have a maximum disk and thus a halo mass fraction within 3.5 kpc of just a few percent. This is in stark contrast to the predictions of cold dark matter theories which imply that the halo mass fraction within 3.5 kpc is between 22% and 30%. This finding is in line with many investigations of dark

matter in the central regions of barred spiral galaxies, and needs to be addressed by theories of dark matter and galaxy formation. The story for unbarred galaxies, however, is not so clear cut. It might seem perverse for barred galaxies to be disk dominated and unbarred galaxies to be halo dominated in their inner parts, but, amongst others, the simulations of Kranz et al. (2001) find that a significant halo is needed to reproduce the gas velocity field of NGC 4254. They may not be justified in their assumption of a fixed pattern speed for an unbarred galaxy, but it shows that dark matter in unbarred galaxies needs to be investigated more closely.

The dynamics of stars and gas: modelling dual morphology in spiral galaxies

The investigations of gas dynamics in the Andromeda galaxy were broadened to the dynamics of both stars and gas in a general Sc-type galaxy in Chapter 5. In particular, I tried to understand the cause of the significant differences in morphology exhibited by many spiral galaxies when observed in the optical or the near-infrared. The DUAL code was used again, this time combining N-body and finite-difference hydrodynamic techniques. N-body methods, particularly the particle-mesh family of algorithms, are described in Section 2.3. I argue there that softened gravity provides 2D simulations with an effective scale height (which includes the effect of a grid) and show how that relates to the softening length.

The use of N-body techniques, instead of an analytic potential, proved to be essential in reproducing the dual nature of stellar and gaseous morphologies. Much of the differences in form arise from the difference in the time scales involved in creating and dispersing structure. Noise seeds in the stellar distribution generate non-axisymmetric features and the large stellar velocity dispersion ensures that these structures are relatively broad. Conversely, the low sound speed and the appearance of shocks in the gas, amplifies and narrows the stellar structures, and also prevents them from disappearing too quickly. This means that the history of a simulation has a large effect on the morphology at any one time. Applying these ideas to real galaxies, implies that it may be difficult to determine the past history of a galaxy from a single snapshot in time.

Chapter 5 began by reviewing the observations of spiral galaxies in the optical and near-infrared wavebands. This highlighted some of the decoupled morphology that can be observed in some

spiral galaxies. At present, however, there are no statistics for how prevalent this decoupling is. Hence, the extremes of the observations mentioned in Chapter 5 should be viewed as instances of a common, but possibly not dominant, phenomena. The possible lack of decoupling in many spiral galaxies may account for the difficulty in creating simulations with radically different morphologies in the stellar and gaseous components. It would be very helpful to analyze a statistically meaningful sample of galaxies in B and K' bands in order to determine the distribution of decoupling in real galaxies, particularly the differences in flocculence and pitch angle.

The simulations in Chapter 5 are performed under the assumptions that N-body models are able to reproduce the dynamics of the underlying stellar population seen in the K' band and that the morphologies produced by hydrodynamic modelling are representative of those seen in B band images. These assumptions are tempered by the caveats that K' band images are contaminated by the light of young stars and that, particularly outside of high luminosity regions, B band images combine light from star formation, dust and older stars. With these assumptions in mind, a selection of models were run, defined by the fraction of mass in the disk f_d and the minimum value of the stability parameter Q_{min}. The extent of the parameter space was limited to the region in which it was possible to produce two armed spiral structure in the stellar component.

Previously, it had been thought that dust and star formation were needed to explain most of the differences in optical and near-infrared observations. The results of these simulations suggest that a significant part of the story of morphological decoupling in spiral galaxies is due to the difference in the dynamics of stars and gas. Gaseous spiral arms are shown to have tighter pitch angles than stellar spirals. This was tentatively attributed to a phase shift between the stellar density and the potential, which was shown to exist in at least one of the models. This is an appealing idea which needs further investigation. Is this effect present in real galaxies with pitch angle differences? Do my simulations posses the characteristic accretion and excretion of stellar particles from the corotation radius? Gaseous images are also shown to be more asymmetric and less smooth than their stellar counterparts and more likely to have multiple arms. These results are all reminiscent of the differences between optical and infrared observations.

I show that morphological decoupling increases as the stellar arm-interarm contrast falls. Hence

flocculent galaxies contain low amplitude stellar arms. I also show that high arm-interarm contrasts lead to hotter disks, and so predictions can be made about Q from the strength of the stellar arms. However, the non-linear evolution of N-body simulations means that it is not possible to draw conclusions about the initial value of Q from the end state of the run.

A major tool in the analysis of the models is the 2D Fourier decomposition of the resulting images into logarithmic spirals. This procedure is discussed in some depth with attention paid to the choice of the minimum radius and the interpretation of the spectra. I introduce a numerical value to characterize the flocculence of a galaxy, $\mathcal{F}$, which is closely related to the uniformity of the Fourier spectra. It would be useful to calculate the value of $\mathcal{F}$ for many more real galaxies and compare the results to the arm classes of Elmegreen and Elmegreen (1982).

The amount of morphological decoupling explained by the simulations remains to be quantified. Obviously dust, star formation and galaxy interactions all have a large role to play in defining the morphology of B band images. Hence, much of the morphological complexity and variety of spiral galaxies could not be explained by purely dynamical models. They have certainly not been able to reproduce the gross decoupling seen in galaxies such as NGC 5195 (Figure 5.1).

In order to determine the relative importance of star formation, dust and the underlying dynamics in defining the morphology of a galaxy, a better understanding is needed of how the simulations relate to B band images. This may be approached by adding star formation and dust creation algorithms. Care must be taken to correctly simulate the relationship between stellar age and luminosity. Alternatively, it may be more profitable to compare the model gaseous images with H I or CO data rather than B band observations. Alongside this, it remains to be seen whether the lack of truly flocculent and multi armed model galaxies is due to stellar arm-interarm contrasts that are artificially too high or because it is not possible to produce them without extra physics in the models. In any case, the reason for the systematic difference between model and observed arm-interarm contrasts of Figure 5.18 needs explaining. It would be helpful to understand how the various morphological parameters defined in Chapter 5, particularly the pitch angle and the flocculence, respond to the addition of dusty structures and star forming knots, such as those seen in real galaxies, to the simulated images.

These models need to be extended to cover the full range of Hubble galaxies. Adding a bulge component would enable results for Sc galaxies to be compared with models of Sb and Sa galaxies. This would also provide a good opportunity to search for unbarred or weakly barred models. Assuming that strong bars lead to high arm-interarm contrasts, the argument presented in Section 5.4.4 relating high $\mathcal{C}$ to dynamically hot stellar disks implies that weaker bars may also lead to longer lasting spiral structure.

Since gas mass is the discriminating factor between galaxies along the Hubble sequence, an essential part of the investigation into dual morphologies is the inclusion of gaseous self gravity. Further, as described in Section 5.5.3, it is not obvious what effect gaseous self gravity will have on the morphology of the stars and gas. A little gas should cool the stars but too much will lead to clumping in the gas and probably hotter stars. As well as giving the gaseous component a mass, adding gas self gravity provides a back reaction on the stellar component from the work done by the stars on the gas. This would make the DUAL code fully self consistent and enable it to realize its full potential. The amended FS2 algorithm presented in Section 3.3.3 ensures the necessary symmetry between the hydrodynamic and N-body parts of the code, and is the first step towards this goal.

It is also important to extend the DUAL code into 3D. The structure of gas, and hence dust, in spiral galaxies is likely to be three dimensional, as the work of Chapter 4 shows. Hence, to understand the contribution of dust to B band morphology, gaseous dynamics must be modelled in 3D. On a grander scale, as described at the start of this chapter, the addition of gaseous self gravity and 3D modelling will bring us closer, step by step, to the broader goal of understanding the formation and evolution of galaxies. However, even as it stands, the DUAL code provides a powerful and publicly available[1] tool with which to further investigate many aspects of galaxy dynamics.

[1]The DUAL code can be downloaded from `www-thphys.physics.ox.ac.uk/users/SimonBerman`.

Bibliography

Abd al-Rahman b. Umar al-Sufi, 964, Book of the Fixed Stars, Bodleian Library, Oxford

Allen C. W., 1973, Astrophysical quantities, Athlone Press, London, 3rd ed.

Arnholm C. A., 1997, Mixed language programming, `home.online.no/~arnholm/cppf77.htm`

Athanassoula E., 1992a, MNRAS, **259**, 328

Athanassoula E., 1992b, MNRAS, **259**, 345

Athanassoula E. and Bureau M., 1999, ApJ, **522**, 699

Baade W. and Arp H., 1964, ApJ, **139**, 1027

Barnes J. and Efstathiou G., 1987, ApJ, **319**, 575

Bertin G. and Lin C. C., 1996, Spiral structure in galaxies, MIT Press, Cambridge

Bertin G., Lin C. C., Lowe S. A., and Thurstans R. P., 1989, ApJ, **338**, 78

Bertola F., Vietri M., and Zeilinger W. W., 1988, The Messenger, **52**, 24

Binney J. and Merrifield M., 1998, Galactic Astronomy, Princeton University Press, Princeton

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

Binney J., Bissantz N., and Gerhard O., 2000, ApJ, **537**, L99

Binney J., Gerhard O., and Silk J., 2001, MNRAS, **321**, 471

Binney J. J. and Evans N. W., 2001, MNRAS, **327**, L27

Block D., Bertin G., Stockton A., Grosbol P., Moorwood A., and Peletier R., 1994, A&A, **288**, 365

Block D. L. and Puerari I., 1999, A&A, **342**, 627

Block D. L. and Wainscoat R. J., 1991, Nature, **353**, 48

Block D. L., Elmegreen B. G., and Wainscoat R. J., 1996, Nature, **381**, 674

Bowers R. L. and Deeming T., 1984, Astrophysics, Vol. 2, Jones and Bartlett, Boston

Brinks E. and Burton W. B., 1984, A&A, **141**, 195

Brinks E. and Shane W. W., 1984, A&AS, **55**, 179

Burstein D. and Heiles C., 1982, AJ, **87**, 1165

Burstein D. and Rubin V. C., 1985, ApJ, **297**, 423

Buta R. and Block D. L., 2001, ApJ, **550**, 243

Carlberg R. G. and Freedman W. L., 1985, ApJ, **298**, 486

Ciardullo R., Rubin V. C., Ford W. K. J., Jacoby G. H., and Ford H. C., 1988, AJ, **95**, 438

Comins N. F. and Zeltwanger T. M., 2000, in AAS Meeting 197, p. 7609

Conselice C. J., 1997, PASP, **109**, 1251

Contopoulos G., 1956, Zeitschrift Astrophysics, **39**, 126

Contopoulos G. and Papayannopoulos T., 1980, A&A, **92**, 33

Courant R., Isaacson E., and Rees M., 1952, Comm. Pure. Appl. Math., **5**, 243

Cowie L. L., 1980, ApJ, **236**, 868

Dame T. M. and Thaddeus P., 1994, ApJ, **436**, L173

Dame T. M., Koper E., Israel F. P., and Thaddeus P., 1993, ApJ, **418**, 730

Danver C. G., 1942, Annals of the Observatory of Lund, **10**, 7

de Grijs R. and Peletier R. F., 1997, A&A, **320**, L21

de Vaucouleurs G., 1958, ApJ, **128**, 465

Debattista V. P. and Sellwood J. A., 2000, ApJ, **543**, 704

Debattista V. P., Corsini E. M., and Aguerri J. A. L., 2001, astro-ph, 0112288, in press

Dehnen W., 1999a, AJ, **118**, 1201

Dehnen W., 1999b, ApJ, **524**, L35

Dehnen W., 2001, MNRAS, **324**, 273

Dehnen W. and Binney J. J., 1998, MNRAS, **298**, 387

Drimmel R. and Spergel D. N., 2001, ApJ, **556**, 181

Elmegreen B. G. and Thomasson M., 1993, A&A, **272**, 37

Elmegreen D. M. and Elmegreen B. G., 1982, MNRAS, **201**, 1021

Englmaier P. and Gerhard O., 1997, MNRAS, **287**, 57

Englmaier P. and Gerhard O., 1999, MNRAS, **304**, 512

Englmaier P. and Shlosman I., 2000, ApJ, **528**, 677

Eskridge P., Frogel J., Pogge R., Quillen A. C., Davies R., DePoy D., Houdashelt M. L., Kuchinski L., Ramírez S., Sellgren K., Terndrup D., and Tiede G. P., 2000, AJ, **119**, 536

Fuchs B., 2000, astro-ph, 0010358, in press

Fux R., 1999, A&A, **345**, 787

Garcia-Burillo S., Guelin M., Cernicharo J., and Dahlem M., 1992, A&A, **266**, 21

Gerhard O., Kronawitter A., Saglia R. P., and Bender R., 2001, AJ, **121**, 1936

Gerhard O. E. and Vietri M., 1986, MNRAS, **223**, 377

Gerhard O. E., Vietri M., and Kent S. M., 1989, ApJ, **345**, L33

Gerola H. and Seiden P. E., 1978, ApJ, **223**, 129

Gerritsen J. P. E. and Icke V., 1997, A&A, **325**, 972

Gerssen J., Kuijken K., and Merrifield M. R., 1999, MNRAS, **306**, 926

Godunov S. K., 1959, Math. Sbornik., **47**, 357

Grosbol P. J. and Patsis P. A., 1998, A&A, **336**, 840

Hernquist L. and Katz N., 1989, ApJS, **70**, 419

Hodge P., 1992, The Andromeda Galaxy, Kluwer Acad. Pub., Dordrecht

Hodge P. W. and Kennicutt R. C., 1982, AJ, **87**, 264

Hohl F., 1971, ApJ, **168**, 343

Huchra J. P., 1993, in ASP Conf. Ser. 48, The Globular Cluster-Galaxy Connection, p. 420

Hunter J. H., England M. N., Gottesman S. T., Ball R., and Huntley J. M., 1988, ApJ, **324**, 721

Iye M., Okamura S., Hamabe M., and Watanabe M., 1982, ApJ, **256**, 103

Jaynes E. T., 1958, Probability theory in science and engineering, Colloquium Lectures in Pure and Applied Science, **4**, Socony-Mobil Oil Co., Dallas

Johnson H. L., 1966, ApJ, **143**, 187

Junqueira S. and Combes F., 1996, A&A, **312**, 703

Kennicutt R. C., 1981, AJ, **86**, 1847

Kent S. M., 1989, AJ, **97**, 1614

Knebe A., Green A., and Binney J., 2001, MNRAS, **325**, 845

Kormendy J., 1981, in Structure and Evolution of Normal Galaxies, ed. S. M. Fall, p. 85

Kormendy J., 1982, ApJ, **257**, 75

Kormendy J. and Illingworth G., 1982, ApJ, **256**, 460

Kranz T., Slyz A., and Rix H., 2001, ApJ, **562**, 164

Kuijken K. and Gilmore G., 1991, ApJ, **367**, L9

Landau L. D. and Lifshitz E. M., 1959, Fluid mechanics, Pergamon Press, Oxford

Laney C. B., 1998, Computational Gasdynamics, Cambridge University Press, Cambridge

Lequeux J., 2000, in WE-Heraeus Seminar, 232, ed. E. M. Berkhuijsen, Shaker, Aachen, p. 63

Lindblad B., 1956, Stockholms Observatoriums Annaler, 2

Lindblad P. A. B. and Kristen H., 1996, A&A, **313**, 733

Lindblad P. A. B., Lindblad P. O., and Athanassoula E., 1996, A&A, **313**, 65

Loinard L. and Allen R. J., 1998, ApJ, **499**, 227

Loinard L., Allen R. J., and Lequeux J., 1995, A&A, **301**, 68

Loinard L., Dame T. M., Heyer M. H., Lequeux J., and Thaddeus P., 1999, A&A, **351**, 1087

Maciejewski W., Teuben P. J., Sparke L. S., and Stone J. M., 2002, MNRAS, **329**, 502

Martos M. A. and Cox D. P., 1998, ApJ, **509**, 703

Melchior A., Viallefond F., Guélin M., and Neininger N., 2000, MNRAS, **312**, L29

Merrifield M. R. and Kuijken K., 1995, MNRAS, **274**, 933

Milgrom M., 1983, ApJ, **270**, 365

Miller R. H., 1971, Ap&SS, **14**, 73

Miller R. H., 1974, ApJ, **190**, 539

Miller R. H., Prendergast K. H., and Quirk W. J., 1970, ApJ, **161**, 903

Mueller M. W. and Arnett W. D., 1976, ApJ, **210**, 670

Mulder W. A., 1986, A&A, **156**, 354

Navarro J. F., Frenk C. S., and White S. D. M., 1996, ApJ, **462**, 563

Navarro J. F., Frenk C. S., and White S. D. M., 1997, ApJ, **490**, 493

Neininger N., Guelin M., Ungerechts H., Lucas R., and Wielebinski R., 1998, Nature, **395**, 871

Pellet A., 1976, A&A, **50**, 421

Piner B. G., Stone J. M., and Teuben P. J., 1995, ApJ, **449**, 508

Pringle J. E., Allen R. J., and Lubow S. H., 2001, MNRAS, **327**, 663

Regan M. W., Sheth K., and Vogel S. N., 1999, ApJ, **526**, 97

Rhoads J. E., 1998, AJ, **115**, 472

Rieke G. H. and Lebofsky M. J., 1985, ApJ, **288**, 618

Rix H., 1993, PASP, **105**, 999

Rix H. and Rieke M. J., 1993, ApJ, **418**, 123

Rix H. and Zaritsky D., 1995, ApJ, **447**, 82

Roberts W. W., Huntley J. M., and van Albada G. D., 1979, ApJ, **233**, 67

Ruiz M. T., 1976, ApJ, **207**, 382

Sandage A. and Bedke J., 1994, The Carnegie atlas of galaxies, Carnegie Institution, Washington

Sanders R. H. and Prendergast K. H., 1974, ApJ, **188**, 489

Sanders R. H. and Tubbs A. D., 1980, ApJ, **235**, 803

Saraiva Schroeder M. F., Pastoriza M. G., Kepler S. O., and Puerari I., 1994, A&AS, **108**, 41

Schaap W. E., Sancisi R., and Swaters R. A., 2000, A&A, **356**, L49

Schroeder M. C., 1989, D.Phil thesis, Maine University

Schroeder M. C. and Comins N. F., 1986, Lecture Notes in Physics, **267**, 223

Schwarz M. P., 1981, ApJ, **247**, 77

Seigar M. S. and James P. A., 1998, MNRAS, **299**, 685

Sellwood J. A., 1981, A&A, **99**, 362

Sellwood J. A., 1987, ARA&A, **25**, 151

Sellwood J. A. and Carlberg R. G., 1984, ApJ, **282**, 61

Sellwood J. A. and Sparke L. S., 1988, MNRAS, **231**, 25P

Simien F., Pellet A., and Monnet G., 1979, A&A, **72**, 12

Spergel D. N. and Steinhardt P. J., 2000, Physical Review Letters, **84**, 3760

Spergel D. N., Malhotra S., and Blitz L., 1996, in Spiral Galaxies in the Near-IR, ed. D. Minniti and H-W. Rix, Springer-Verlag, Berlin, p. 128

Spitzer L. J., 1990, ARA&A, **28**, 71

Stanek K. Z. and Garnavich P. M., 1998, ApJ, **503**, L131

Stark A. A., 1977, ApJ, **213**, 368

Stark A. A., 1985, in IAU Symp. 106, The Milky Way Galaxy, p. 445

Stark A. A. and Binney J., 1994, ApJ, **426**, L31

Steinmetz M. and Bartelmann M., 1995, MNRAS, **272**, 570

Strom S. E., Strom K. M., and Jensen E. B., 1976, ApJ, **206**, L11

Swaters R. A., Madore B. F., and Trewhella M., 2000, ApJ, **531**, L107

Swaters R. A., Sancisi R., and van der Hulst J. M., 1997, ApJ, **491**, 140

Thomasson M., Elmegreen B. G., Donner K. J., and Sundelius B., 1990, ApJ, **356**, L9

Thornley M. D., 1996, ApJ, **469**, L45

Toomre A., 1964, ApJ, **139**, 1217

Toro E. F., 1999, Riemann solvers and numerical methods for fluid dynamics, Springer-Verlag

Trager S. C., Faber S. M., Worthey G., and González J. J., 2000, AJ, **119**, 1645

Tremaine S. and Weinberg M. D., 1984, ApJ, **282**, L5

van Albada G. D. and Roberts W. W., 1981, ApJ, **246**, 740

van Albada G. D., van Leer B., and Roberts W. W., 1982, A&A, **108**, 76

van Albada T. S., Bahcall J. N., Begeman K., and Sancisi R., 1985, ApJ, **295**, 305

van den Bergh S., 2000, The Galaxies of the Local Group, Cambridge University Press, Cambridge

van der Kruit P. C. and Searle L., 1981, A&A, **95**, 105

van Genderen A. M., 1973, A&A, **24**, 47

van Leer B., 1982, Lecture Notes in Physics, **170**, 507

van Leer B., 1983, Computational methods for ideal compressible flow, ICASE report 83-38

Veldhuizen T., 2000, Blitz++ class library, `www.oonumerics.org/blitz/`

Walterbos R. A. M., 2001, in Encyclopedia of Astronomy and Astrophysics, Nature Publishing

Walterbos R. A. M. and Kennicutt R. C., 1988, A&A, **198**, 61

Weiner B. J. and Sellwood J. A., 1999, ApJ, **524**, 112

Weiner B. J., Williams T. B., van Gorkom J. H., and Sellwood J. A., 2001, ApJ, **546**, 916

Whitehurst R. N. and Roberts M. S., 1972, ApJ, **175**, 347

Wilkinson M. I. and Evans N. W., 1999, MNRAS, **310**, 645

Zaritsky D. and Lo K. Y., 1986, ApJ, **303**, 66

Zeltwanger T., 1999, in AAS Meeting 195, p. 9704

Zhang X., 1996, ApJ, **457**, 125

Zwicky F., 1955, PASP, **67**, 232

Zwicky F., 1957, Morphological astronomy, Springer, Berlin

Appendix A

The $\mathcal{X}$ sweep of the FS2 algorithm

A.1 Overview

The $\mathcal{X}^{\Delta t}$ operator in conservative form is

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \frac{\Delta t}{\Delta x}(\hat{\mathbf{F}}_{i+1/2}^{n+1/2} - \hat{\mathbf{F}}_{i-1/2}^{n+1/2}), \tag{A.1}$$

where $\mathbf{U}_i^n$ is the value of $\mathbf{U}$ at time $t = n\,\Delta t$ and position $x = i\,\Delta x$ and $\hat{\mathbf{F}}_{i+1/2}^{n+1/2}$ is a flux function to be determined. The flux is defined on cell boundaries $i+1/2$ and at an intermediate time $n+1/2$. The $\mathcal{Y}^{\Delta t}$ operator can be obtained by replacing i by j, x by y and $\mathbf{F}$ by $\mathbf{G}$.

A.2 Evolving the cell centred values

According to van Albada et al. (1982), any first order upwind differencing method can achieve second order accuracy in time by advancing the cell boundary values to the intermediate time level $t^{n+1/2}$. In obtaining these values, the interaction between cells can be fully ignored.

In order to calculate the cell boundary values at $(x_{i\pm1/2}, t^{n+1/2})$, it is necessary to evolve the values at the cell centres to $(x_i, t^{n+1/2})$ and then interpolate these values to the cell boundaries. FS2 uses primitive variables $\mathbf{Q} = (\Sigma, u, v)^T$ to evolve the cell centered values to the intermediate time step $t^{n+1/2}$ by

$$\mathbf{Q}_i^{n+1/2} = \mathbf{Q}_i^n + \frac{\Delta t}{2}\left(\frac{\partial \mathbf{Q}}{\partial t}\right)_i^n. \tag{A.2}$$

The partial differential $(\mathbf{Q}_t)_i^n$ can be deduced by rewriting the discrete Euler equations without

source terms in terms of primitive, instead of conserved, variables. Hence

$$\left(\frac{\partial \mathbf{Q}}{\partial t}\right)_i^n = -\frac{1}{\Delta x}\begin{pmatrix} u_i^n(\delta\Sigma)_i^n + \Sigma_i^n(\delta u)_i^n \\ (c^2/\Sigma_i^n)(\delta\Sigma)_i^n + u_i^n(\delta u)_i^n \\ u_i^n(\delta v)_i^n \end{pmatrix} \qquad (x-\text{sweep}),$$

$$\left(\frac{\partial \mathbf{Q}}{\partial t}\right)_i^n = -\frac{1}{\Delta y}\begin{pmatrix} v_i^n(\delta\Sigma)_i^n + \Sigma_i^n(\delta v)_i^n \\ v_i^n(\delta u)_i^n \\ (c^2/\Sigma_i^n)(\delta\Sigma)_i^n + v_i^n(\delta v)_i^n \end{pmatrix} \qquad (y-\text{sweep}), \tag{A.3}$$

where $(\delta\Sigma)_i^n$, $(\delta u)_i^n$ and $(\delta v)_i^n$ are defined below.

A.3 Reconstructing the intra-cell data

To achieve second order accuracy in space, the cell centred values $\mathbf{Q}_i^n$ (equal to the mean values across the cell) are replaced within each cell by piecewise linear functions of the primitive variables, such that

$$\mathbf{Q}^n(x) = \mathbf{Q}_i^n + \frac{x - x_i}{\Delta x}(\delta\mathbf{Q})_i^n, \qquad x_{i-1/2} < x < x_{i+1/2} \tag{A.4}$$

where $(\delta\mathbf{Q})_i^n$ is a slope vector defined by

$$(\delta\Sigma)_i^n = \Sigma_i^n \,\text{ave}\left(2\frac{\Sigma_{i+1}^n - \Sigma_i^n}{\Sigma_{i+1}^n + \Sigma_i^n}, 2\frac{\Sigma_i^n - \Sigma_{i-1}^n}{\Sigma_i^n + \Sigma_{i-1}^n}\right),$$

$$(\delta\mathbf{v})_i^n = c\,\text{ave}\left(\frac{\mathbf{v}_{i+1}^n - \mathbf{v}_i^n}{c}, \frac{\mathbf{v}_i^n - \mathbf{v}_{i-1}^n}{c}\right), \tag{A.5}$$

$$\text{ave}(a,b) = \frac{(b^2+\epsilon^2)a + (a^2+\epsilon^2)b}{a^2 + b^2 + 2\epsilon^2}. \tag{A.6}$$

The formulation of $(\delta\Sigma)_i^n$ ensures positivity for the density. The averaging function, $\text{ave}(a,b)$, is the van Leer limiter defined in Equation A.6. ϵ is a small bias taken to be the average of the differences in the characteristic variables $\mathbf{W} = (W^+, W^-)$ (the Riemann invariants of Section 2.1.2) over a

smooth part of the flow (Mulder 1986). For the x-sweep

$$W^{\pm} = u \pm c \ln \Sigma, \tag{A.7}$$

and for the y-sweep simply replace u by v. Then

$$\epsilon = \frac{1}{3N} \sum_{i=0}^{N-1} |W^{+}_{i+1} - W^{+}_{i}| + |W^{-}_{i+1} - W^{-}_{i}|, \tag{A.8}$$

where N is the number of grid cells in the x or y direction.

A.4 Calculating the cell boundary values

The cell boundary values of the primitive variables $\mathbf{Q}^{n+1/2}_{(i\pm1/2)\mp}$ are calculated by interpolation from the cell centred values using the above prescription. Thus

$$\mathbf{Q}^{n+1/2}_{(i+1/2)-} = \mathbf{Q}^{n+1/2}_{i} + \frac{(\delta\mathbf{Q})^{n}_{i}}{2},$$

$$\mathbf{Q}^{n+1/2}_{(i+1/2)+} = \mathbf{Q}^{n+1/2}_{i+1} - \frac{(\delta\mathbf{Q})^{n}_{i+1}}{2}. \tag{A.9}$$

The reason that $\mathbf{Q}^{n+1/2}_{(i+1/2)-}$ is not simply written as $\mathbf{Q}^{n+1/2}_{i+1/2}$ is because the cell boundary values of the vector $\mathbf{Q}$ are derived from cell centered values $\mathbf{Q}_i$. Therefore $\mathbf{Q}^{n+1/2}_{i+1/2}$ could be two different values depending whether it is derived from the left or right hand cell.

The values of the conserved variables at the cell boundaries, $\mathbf{U}^{n+1/2}_{(i\pm1/2)\mp}$, can now be calculated from the values of the primitive variables at the same position and time,

$$\mathbf{U}^{n+1/2}_{(i\pm1/2)\mp} = \mathbf{U}(\mathbf{Q}^{n+1/2}_{(i\pm1/2)\mp}). \tag{A.10}$$

A.5 Determining the flux splitting function

The technique of flux vector splitting (or just flux splitting) ensures that the difference scheme used in the FS2 algorithm is always upwind by splitting the flux vector into a positive and a negative part containing the positive and negative running characteristics respectively. This is

possible due to the homogeneity of the Euler equations, i.e. $\mathbf{F}(\mathbf{U}) = \mathbf{A}(\mathbf{U})\mathbf{U}$, where $\mathbf{A}(\mathbf{U}) = \frac{d\mathbf{F}}{d\mathbf{U}}$ is the Jacobian matrix.

Using flux vector splitting, the flux functions become

$$\hat{\mathbf{F}}^{n+1/2}_{i\pm1/2} = \mathbf{F}^{+}(\mathbf{U}^{n+1/2}_{(i\pm1/2)-}) + \mathbf{F}^{-}(\mathbf{U}^{n+1/2}_{(i\pm1/2)+}). \tag{A.11}$$

Fluxes at cell boundaries $x_{i\pm1/2}$ have been split into two components: $\mathbf{F}^{+}(\mathbf{U}^{n+1/2}_{(i\pm1/2)-})$, the value of the flux travelling to the right from the cell to the left of the boundary and $\mathbf{F}^{-}(\mathbf{U}^{n+1/2}_{(i\pm1/2)+})$, the value of the flux travelling to the left from the cell to the right of the boundary.

By performing this splitting correctly, the Jacobian matrices of the positive and negative fluxes will contain only positive or negative eigenvalues respectively. Each component of the flux can then be subjected to upwind differencing. Using the van Leer (1982) method of flux vector splitting, the flux vector is rewritten in terms of the Mach number M,

$$\mathbf{F} = \begin{pmatrix} \Sigma cM \\ \Sigma c^2(M^2+1) \\ \Sigma cMv \end{pmatrix} \qquad (x\text{−sweep}),$$

$$\mathbf{G} = \begin{pmatrix} \Sigma cM \\ \Sigma cMu \\ \Sigma c^2(M^2+1) \end{pmatrix} \qquad (y\text{−sweep}), \tag{A.12}$$

where $M = u/c$ or $M = v/c$ for x and y sweeps respectively. The positive and negative components

of the flux function are then defined as

$$
\begin{aligned}
&\mathbf{F}^+ = \mathbf{F}, \qquad \mathbf{F}^- = \mathbf{0} && \text{if } M \geq 1, \\
&\mathbf{F}^+ = \mathbf{0}, \qquad \mathbf{F}^- = \mathbf{F} && \text{if } M \leq -1, \\
&\mathbf{F}^{\pm} = \pm\frac{\Sigma c}{4}(M \pm 1)^2 \begin{pmatrix} 1 \\ \pm 2c \\ v \end{pmatrix} && \text{if } |M| < 1 \qquad (x\text{–sweep}), \\
&\mathbf{G}^{\pm} = \pm\frac{\Sigma c}{4}(M \pm 1)^2 \begin{pmatrix} 1 \\ u \\ \pm 2c \end{pmatrix} && \text{if } |M| < 1 \qquad (y\text{–sweep}).
\end{aligned} \tag{A.13}
$$

Van Leer's splitting ensures that the split fluxes $\mathbf{F}^{\pm}$, and their differentials $\mathbf{A}(\mathbf{U})^{\pm}$, are continuous, leading to standard upwind differencing in supersonic regions and smoother numerical results generally. Furthermore, it is apparent that one of the eigenvalues of $\mathbf{A}(\mathbf{U})^{\pm}$ will vanish for $|M| < 1$ which makes it possible to build stationary shock structures with no more than two interior zones (van Leer 1982).

Appendix B

Description of models of the bulge of M31

Table B.1: Description of all bulge models

Number	$\mathcal{R}$	$\Upsilon_B\,(M_\odot/L_\odot)$	ϕ	χ^2/N	$\Omega_p\,(\mathrm{km\ s^{-1}\ kpc^{-1}})$
700	1.0	5.5	9	3.08	55.5
701	1.0	7.5	9	5.30	59.5
702	1.0	5.5	15	4.51	63.5
703	1.0	7.5	15	4.33	69.5
704	1.0	5.5	21	3.78	69.6
705	1.0	7.5	21	4.08	77.1
706	1.0	6.5	15	4.67	66.6
707	1.0	5.5	12	3.85	59.8
708	1.0	6.0	12	4.17	61.1
709	1.0	6.0	9	2.89	56.5
710	1.0	6.5	9	3.28	57.5
711	1.0	5.0	12	4.07	58.5
712	1.0	5.0	9	3.51	54.5
713	1.0	5.0	6	5.24	49.8
714	1.0	5.5	6	4.65	50.5
715	1.0	6.0	6	3.96	51.1
716	1.0	6.5	6	failed	
717	1.0	6.5	12	4.43	62.4
718	1.0	5.0	15	4.85	61.9

Continued on next page

Table continued from previous page

Number	$\mathcal{R}$	$\Upsilon_B\,(M_\odot/L_\odot)$	ϕ	χ^2/N	$\Omega_p\,(\mathrm{km\ s^{-1}\ kpc^{-1}})$
719	1.0	6.0	15	4.38	65.1
720	1.0	5.0	18	5.00	64.9
721	1.0	5.5	18	3.96	66.7
722	1.0	6.0	18	4.03	68.5
723	1.0	6.5	18	3.88	70.2
724	1.0	4.5	6	5.61	49.2
725	1.0	4.5	9	4.20	53.4
726	1.0	4.5	12	4.82	57.1
727	1.0	4.5	15	5.15	60.3
411	1.2	6.5	24	4.36	60.4
412	1.2	7.5	18	4.28	58.0
413	1.2	6.5	30	4.86	63.8
414	1.2	6.5	18	3.06	56.2
415	1.2	5.5	18	4.46	54.3
416	1.2	6.5	12	3.56	51.0
417	1.2	7.5	12	4.84	52.1
418	1.2	7.5	24	4.41	62.6
419	1.2	5.5	12	4.18	49.8
420	1.2	5.5	24	4.62	58.0
421	1.2	6.0	24	4.63	59.2
422	1.2	6.0	18	3.35	55.3
423	1.2	6.0	12	3.73	50.4
424	1.2	7.0	24	5.50	61.5
425	1.2	7.0	18	3.65	57.1

Continued on next page

Table continued from previous page

Number	$\mathcal{R}$	$\Upsilon_B\,(M_\odot/L_\odot)$	ϕ	χ^2/N	$\Omega_p\,(\mathrm{km\ s^{-1}\ kpc^{-1}})$
426	1.2	7.0	12	3.25	51.6
427	1.2	5.5	15	3.49	52.2
428	1.2	6.0	15	2.99	53.0
429	1.2	6.5	15	2.66	53.7
430	1.2	7.0	15	2.68	54.5
431	1.2	7.5	15	3.32	55.2
432	1.2	5.5	21	4.13	56.3
433	1.2	6.0	21	3.88	57.3
434	1.2	6.5	21	4.10	58.4
435	1.2	7.0	21	4.65	59.4
436	1.2	7.5	21	5.62	60.4
600	1.4	5.5	24	4.69	50.9
601	1.4	6.5	24	3.50	52.2
602	1.4	7.5	24	3.75	53.6
603	1.4	5.5	18	4.79	48.4
604	1.4	6.5	18	4.08	49.3
605	1.4	7.5	18	4.65	50.3
606	1.4	5.5	12	4.48	45.4
607	1.4	6.5	12	4.38	45.8
608	1.4	7.5	12	5.40	46.3
609	1.4	5.5	30	4.95	53.0
610	1.4	6.5	30	3.70	54.7
611	1.4	7.5	30	4.54	56.3
612	1.4	5.5	36	4.88	54.9

Continued on next page

Table continued from previous page

Number	$\mathcal{R}$	$\Upsilon_B\,(M_\odot/L_\odot)$	ϕ	χ^2/N	$\Omega_p\,(\mathrm{km\ s^{-1}\ kpc^{-1}})$
613	1.4	6.5	36	3.80	56.8
614	1.4	7.5	36	5.40	58.7
615	1.4	8.5	30	6.60	57.9
616	1.4	8.5	36	7.06	60.5
617	1.4	6.0	18	4.43	48.9
618	1.4	7.0	18	4.32	49.8
619	1.4	5.5	21	4.78	49.7
620	1.4	6.0	21	4.37	50.3
621	1.4	6.5	21	3.76	50.9
622	1.4	7.0	21	3.71	51.4
623	1.4	7.5	21	4.06	52.0
624	1.4	6.0	24	4.02	51.6
625	1.4	7.0	24	3.34	52.9
626	1.4	5.5	27	4.57	52.0
627	1.4	6.0	27	3.89	52.8
628	1.4	6.5	27	3.47	53.5
629	1.4	7.0	27	3.37	54.2
630	1.4	7.5	27	3.86	55.0
631	1.4	6.0	30	4.37	53.9
632	1.4	7.0	30	3.76	55.5

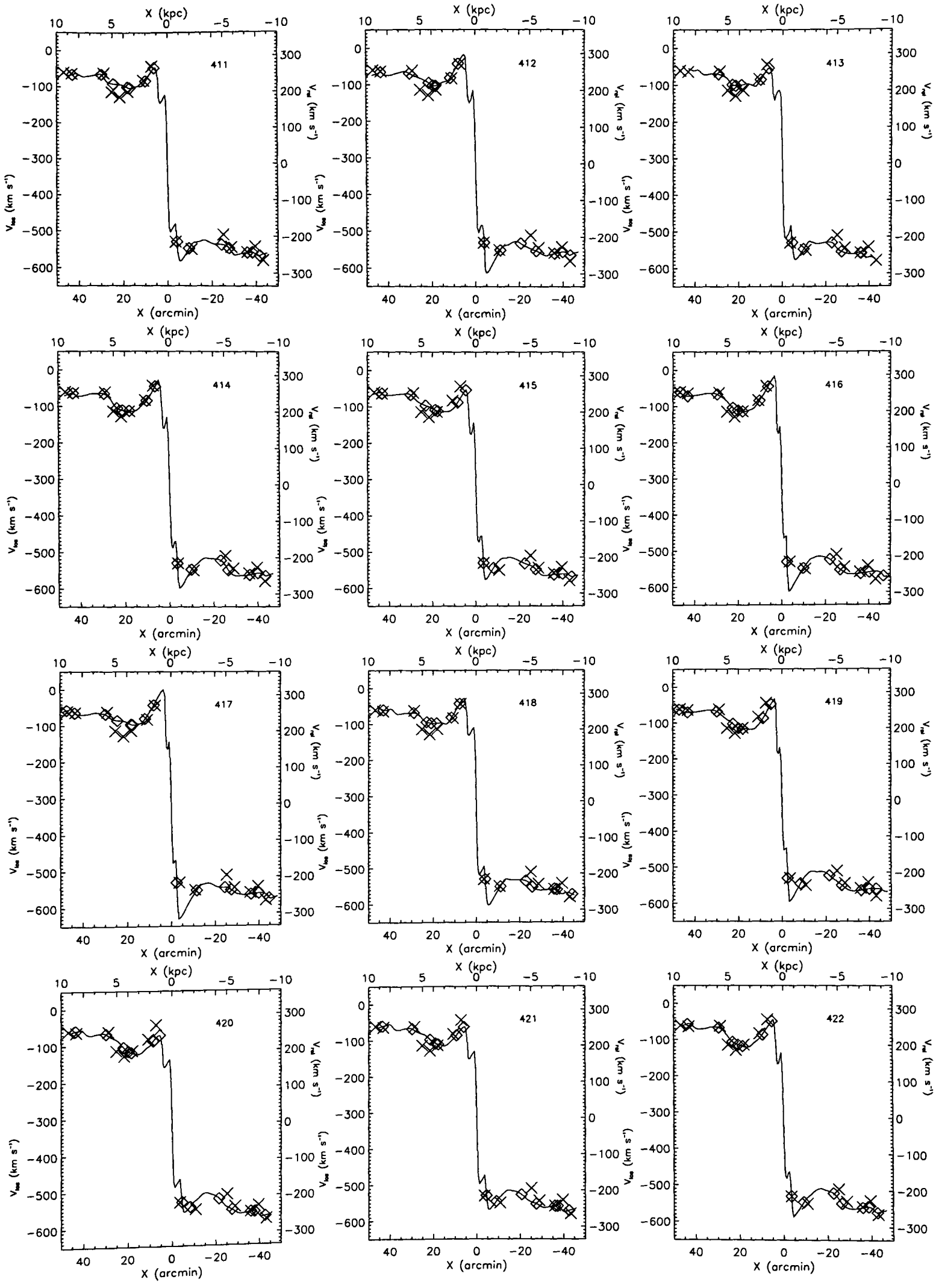
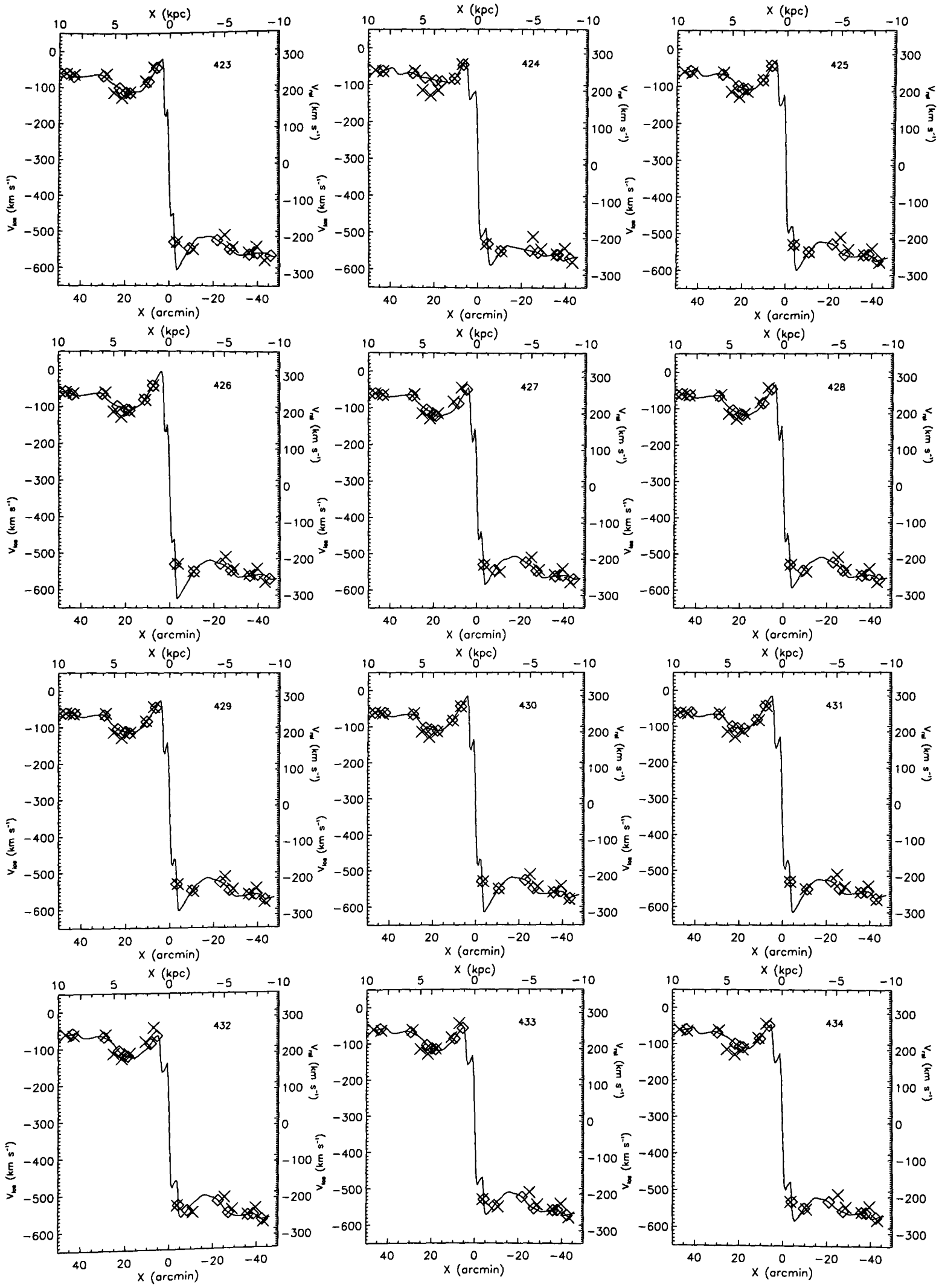
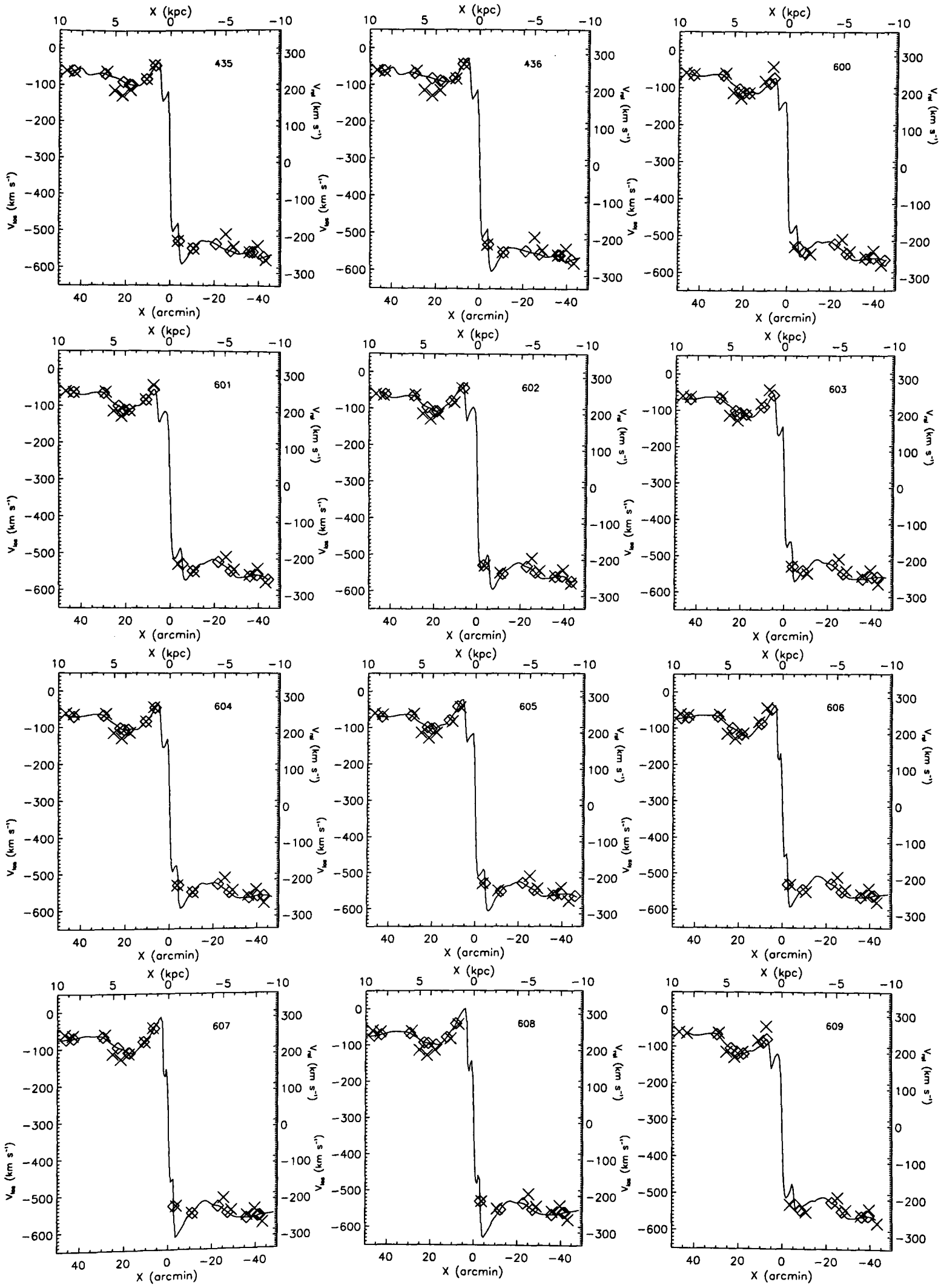
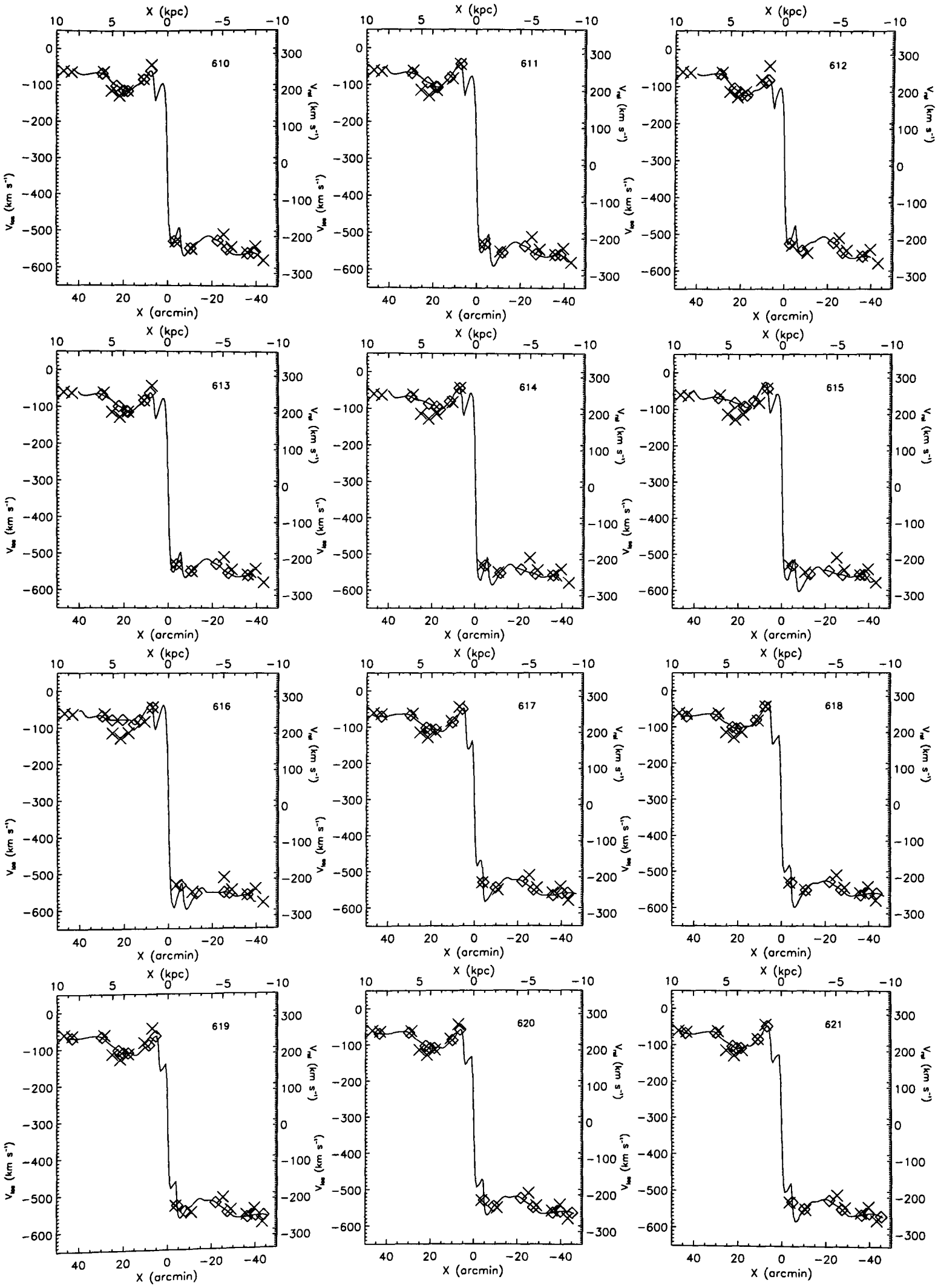
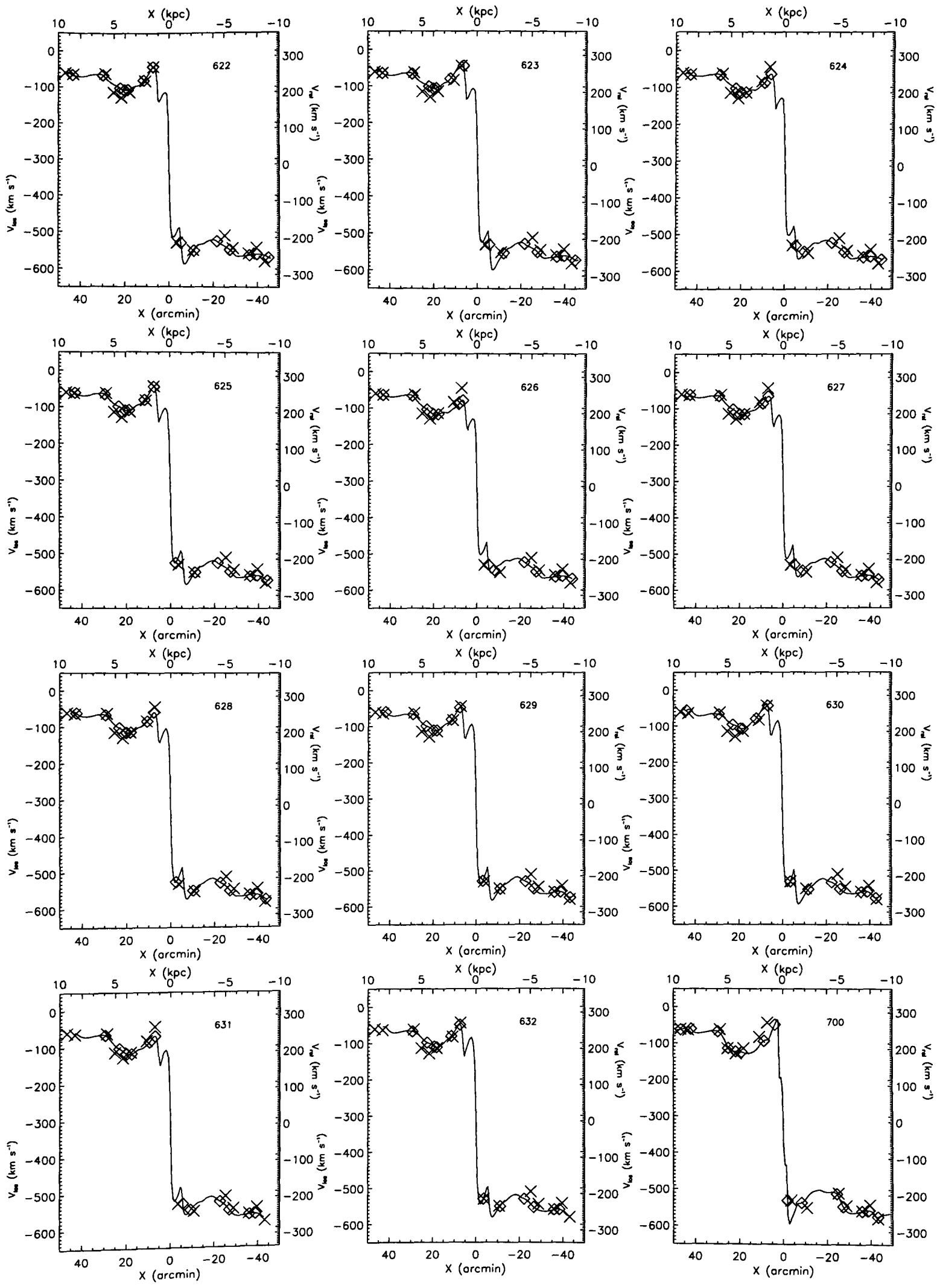


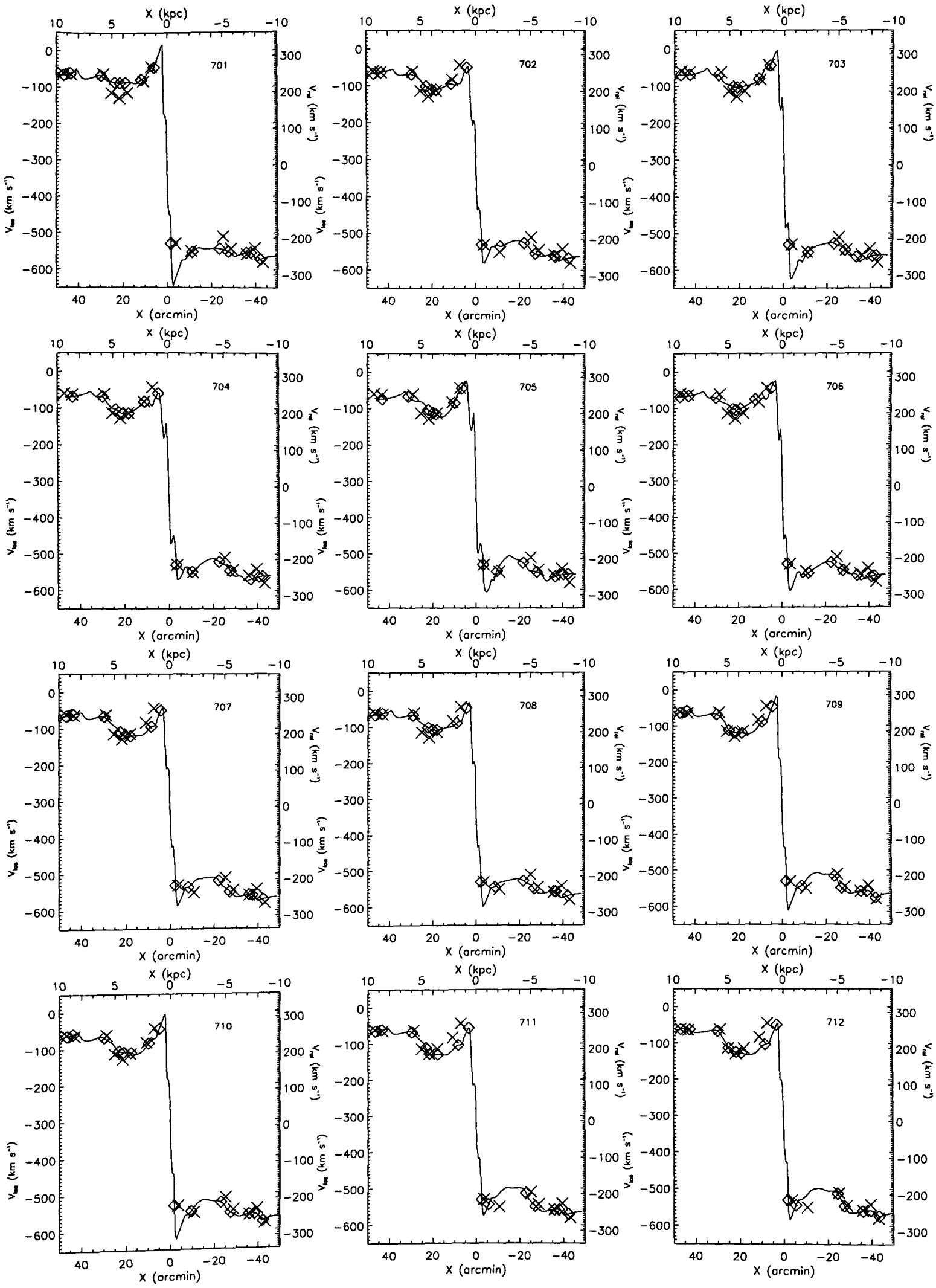
Figure B.1: Comparison with CO observations on major axis for all M31 bulge models.

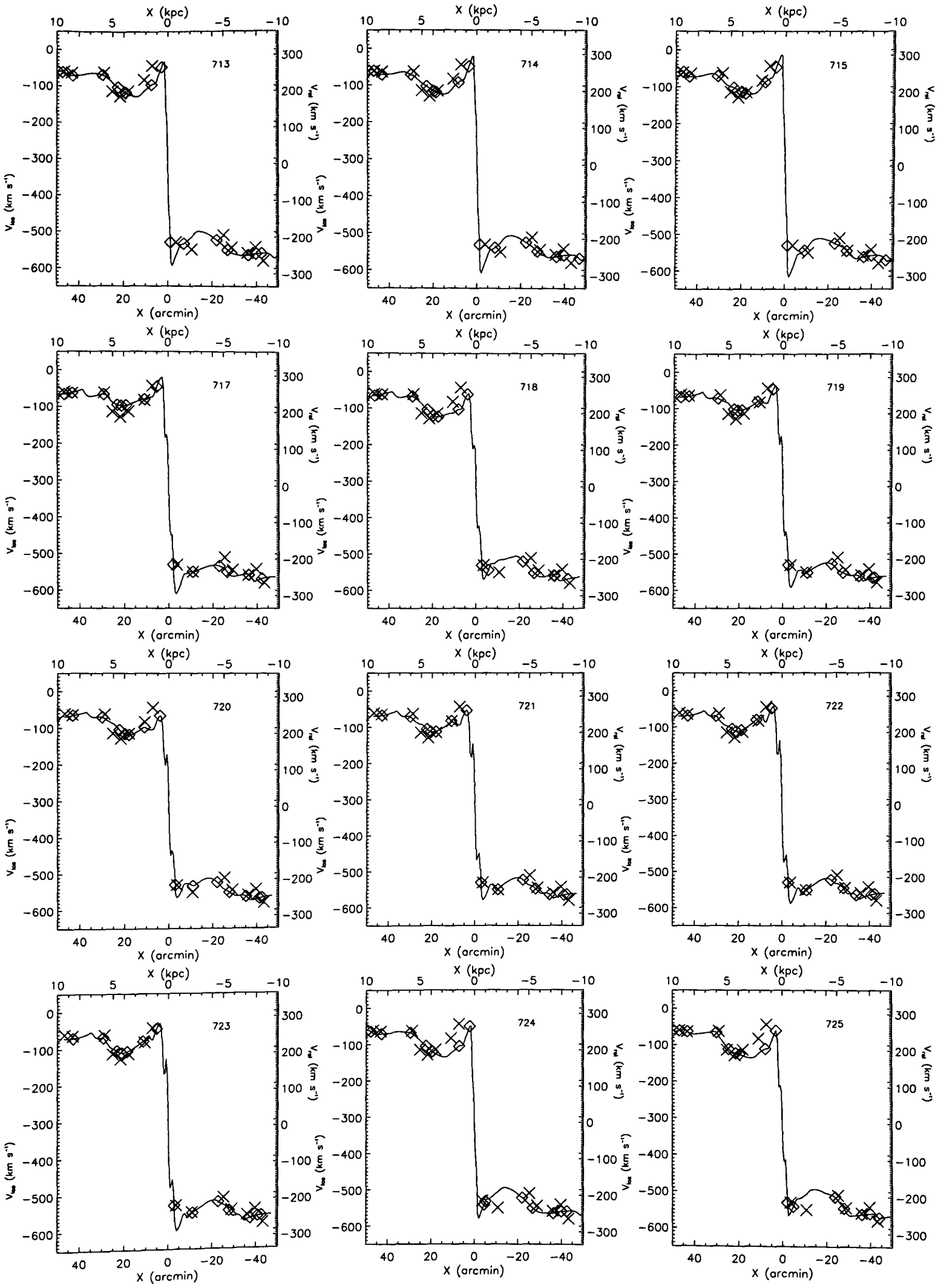

X (kpc)
X (arcmin)
V_{los} (km s^{-1})
V_{rot} (km s^{-1})
423
424
425
426
427
428
429
430
431
432
433
434

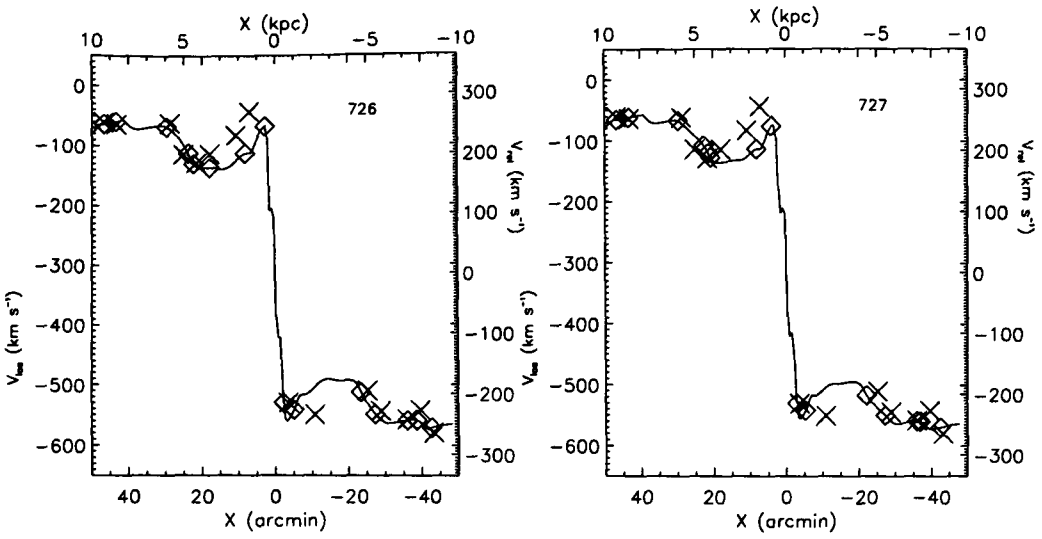

435
436
600
601
602
603
604
605
606
607
608
609
X (kpc)
X (arcmin)
V_{los} (km s^{-1})
V_{rot} (km s^{-1})


610
611
612
613
614
615
616
617
618
619
620
621
X (kpc)
X (arcmin)
V_{los} (km s^{-1})
V_{rot} (km s^{-1})


622
623
624
625
626
627
628
629
630
631
632
700
X (kpc)
X (arcmin)
V_{los} (km s^{-1})
V_{rot} (km s^{-1})


701
702
703
704
705
706
707
708
709
710
711
712
X (kpc)
X (arcmin)
V_{los} (km s^{-1})
V_{rot} (km s^{-1})


713
714
715
717
718
719
720
721
722
723
724
725
X (kpc)
X (arcmin)
V_{los} (km s^{-1})
V_{rot} (km s^{-1})


X (kpc)
10
5
0
−5
−10
726
727
0
−100
−200
−300
−400
−500
−600
300
200
100
0
−100
−200
−300
V_{sys} (km s^{-1})
40
20
0
−20
−40
X (arcmin)

Appendix C

Description of dual morphology models

Galaxy	NGC 2997	NGC 3223	NGC 5085
R_{min}	34″	15″	25″
R_{max}	122″	48″	50″
ψ_{sta}	26.6	9.46	38.7
ψ_{gas}	14.0	9.27	33.7
$\mathcal{F}_{\mathrm{sta}}$	0.44	0.57	0.47
$\mathcal{F}_{\mathrm{gas}}$	0.48	0.62	0.56
$\mathcal{A}_{\mathrm{sta}}$	0.29	0.24	0.07
$\mathcal{A}_{\mathrm{gas}}$	0.27	0.44	0.29
$\mathcal{C}$	1.49	1.11	1.38

Table C.1: Results are shown for a selection of real galaxies, including the length of the bar used as the minimum radius in logarithmic Fourier decompositions R_{min}, the pitch angle of the stellar spiral arms ψ_{sta}, the pitch angle of the gaseous spiral arms ψ_{gas}, the stellar flocculence $\mathcal{F}_{\mathrm{sta}}$, the gaseous flocculence $\mathcal{F}_{\mathrm{gas}}$, the stellar asymmetry $\mathcal{A}_{\mathrm{sta}}$, the gaseous asymmetry $\mathcal{A}_{\mathrm{gas}}$ and the stellar arm-inter arm contrast $\mathcal{C}$.

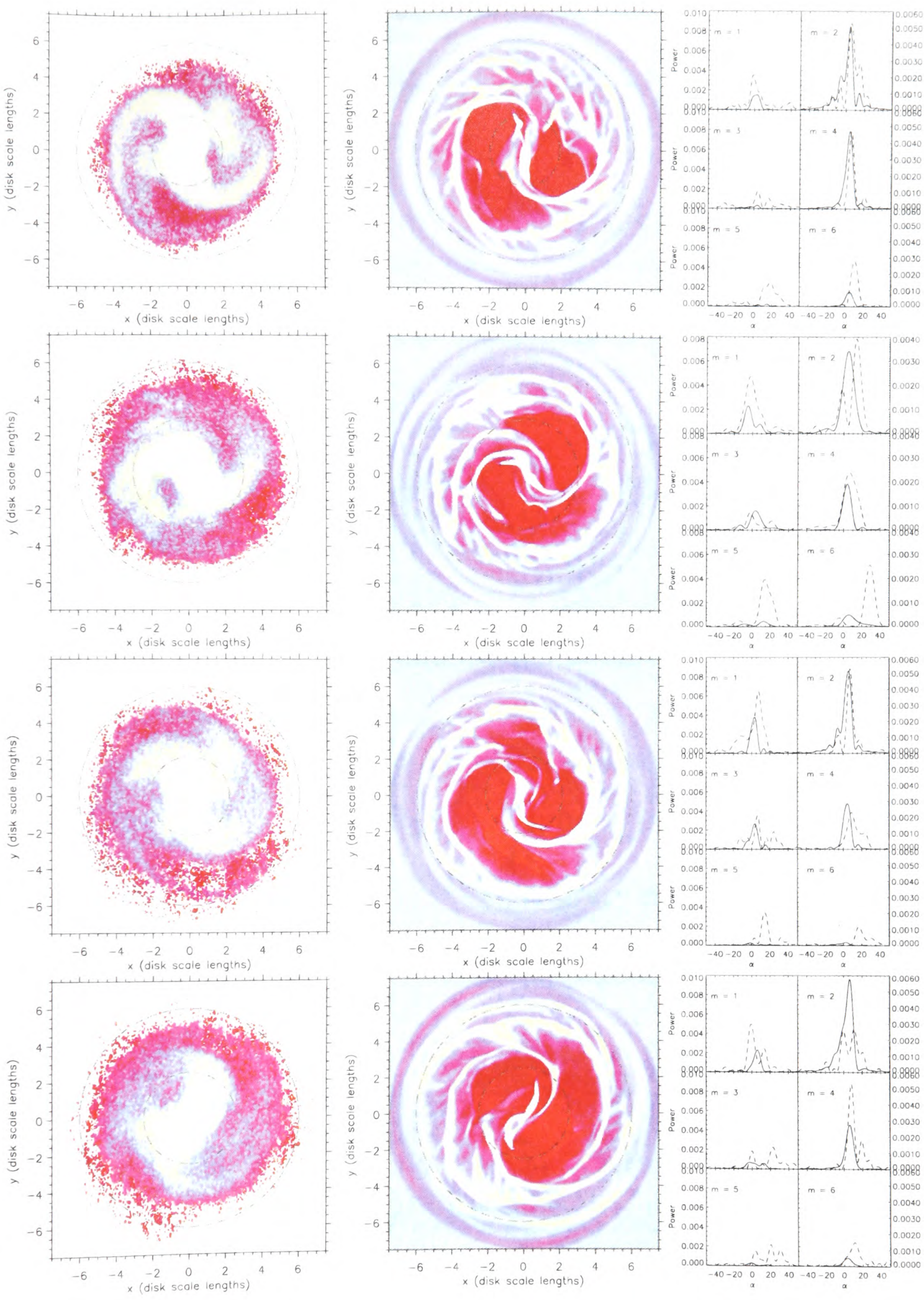


Figure C.1: Density fields and Fourier spectra for **sta13** and **gas13** at 80, 96, 108 and 120 time units or 4.1, 5.0, 5.6 and 6.2 bar rotation periods. In the Fourier spectra, solid lines indicate the stars and dashed lines the gas.

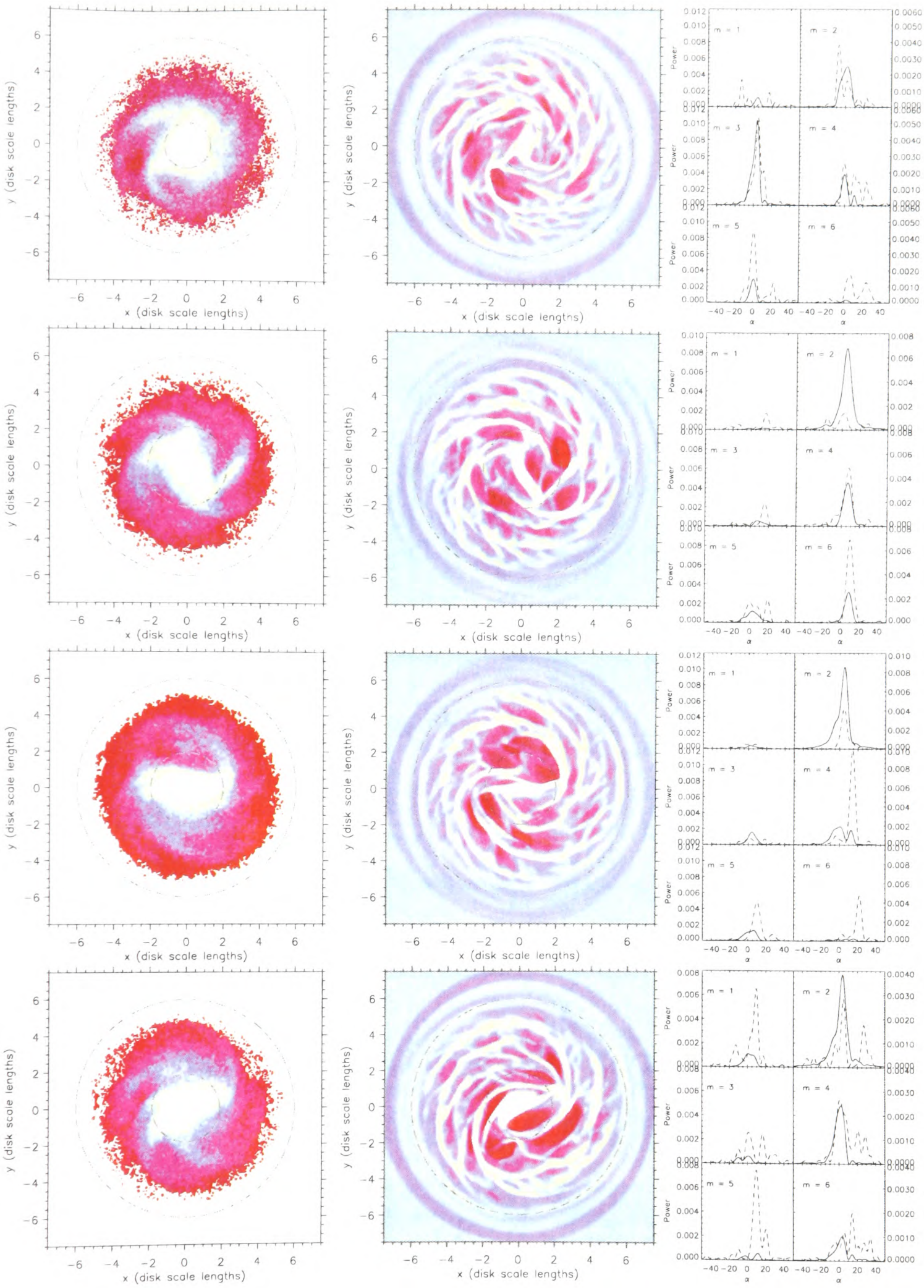


Figure C.2: Density fields and Fourier spectra for `sta16` and `gas16` at 80, 100, 108 and 118 time units or 3.4, 4.2, 4.6 and 5.0 bar rotation periods. In the Fourier spectra, solid lines indicate the stars and dashed lines the gas.

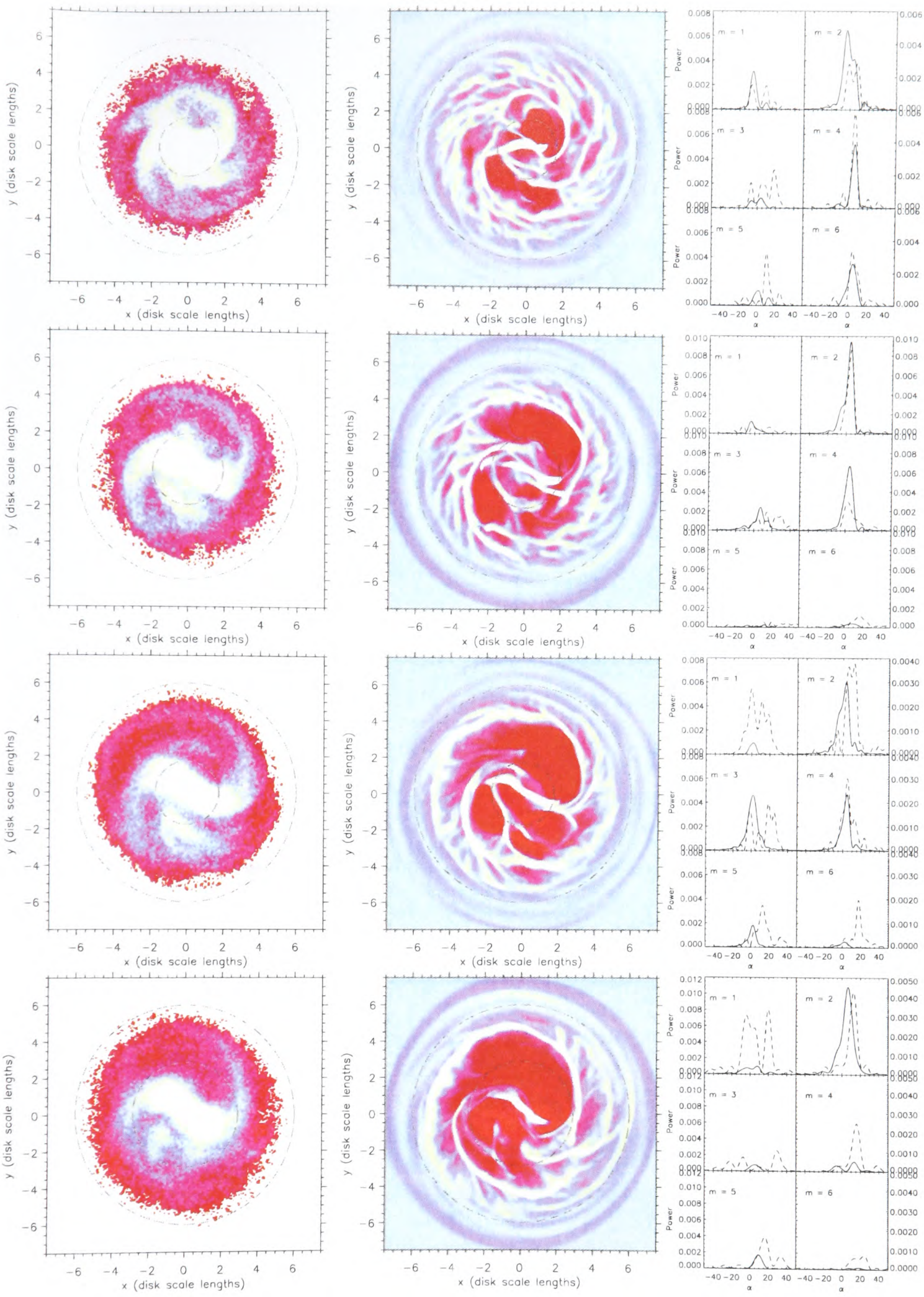


Figure C.3: Density fields and Fourier spectra for **sta17** and **gas17** at 58, 76, 94 and 112 time units or 3.3, 4.3, 5.3 and 6.3 bar rotation periods. In the Fourier spectra, solid lines indicate the stars and dashed lines the gas.

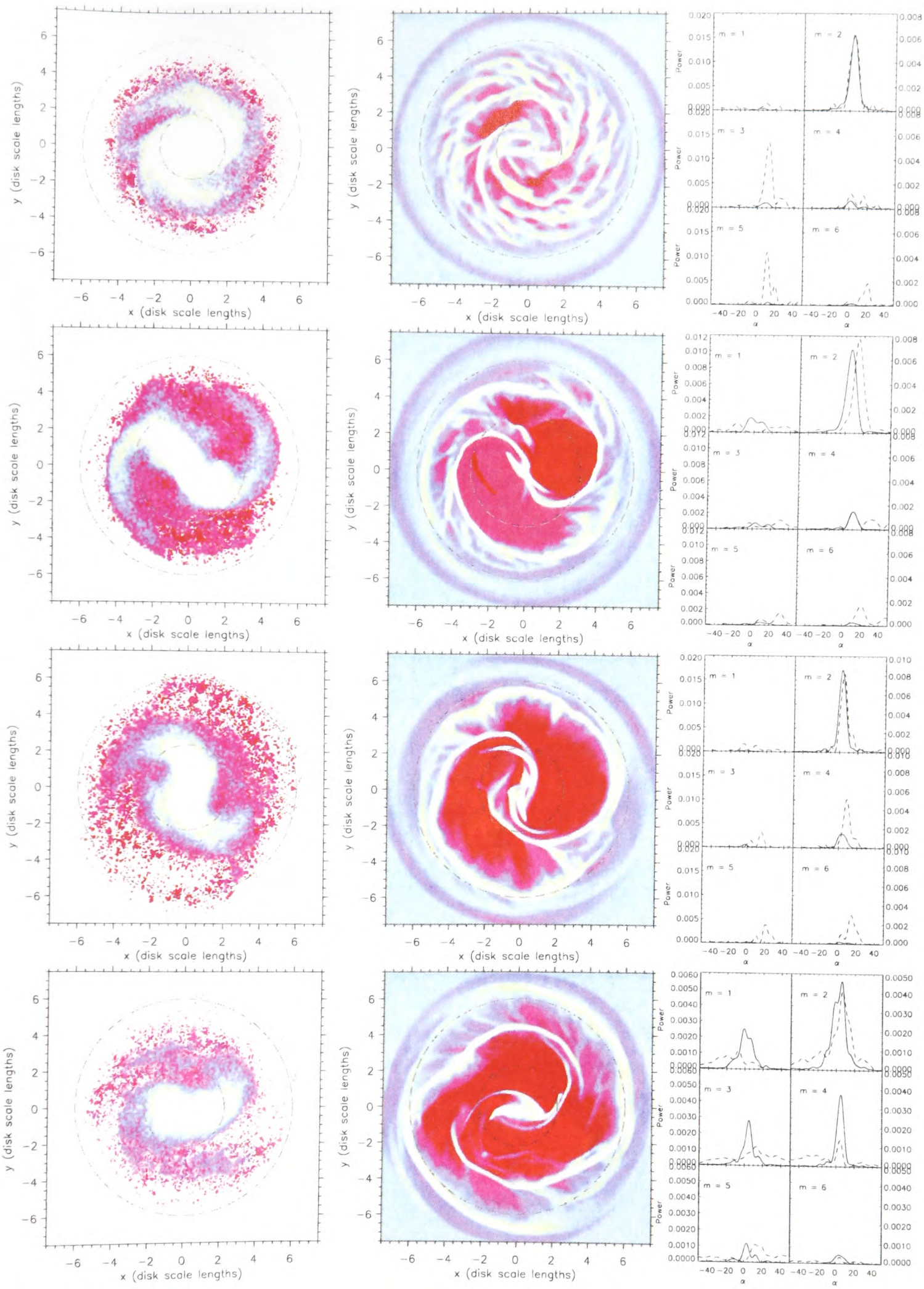


Figure C.4: Density fields and Fourier spectra for **sta19** and **gas19** at 44, 80, 94 and 110 time units or 2.0, 3.7, 4.4 and 5.1 bar rotation periods. In the Fourier spectra, solid lines indicate the stars and dashed lines the gas.

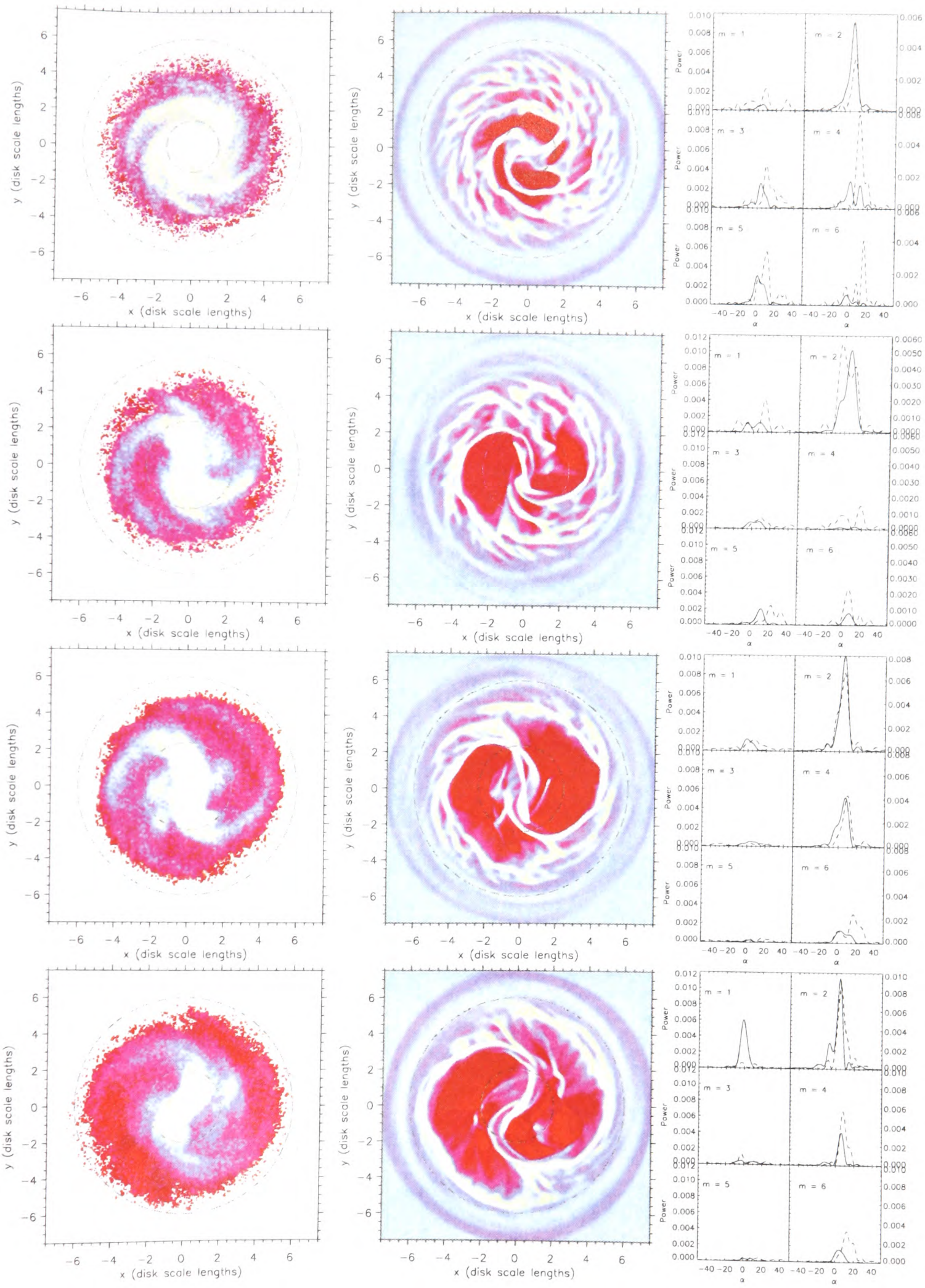


Figure C.5: Density fields and Fourier spectra for **sta21** and **gas21** at 38, 60, 70 and 82 time units or 1.9, 3.1, 3.6 and 4.2 bar rotation periods. In the Fourier spectra, solid lines indicate the stars and dashed lines the gas.

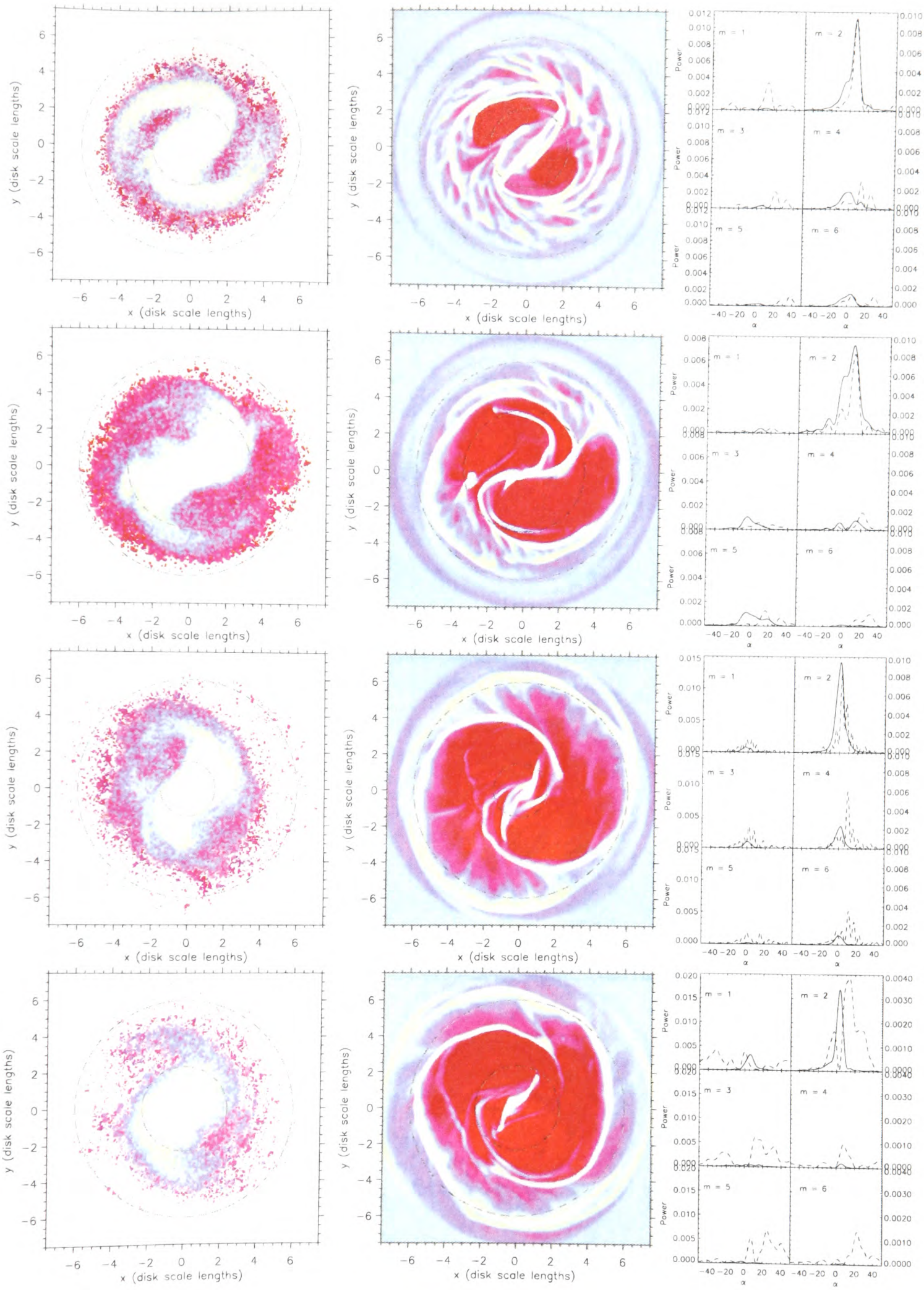


Figure C.6: Density fields and Fourier spectra for **sta26** and **gas26** at 54, 78, 102 and 126 time units or 1.9, 3.1, 3.6 and 4.2 bar rotation periods. In the Fourier spectra, solid lines indicate the stars and dashed lines the gas.

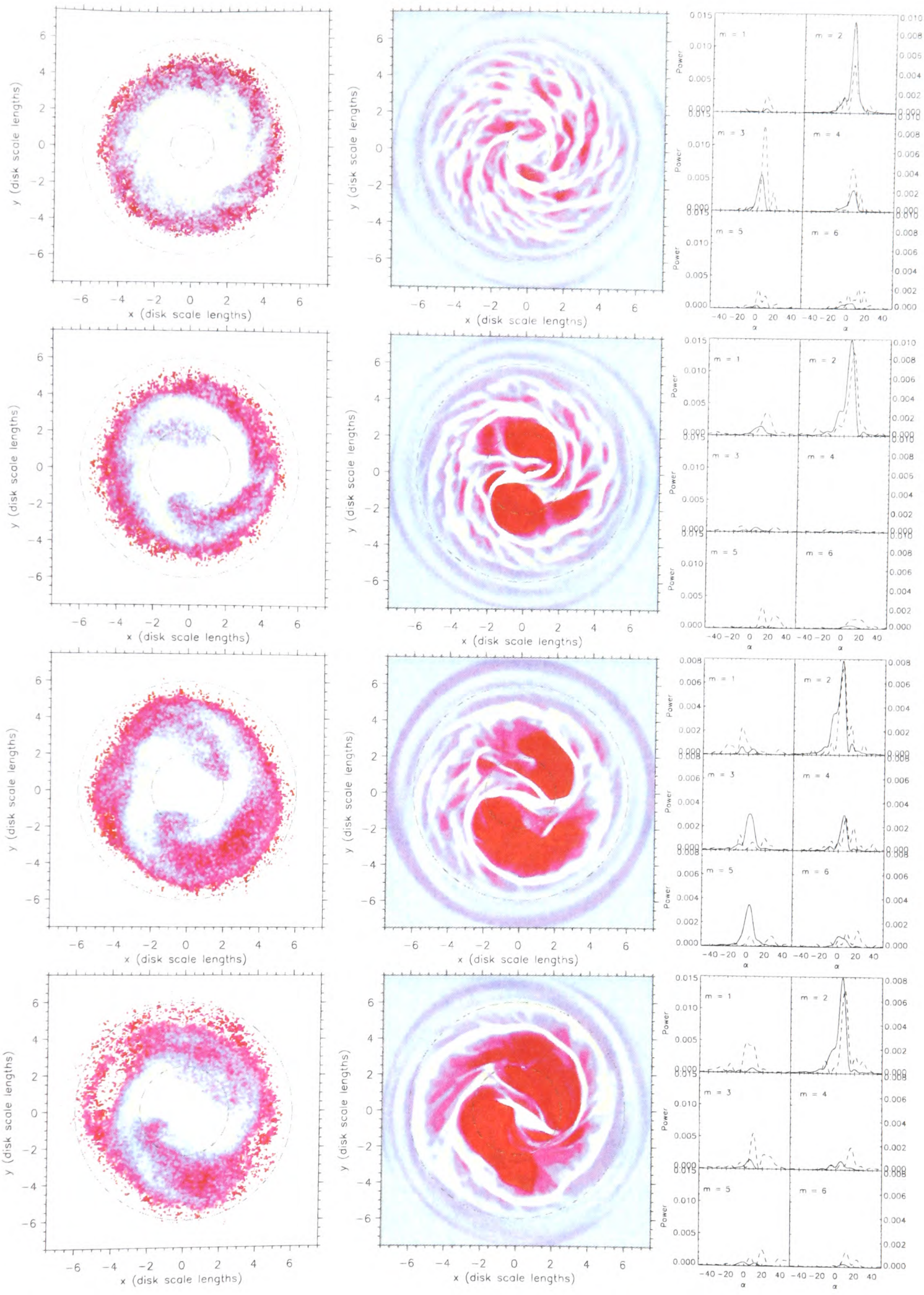


Figure C.7: Density fields and Fourier spectra for **sta28** and **gas28** at 54, 94, 114 and 134 time units or 2.9, 5.0, 6.1 and 7.1 bar rotation periods. In the Fourier spectra, solid lines indicate the stars and dashed lines the gas.

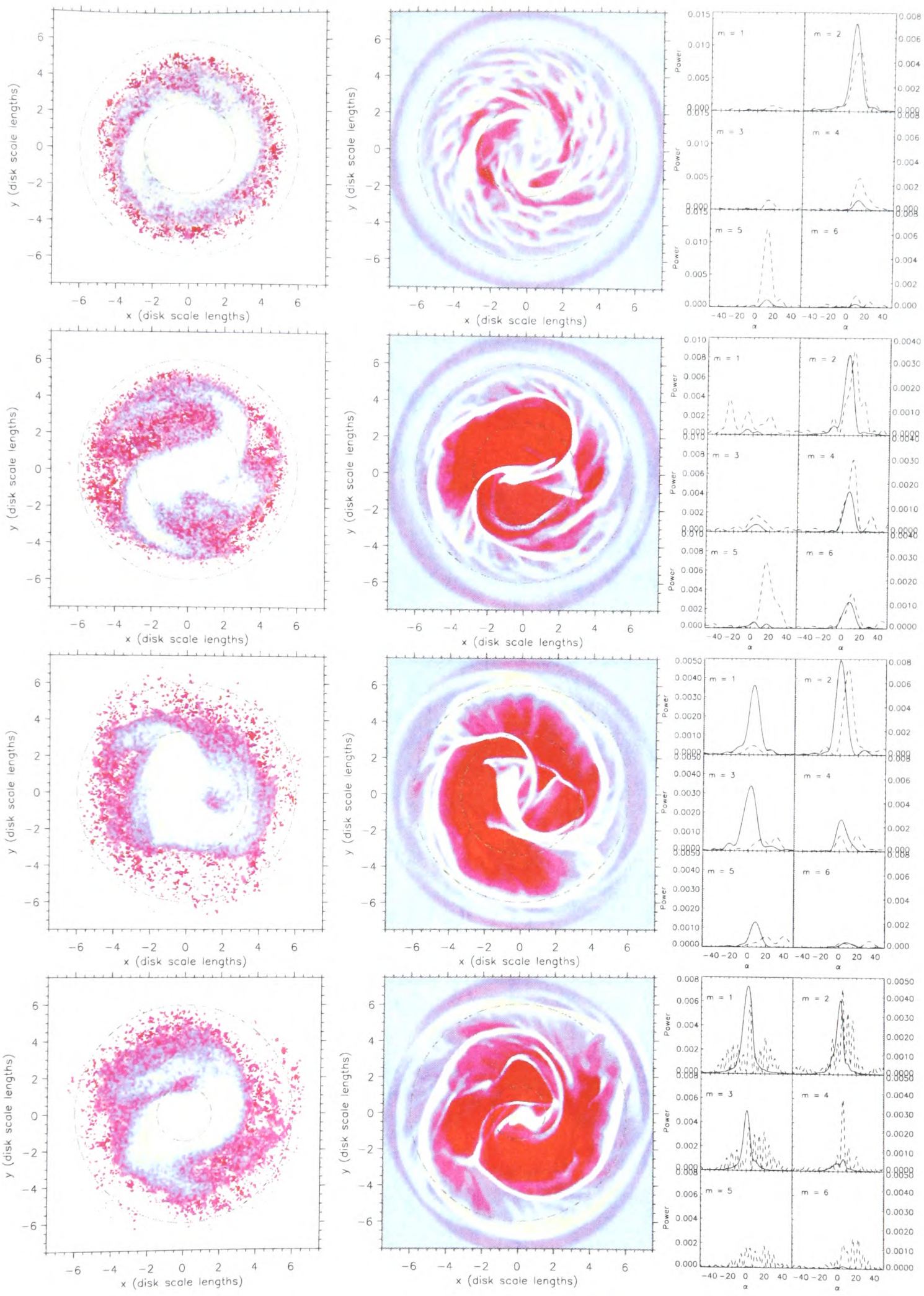


Figure C.8: Density fields and Fourier spectra for `sta30` and `gas30` at 44, 82, 100 and 116 time units or 2.0, 3.7, 4.6 and 5.3 bar rotation periods. In the Fourier spectra, solid lines indicate the stars and dashed lines the gas.

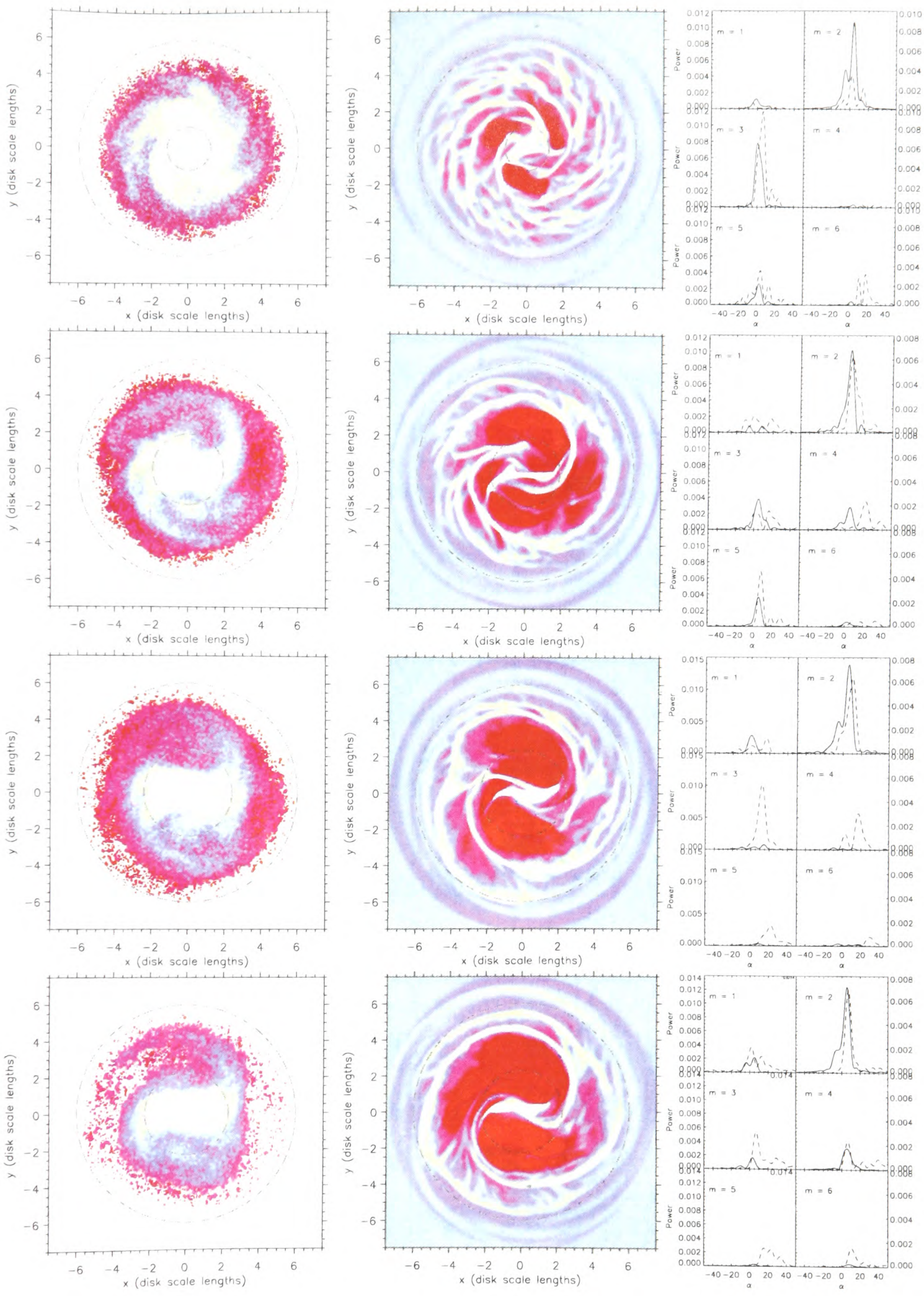


Figure C.9: Density fields and Fourier spectra for `sta36` and `gas36` at 56, 96, 116 and 136 time units or 2.8, 4.8, 5.8 and 6.8 bar rotation periods. In the Fourier spectra, solid lines indicate the stars and dashed lines the gas.

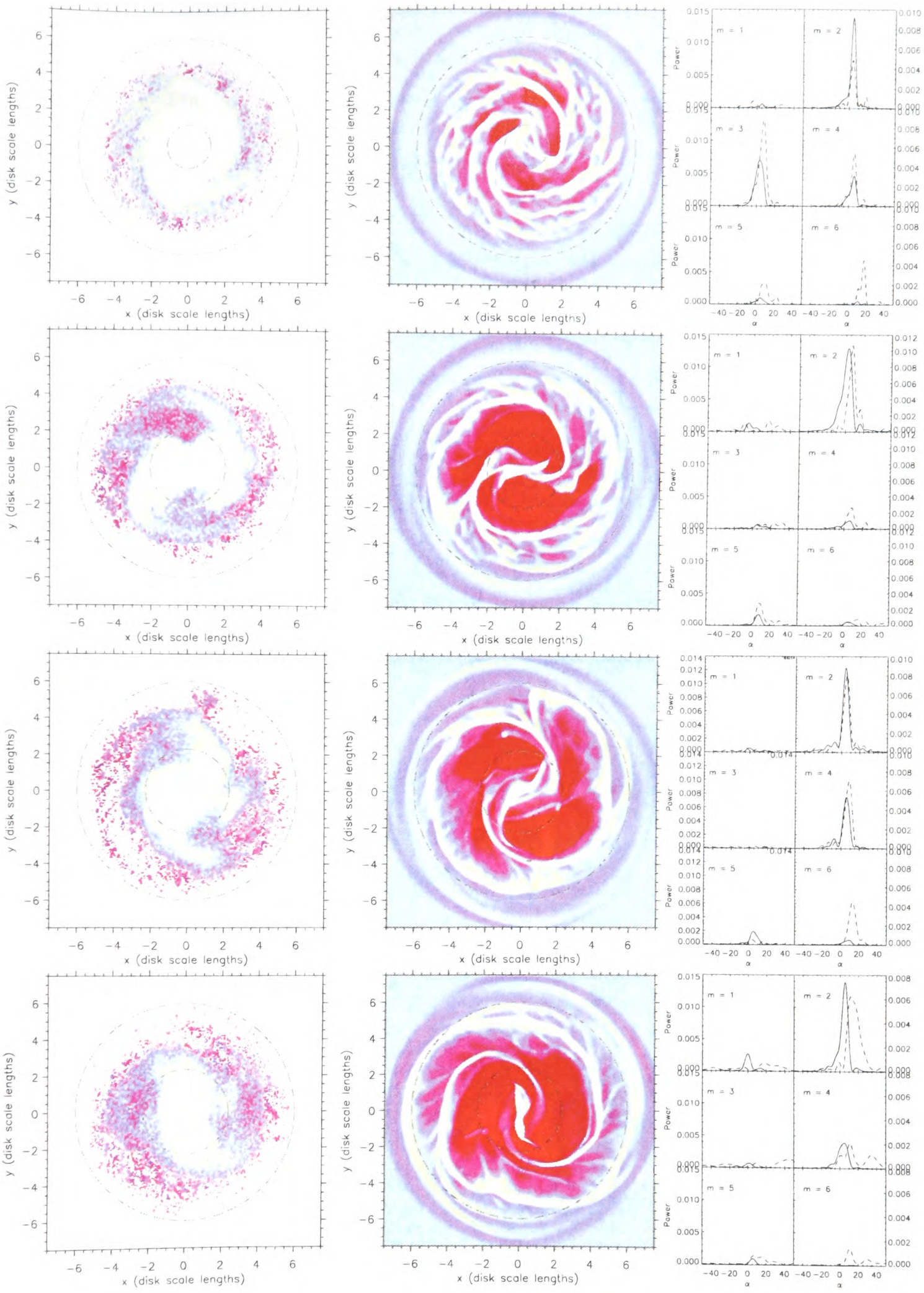


Figure C.10: Density fields and Fourier spectra for `sta38` and `gas38` at 40, 80, 100 and 118 time units or 1.9, 3.7, 4.7 and 5.5 bar rotation periods. In the Fourier spectra, solid lines indicate the stars and dashed lines the gas.

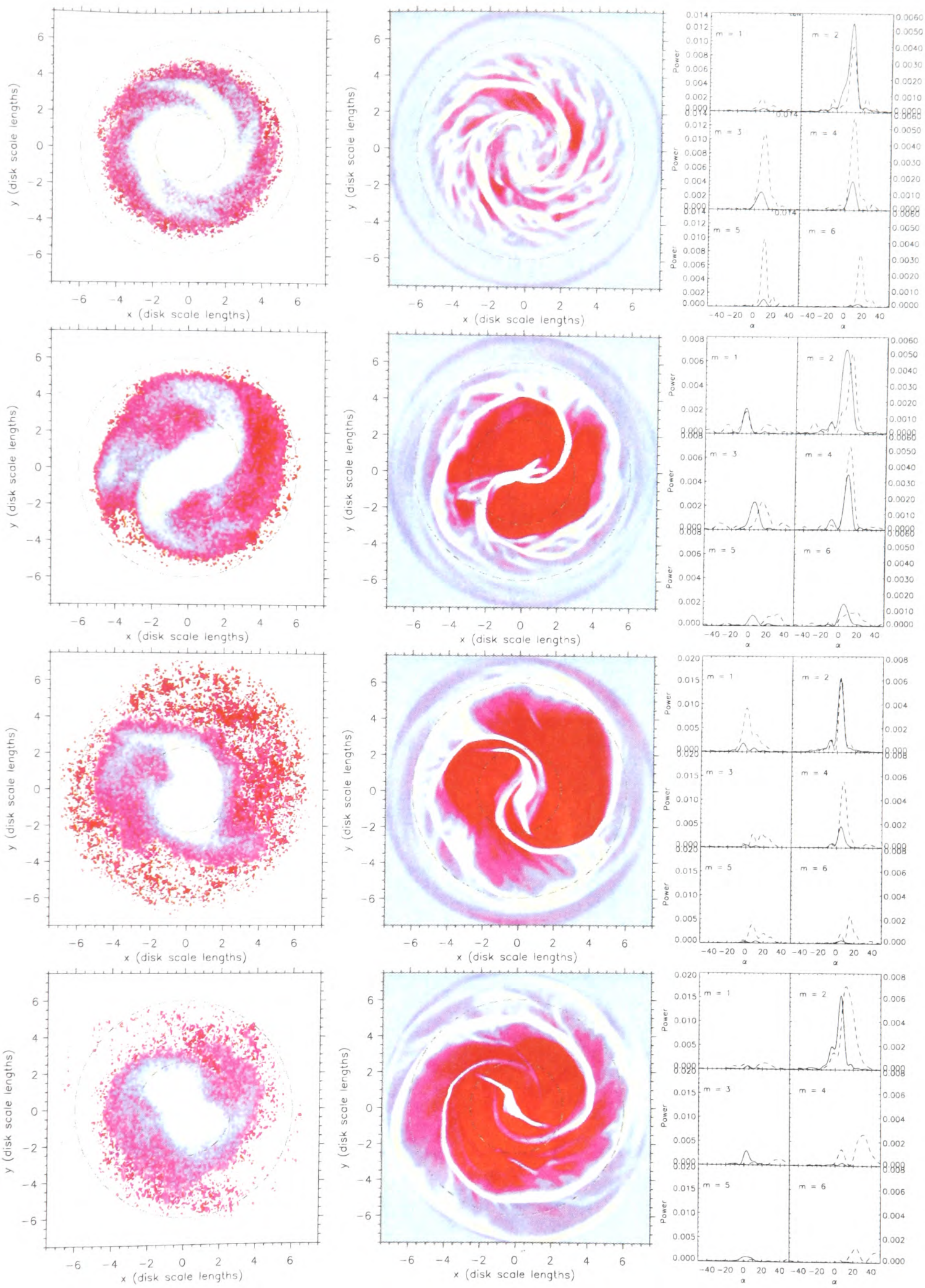


Figure C.11: Density fields and Fourier spectra for **sta40** and **gas40** at 32, 72, 92 and 112 time units or 1.4, 3.1, 4.0 and 4.9 bar rotation periods. In the Fourier spectra, solid lines indicate the stars and dashed lines the gas.

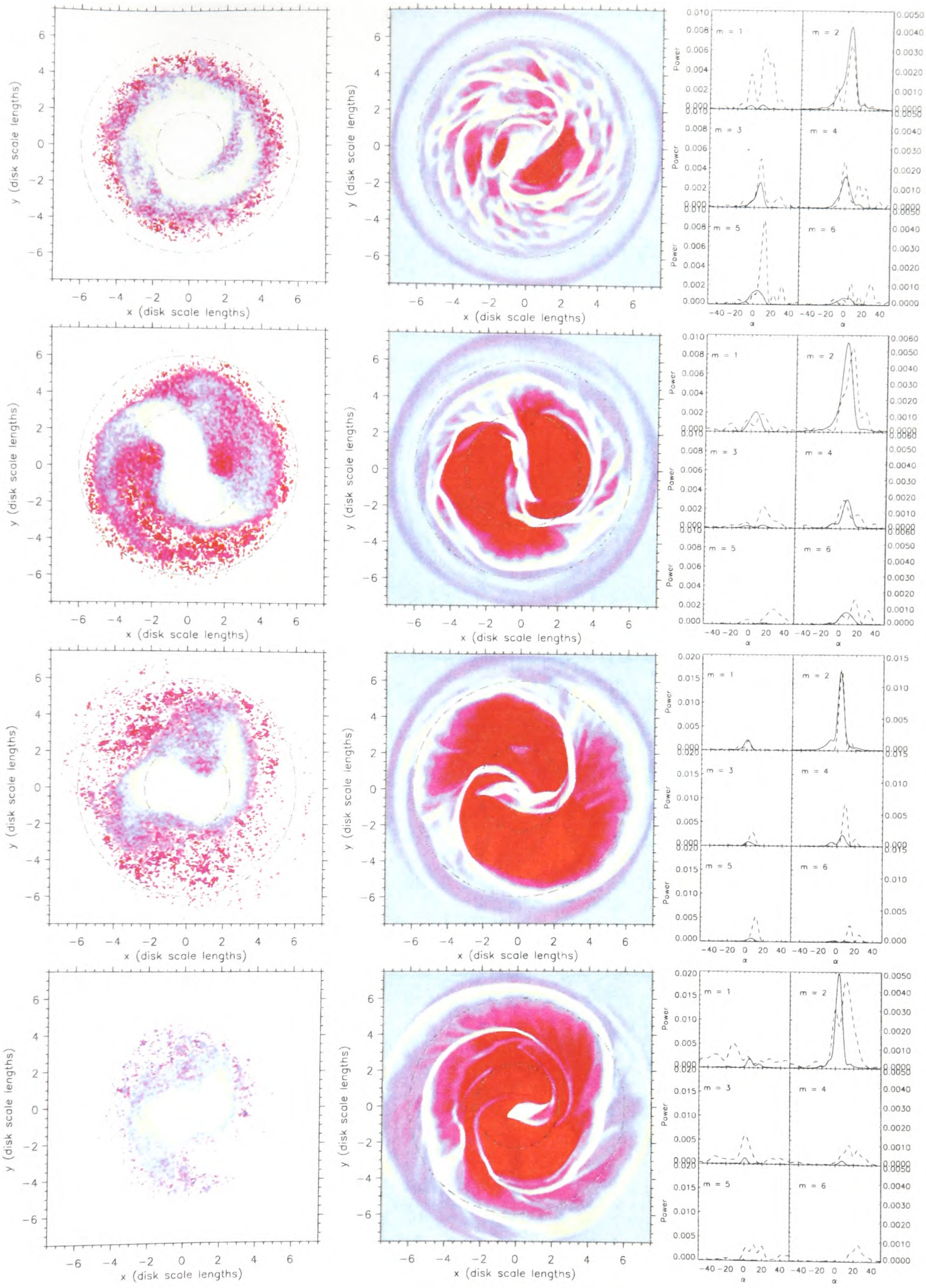


Figure C.12: Density fields and Fourier spectra for **sta42** and **gas42** at 48, 78, 98 and 118 time units or 2.1, 3.5, 4.3 and 5.2 bar rotation periods. In the Fourier spectra, solid lines indicate the stars and dashed lines the gas.

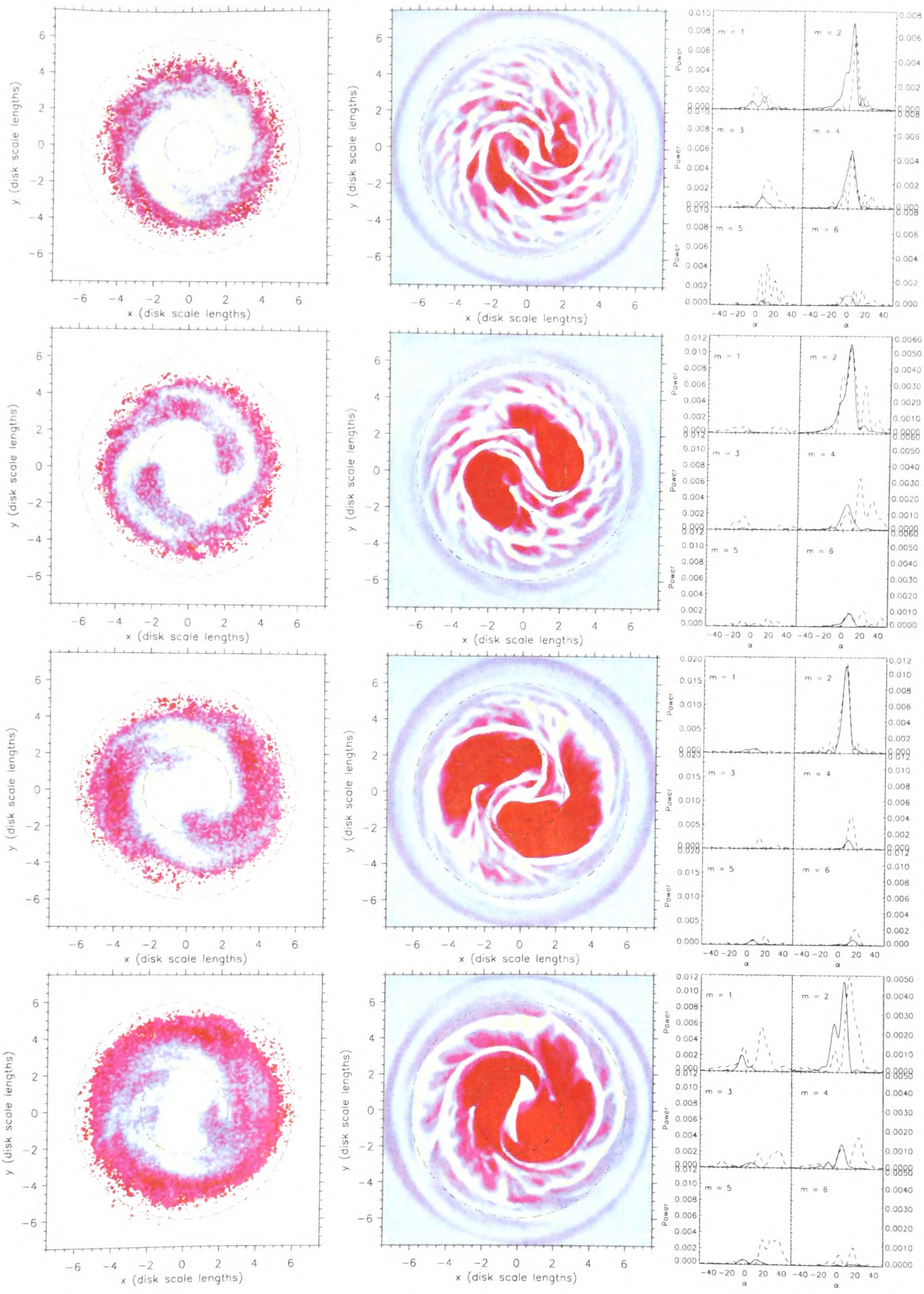


Figure C.13: Density fields and Fourier spectra for **sta43** and **gas43** at 44, 60, 74 and 90 time units or 2.4, 3.3, 4.1 and 5.0 bar rotation periods. In the Fourier spectra, solid lines indicate the stars and dashed lines the gas.

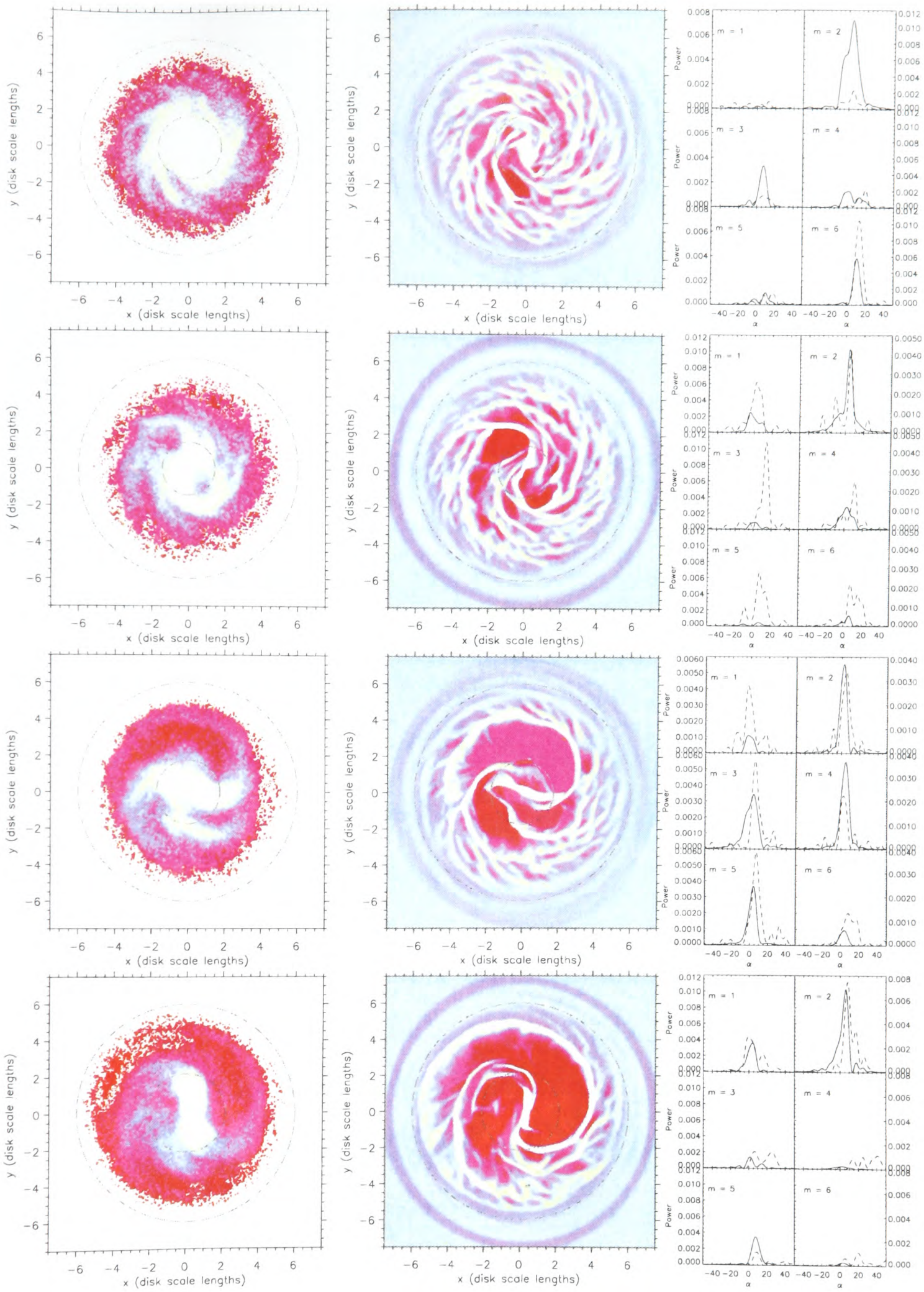


Figure C.14: Density fields and Fourier spectra for **sta44** and **gas44** at 60, 80, 100 and 120 time units or 3.6, 4.8, 6.0 and 7.2 bar rotation periods. In the Fourier spectra, solid lines indicate the stars and dashed lines the gas.

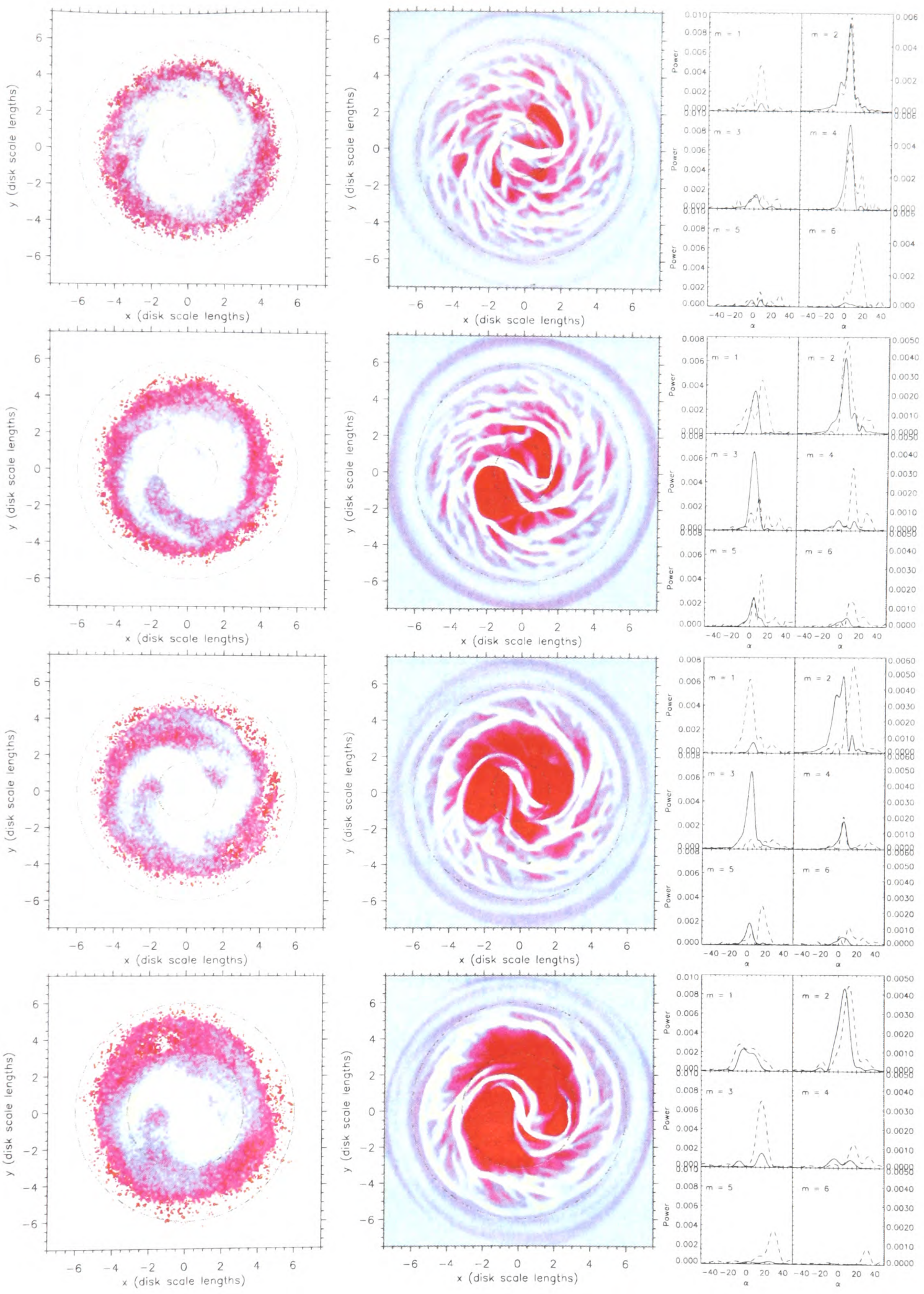


Figure C.15: Density fields and Fourier spectra for **sta45** and **gas45** at 52, 78, 104 and 130 time units or 3.1, 4.7, 6.2 and 7.8 bar rotation periods. In the Fourier spectra, solid lines indicate the stars and dashed lines the gas.

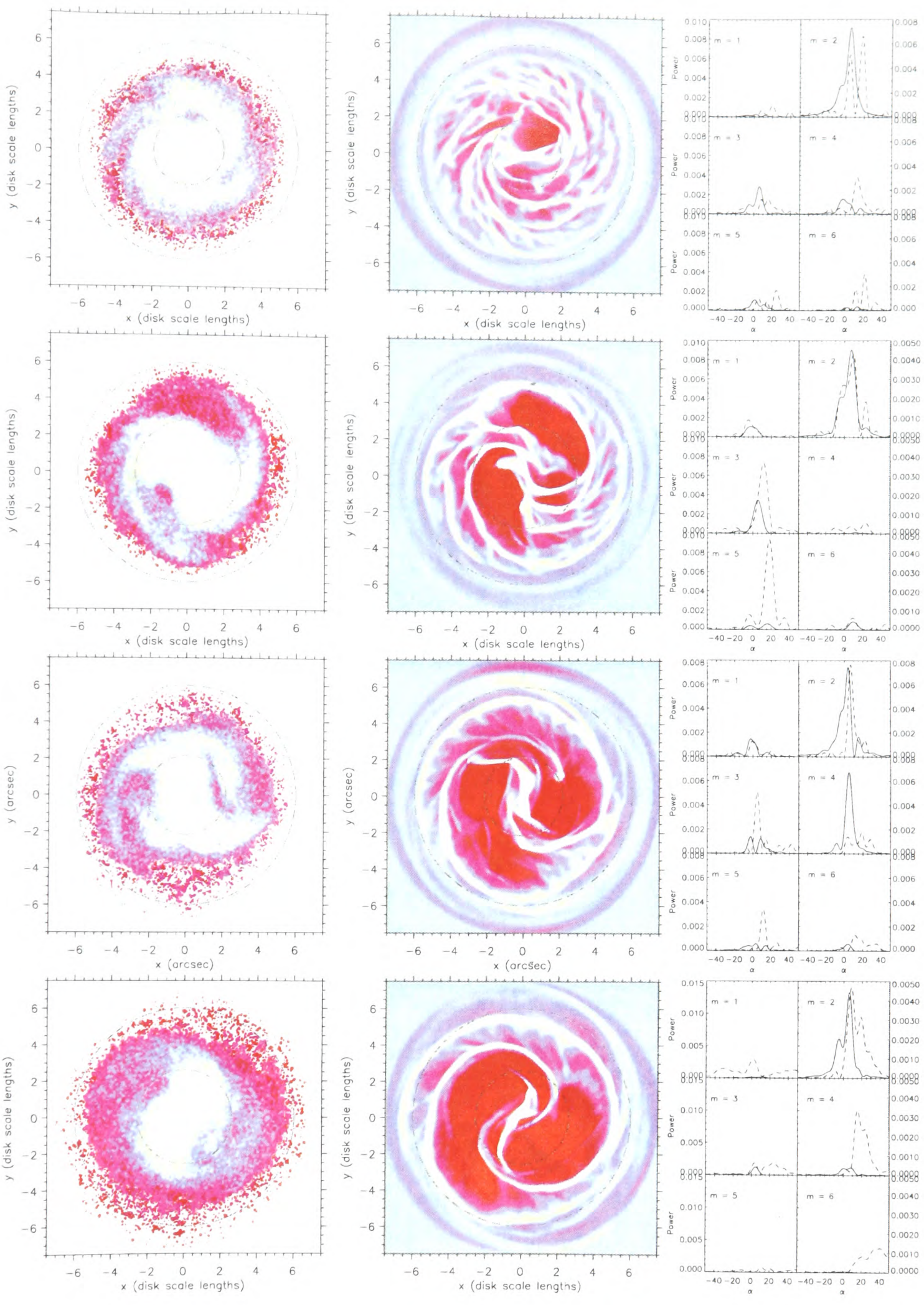


Figure C.16: Density fields and Fourier spectra for **sta46** and **gas46** at 80, 104, 128 and 154 time units or 3.7, 4.8, 5.9 and 7.0 bar rotation periods. In the Fourier spectra, solid lines indicate the stars and dashed lines the gas.

Run	13	16	17	19	21	26	28	30	36	38	40	42	43	44	45	46
f_d	0.400	0.400	0.350	0.450	0.400	0.500	0.370	0.470	0.425	0.450	0.500	0.475	0.350	0.330	0.350	0.425
$Q_{\rm min}$	1.193	1.256	1.193	1.193	1.130	1.200	1.196	1.204	1.193	1.170	1.170	1.182	1.170	1.194	1.225	1.225
z_0	0.120	0.133	0.108	0.131	0.108	0.144	0.114	0.138	0.126	0.126	0.137	0.134	0.104	0.104	0.114	0.133
$t_{\rm ini}$	48	50	38	39	30	39	47	41	48	40	36	39	30	40	39	52
$t_{\rm fin}$	60	59	56	55	41	63	67	58	68	59	56	59	45	60	65	80
$R_{\rm min}$	2.31	1.99	2.04	2.58	2.28	2.54	2.19	2.49	2.23	2.51	2.55	2.46	2.35	1.70	1.99	2.63
$R_{\rm co}$	2.33	2.96	2.23	2.53	2.36	2.81	2.36	2.53	2.38	2.51	2.61	2.61	2.31	2.11	2.06	2.61
$R_{\rm olr}$	4.16	5.02	4.09	4.41	4.21	4.84	4.21	4.41	4.21	4.39	4.52	4.52	4.16	3.91	3.84	4.52
$\mathcal{R}$	1.01	1.49	1.09	0.98	1.04	1.11	1.08	1.02	1.07	1.00	1.02	1.06	0.98	1.24	1.04	0.99
Ω_p	0.326	0.266	0.354	0.291	0.322	0.254	0.334	0.286	0.313	0.294	0.273	0.278	0.347	0.380	0.377	0.291
$\bar{Q}$	2.75	1.91	2.57	3.14	2.60	3.34	2.62	3.01	2.51	2.76	3.23	3.47	2.75	2.35	2.33	2.81
$\psi_{\rm sta}$	21.0	23.1	24.9	21.4	22.0	28.8	20.3	29.6	20.5	20.5	24.0	27.4	19.6	29.1	23.1	18.9
$\psi_{\rm gas}$	9.1	26.5	14.6	16.2	22.6	18.8	16.2	17.7	15.1	14.3	16.9	20.8	14.8	19.8	16.9	14.1
$\mathcal{F}_{\rm sta}$	0.57	0.55	0.60	0.53	0.54	0.54	0.53	0.56	0.53	0.51	0.53	0.54	0.53	0.63	0.60	0.53
$\mathcal{F}_{\rm gas}$	0.66	0.68	0.64	0.53	0.54	0.56	0.57	0.56	0.60	0.54	0.59	0.59	0.59	0.65	0.64	0.59
$\mathcal{A}_{\rm sta}$	0.28	0.18	0.35	0.18	0.21	0.19	0.23	0.35	0.21	0.13	0.22	0.24	0.12	0.38	0.38	0.18
$\mathcal{A}_{\rm gas}$	0.45	0.34	0.50	0.26	0.24	0.31	0.39	0.34	0.42	0.27	0.32	0.43	0.26	0.48	0.47	0.39
$\mathcal{C}$	1.70	1.68	1.73	2.10	1.97	2.19	1.85	2.05	1.65	1.90	1.96	2.09	2.01	1.78	1.62	1.88
$N_{\rm lost}$	1	0	1	21	0	662	9	45	26	8	229	605	0	0	0	37
$C_{\rm mass}$	0.09	0.04	0.05	0.12	0.06	0.12	0.03	0.15	0.10	0.11	0.09	0.18	0.08	0.07	0.05	0.09

Table C.2: Results are shown for all of the simulations, including the disk mass fraction at disk four scale lengths f_d, the minimum value of the initial Q parameter $Q_{\rm min}$, the effective scale height of the stellar particles z_0, the initial and final time slices defining the set of contiguous two armed stellar spiral structure $t_{\rm ini}$ and $t_{\rm fin}$, the time averaged length of the bar used as the minimum radius in logarithmic Fourier decompositions $R_{\rm min}$, the corotation radius $R_{\rm co}$, the Outer Lindblad Resonance $R_{\rm olr}$, the rotation ratio $\mathcal{R}$, the pattern speed Ω_p, the time averaged mean value of the Q parameter $\bar{Q}$, the time averaged pitch angle of the stellar spiral arms $\psi_{\rm sta}$, the time averaged pitch angle of the gaseous spiral arms $\psi_{\rm gas}$, the time averaged stellar flocculence $\mathcal{F}_{\rm sta}$, the time averaged gaseous flocculence $\mathcal{F}_{\rm gas}$, the time averaged stellar asymmetry $\mathcal{A}_{\rm sta}$, the time averaged gaseous asymmetry $\mathcal{A}_{\rm gas}$, the time averaged stellar arm-inter arm contrast $\mathcal{C}$, the number of particles lost during the simulation $N_{\rm lost}$ and the maximum displacement of the centre of mass during the simulation $C_{\rm mass}$.